Check for updates

REVIEW SUMMARY

ORIGINS OF LANGUAGE

What enables human language? A biocultural framework

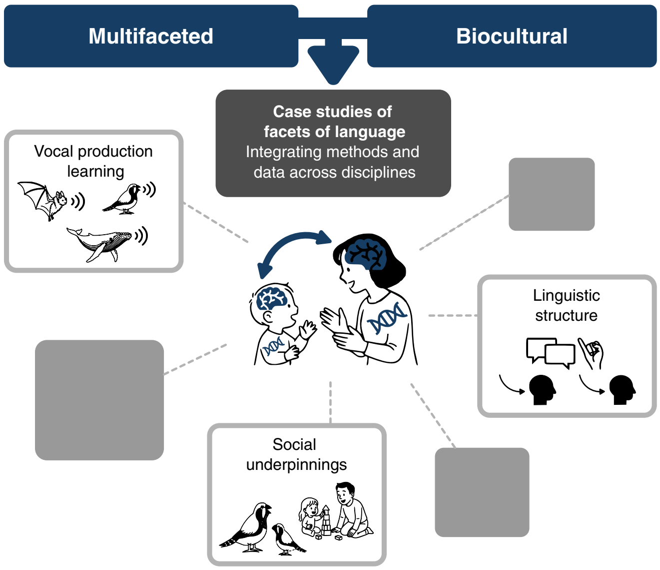
Inbal Arnon*, Liran Carmel, Nicolas Claidière, W. Tecumseh Fitch, Susan Goldin-Meadow, Simon Kirby, Kazuo Okanoya, Limor Raviv, Lucie Wolters, Simon E. Fisher*

Full article and list of author affiliations:
https://doi.org/10.1126/science.adq8303

BACKGROUND: Explaining the origins of language is a key challenge in understanding ourselves as a species. We present an empirical framework that draws on synergies across scientific disciplines to facilitate robust studies of language evolution. The approach is multifaceted, seeing language emergence as dependent on convergence of multiple capacities, each with their own evolutionary trajectories. It is explicitly biocultural, recognizing and incorporating the importance of both biological preparedness and cultural transmission as well as interactions between them. Biocultural and multifaceted perspectives are increasingly appreciated, but there remains a need to integrate them within a unified framework and demonstrate how this advances understanding. We do so in this paper through three case studies examining the evolution of different facets of human language (vocal production learning, linguistic structure, and social underpinnings), each synthesizing the latest findings from multiple fields to generate valuable insights and setting a new agenda for future research.

ADVANCES: Case study 1 considers vocal production learning, an organism's capacity to enlarge and modify its repertoire of vocalizations based on auditory experience. This ability is crucial for learning spoken language and limited in nonhuman primates but has emerged in other branches of the evolutionary tree, including subsets of birds, bats, elephants, cetaceans, and pinnipeds. Bringing together data from molecular investigations of speech and language disorders, genetic manipulations in animal models, and studies of ancient DNA, this case study demonstrates how ancient genetic and neural infrastructures may have been modified and recombined to enable distinctive human capacities. Case study 2 examines the emergence of linguistic structure, a defining property of human language, using data from real-world cases of emergence (e.g., homesign and emerging sign languages); experiments recreating cultural evolution in the lab; and comparative studies of nonhuman animals, including songbirds and primates. This case study highlights the importance of transmission and interaction, suggesting that emergence of structure involves a combination of biological, cognitive, and cultural conditions: Although some (or all) traits are shared with other species, their combination may be specific to humans. Case study 3 focuses on the social underpinnings of communication across species. Social interaction contributes to language learning in humans and learned behaviors in other species with culturally transmitted communication systems, such as songbirds. But humans also demonstrate a strong internal drive to socially share information, which is rarely observed in nonhuman animals.

An empirical framework for the study of language evolution



Our framework is both multifaceted and explicitly biocultural and is grounded in empirical investigations spanning a diverse array of fields and benefiting from major advances in methods, analyses, and theory. We demonstrate the potential of this integrated framework through three example case studies, each focused on a different facet with its own distinctive evolutionary history (other facets relevant to language, not discussed in the present paper, could be similarly investigated under this framework; these facets are represented by the empty gray boxes). Drawing on data from multiple disciplines and several species, including humans, primates, and songbirds, the case studies highlight the importance of both biological preparedness and cultural processes, as well as the interactions between them, in the emergence of language.

OUTLOOK: Drawing on diverse data, the case studies show how modification and recombination of abilities present in nonhumans, combined with intra- and intergenerational cultural transmission, may yield linguistic capacities in our own species. This perspective increases the range of species relevant for understanding language origins, as different abilities may be present across different branches of the evolutionary tree. The case studies also demonstrate the value of explicit biocultural framing, where both biological preparedness and cultural evolution shape language emergence. Language evolution is impacted by three distinct but interacting timescales: the individual (language learning), the community (cultural evolution), and the species (biological evolution). Understanding how the timescales interact and constrain one another requires synergies of data, methods, and fields. A recurring theme and avenue for future research is the role of biological reward systems in language evolution, including the motivation to communicate and endogenous and exogenous rewards for successful imitation and communication. Our integrative framework shows how research across disciplines and methods can advance understanding of a fundamental question in human evolution. □

---

*Corresponding author. Email: inbal.arnon@mail.huji.ac.il (I.A.); simon.fisher@mpi.nl (S.E.F.) Cite this article as I. Arnon *et al*., *Science* **390**, eadq8303 (2025). DOI: 10.1126/science.adq8303

Downloaded from https://www.science.org at Radboud University Nijmegen on February 05, 2026

ORIGINS OF LANGUAGE

What enables human language? A biocultural framework

Inbal Arnon[1]*, Liran Carmel[2], Nicolas Claidière[3,4,5], W. Tecumseh Fitch[6], Susan Goldin-Meadow[7], Simon Kirby[8], Kazuo Okanoya[9], Limor Raviv[10,11], Lucie Wolters[12], Simon E. Fisher[13,14]*

Explaining the origins of language is a key challenge in understanding ourselves as a species. We present an empirical framework that draws on synergies across fields to facilitate robust studies of language evolution. The approach is multifaceted, seeing language emergence as dependent on the convergence of multiple capacities, each with their own evolutionary trajectories. It is explicitly biocultural, recognizing and incorporating the importance of both biological preparedness and cultural transmission as well as interactions between them. We demonstrate this approach through three case studies that examine the evolution of different facets involved in human language (vocal production learning, linguistic structure, and social underpinnings).

Human language is a distinctive trait of our species, yet its origins are still not understood (*1*, *2*). The lack of any fossil record of the first language(s) together with many unknowns about human evolution and animal communication have led some to conclude that this question is scientifically intractable (*3*). We propose instead that studying language evolution lies well within the scope of scientific enquiry when new sources of data and theoretical perspectives are incorporated. We present an empirical biocultural framework for research on language evolution, applying it to three case studies, each examining a different facet involved in human language. Our aim is neither to comprehensively review the many existing theories nor to advocate for our own special one but to set an agenda for future interdisciplinary research, highlighting promising avenues.

This approach is multifaceted in seeing language emergence as dependent on convergence of multiple capacities (physical, cognitive, social, and cultural), each with its own developmental and evolutionary trajectories [see (*4*, *5*)]. Proposed facets include those related to production and perception of signals (e.g., vocal learning), systematic organization of language (e.g., linguistic structure), and communicative motivations (e.g., aspects of social behavior). A facet does not have to be specific to humans or language to offer explanatory value: Similar to the evolution of other complex biological systems (e.g., the eye), the emergence of language can be explained by modifications and recombination of ancestral infrastructures and exaptation of existing structures (*6*, *7*). This reflects a move away from "silver bullet" views of language evolution [e.g., (*8*, *9*)], where human distinctiveness is defined by just one explanatory factor (e.g., a single genetic mutation). Although such accounts have been historically prolific, persisting in some academic discourse and popular science writing, they are untenable in light of modern biology. Considerable evidence from multiple sources indicates that no one thing itself was enough to "give us language" or "make us human" (*10*–*12*). The multifaceted perspective calls for empirical investigations of larger historical windows. Although common wisdom was that language is specific to anatomically modern humans, appearing on the *Homo sapiens* lineage within the past 50 to 150 thousand years (kyr) [e.g., (*8*, *9*)], contemporary data suggest that deeper evolutionary timescales, those of hundreds of thousands (perhaps millions) of years, are more plausible (*10*, *13*). Even if the language system as we know it in present-day humans only emerged recently, then different facets may have evolved over longer timescales, under different selective pressures.

Our approach is also biocultural, recognizing and incorporating biological preparedness, cultural processes, and the interactions between them as key factors in language emergence. Understanding biological preparedness, including innate learning mechanisms and biases, is necessary to explain the distinctiveness of human language and helps guide comparative research on nonhuman species. However, no human infant develops a fully structured language in isolation; such languages arise only after extended social and communicative interaction [e.g., (*14*, *15*)]. Over generations, learners progressively systematize language through communication and cultural transmission (*16*–*18*), processes shaped by properties of the individual and the community (*19*, *20*). Computational simulations, experiments, and real-world cases of emergence identified specific cultural processes necessary for structured language to emerge. One reason nonhuman species lack human-like language may be their limited biological capacity to support these cultural processes. Notably, biology and culture can interact in complex nonintuitive ways. For example, the emergence of more complex communication systems can increase selective pressure on the cognitive mechanisms required to learn and produce complex signals. This could result in virtuous cycles of gene-culture coevolution (Fig. 1), making iterated biocultural processes central to understanding language emergence. Crucially, both classes of phenomena, biological and cultural, along with their interactions, can be empirically investigated in humans, nonhuman animals, and simulated or artificial agents (*21*).

Biocultural and multifaceted perspectives are increasingly appreciated in discussions of language evolution, but there is a need to integrate them in a unifying framework and show concrete examples of how that advances understanding. We demonstrate application of an integrated framework through three case studies, targeting different facets important for language emergence: (i) Vocal production learning: the ability to modify vocalizations based on experience, critical for acquiring spoken language; (ii) language structure: the systematic ways in which linguistic elements relate to one another, underlying the productivity of human language; and (iii) social underpinnings: behaviors and processes that facilitate social interaction, enabling cultural transmission of language. These are not claimed as the sole or primary facets relevant to language evolution but are used to demonstrate the value of a biocultural framework.

Case study 1: Vocal production learning

Human language is inherently multimodal, expressible through speech, sign, writing, or touch (*22*, *23*). However, when available, speech is the primary modality across societies. Its acquisition depends on auditory-guided vocal production learning (VPL): the ability of an organism to flexibly enlarge and modify its repertoire of vocalizations based on auditory experiences (*24*). VPL is critical for learning the sounds and

Downloaded from https://www.science.org at Radboud University Nijmegen on February 05, 2026

---

[1]Department of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel. [2]Department of Genetics, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel. [3]Centre de recherche en psychologie et neurosciences, UMR7077, Université Aix-Marseille/CNRS, Marseille, France. [4]Station de Primatologie-Celphedia, CNRS UAR846, Rousset, France. [5]Institute for Language, Communication and the Brain, Université Aix-Marseille, CNRS, Aix-en-Provence, France. [6]Department of Behavioral and Cognitive Biology, University of Vienna, Vienna, Austria. [7]Departments of Psychology and Comparative Human Development, University of Chicago, Chicago, IL, USA. [8]Centre for Language Evolution, University of Edinburgh, Edinburgh, UK. [9]Advanced Comprehensive Research Organization, Teikyo University, Tokyo, Japan. [10]LEADS group, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands. [11]Donders Center for Cognition, Radboud University, Nijmegen, Netherlands. [12]Department of Cognitive and Brain Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel. [13]Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands. [14]Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, Netherlands. ***Corresponding author. Email: inbal.arnon@mail.huji.ac.il (I.A.); simon.fisher@mpi.nl (S.E.F.)**

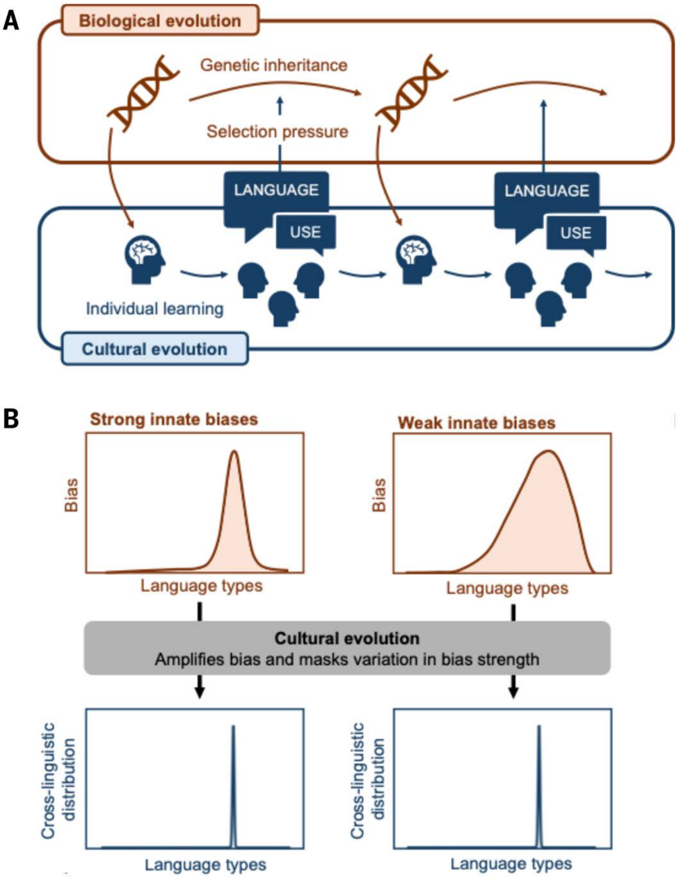


Fig. 1. Gene-culture coevolution model. Interacting processes operating on different timescales, from milliseconds to millennia, shape language emergence. (**A**) Processes of language use operate at the shortest timescale, as individuals comprehend and produce utterances in ongoing conversation. Learning to form these utterances (learning sounds, words, and rules) happens over a lifetime of exposure to the language of the community. Zooming out further, the structure of a specific language emerges and changes through cultural evolution, as knowledge of language is passed from one generation to the next. Lastly, the cognitive and anatomical machinery that allows humans to learn and use language has been subject to genetic evolution over the course of human evolution. The processes of biological and cultural evolution interact to produce a dual-inheritance system (*154*). Features of languages are inherited culturally, and the mechanisms that support such cultural inheritance are themselves inherited genetically. These processes may interact in complex and interesting ways, studied using mathematical and computational models that include all three timescales: individual learning and use, cultural evolution, and biological evolution. (**B**) One prominent approach, iterated Bayesian learning (*155*), treats learning as a process of inductive inference, combining utterances that the learner observes with a prior bias favoring particular types of languages. Cultural evolution is modeled as a process in which the languages inferred by one generation provide data observed by the next generation of learners. Iterated Bayesian learning allows us to compute expected results of cultural evolution for any hypothesized prior bias learners might have along with a model of how language is used for communication (*19*). This approach has been extended to the full dual-inheritance model by assuming that priors for learners are shaped by their genes, and these genes are selected based on communicative effectiveness of the individuals in the population (*156*). One notable finding is that the existence of cultural evolution tends to weaken inductive biases in language learning (*156*). Cultural evolution amplifies weak biases in individual learners, such that weak biases have the same outcome at the population level as strong constraints would. If strong biases are costly to maintain (e.g., by being more subject to mutation pressure), then weak biases are the inevitable consequence. This is surprising given previous work on the evolution of learning, which suggests the opposite: that learning can make evolution of innate constraints more likely (*157*).

open-ended vocabulary of language. Nonhuman primates appear much more limited than humans in their capacity to produce new vocalizations, but these abilities have emerged in other species, including subsets of birds, bats, cetaceans, pinnipeds, and elephants (*24*). There is increasing evidence that the independent appearance of VPL on different branches of the evolutionary tree involves deep homology (*25*, *26*), a phenomenon where convergently evolved traits recruit similar underlying genetic regulatory mechanisms across species (*27*). This aligns with the idea that some facets of language rely on ancient genetic and neural infrastructures, modified and recombined to enable more complex systems or abilities. The relevance of deep homology for understanding VPL is exemplified by studies of the *FOXP2* gene.

FOXP2 was originally discovered by using human genetics tools (Table 1) to investigate the biological bases of developmental speech and language disorders (*28*). Given adequate exposure to spoken language (and in the absence of sensory disorders), most children become proficient language users within the first years of life. However, there are unusual cases where this process goes awry. Before the advent of molecular methods, studies comparing identical and nonidentical twins and documenting recurrence of cases within families suggested that genetic factors play a role in these otherwise unexplained disorders without pinpointing the genes involved. In 2001, a rare pathogenic DNA variation in *FOXP2* was found to disturb development of the coordinated sequencing of mouth and face movements underlying proficient speech (childhood apraxia of speech) in a large family known as the "KE family" (*29*) (Fig. 2A). Multiple independent cases of people carrying *FOXP2* disruptions have since been reported, with developmental speech deficits being the most consistent consequence (*30*).

Once *FOXP2* was identified in humans, researchers looked for versions of the gene in other species, retracing its evolutionary history. Cross-species DNA comparisons (Fig. 2B) revealed that *FOXP2* is not specific to humans but is evolutionarily ancient, with similar versions in disparate vertebrates, including mammals, birds, reptiles, fish, and amphibians (*26*, *31*). There is high species-wide concordance in the places where this gene is active in the developing or adult central nervous system, including in subsets of neurons in the cortex or pallium, basal ganglia, thalamus, and cerebellum. These findings suggested that contributions of *FOXP2* to human speech may be built on ancient evolutionary pathways involved in motor-skill learning and vocal behaviors (*32*). Such deep evolutionary conservation means that genetic manipulations of versions of *FOXP2* in nonhuman species can help elucidate its functions and how these influence brain plasticity and behavior (*33*–*38*).

For example, though mice have very limited VPL capacities (*39*), valuable insights were gained from mouse models engineered to carry *FOXP2* disruptions known to cause speech disorders in humans (Fig. 2C). Mice carrying the pathogenic variant of the KE family show motor skill learning deficits and altered neuronal properties in basal ganglia and cortex (*33*–*36*), among other findings. Investigating nonhuman animals that are vocal learners, such as songbirds, is even more revealing (Fig. 2D). Male zebra finches sing structured songs comprising vocal elements (syllables) arranged in a stereotyped sequence, which they learn as juveniles by listening to adult males (*37*). During this developmental period of plasticity, *FoxP2* (the avian version of *FOXP2*) has elevated activity in Area X, a basal ganglia structure that is crucial for VPL (*37*). Experimentally reducing *FoxP2* Area X activity interferes with song learning and variability, potentially mediated by disturbed dopaminergic signaling (*37, 38*). Thus, impacts of this gene on brain plasticity linked to sensorimotor functions and motor skill learning may have been independently recruited toward VPL in disparate species (i.e., supporting speech in humans and song in zebra finches). Most recently, genome-wide investigations of >200 mammals with different vocal-learning capacities pinpointed multiple additional genetic loci as candidates for cross-species involvement in VPL (*40*).

Downloaded from https://www.science.org at Radboud University Nijmegen on February 05, 2026

Table 1: Identifying genetic links to language through genomic studies in modern humans. Researchers can investigate genetics of relevant pathologies (childhood apraxia of speech, developmental language disorders, etc.) by identifying genetic correlates of individual differences in language-related skills in the general population, exploiting advances in molecular methods and analytic approaches.

Type of DNA variation	Biological impact	Molecular methods	Typical study designs	Examples from the literature	Linking to evolution
Rare gene disruptions	Rarely, a change at a single genetic locus can be sufficient to substantially derail language development.	Advances in next-generation sequencing now allow rapid reading of almost all of a person's genome at high resolution at a fraction of the cost of classical methods.	Pathogenic variants can be identified by analyzing DNA of relatives in multigenerational families where multiple individuals have a developmental speech and/or language disorder.	The first rare gene variants in childhood apraxia of speech were discovered by studying a three-generation family before the advent of next-generation sequencing (*29*).	The evolutionary history of genes implicated in speech and/or language disorders can be retraced by comparing to versions found in extinct archaic hominins and extant apes and testing for evidence of Darwinian selection at these genomic loci on the lineage leading to *H. sapiens* (*177*).
			A complementary approach investigates de novo cases of disorder (where parents or siblings are unaffected) to identify pathogenic DNA variants that are only present in the affected child.	Whole-genome sequencing in speech apraxia has since identified pathogenic de novo disruptions of multiple candidate genes, with regulatory roles in early brain development (*178*).	
Common variation	Many studies focus on single-nucleotide polymorphisms (SNPs) found at >1% frequency in the general population. Any one SNP by itself has little impact, but combinations of many such variants across the genome may jointly explain a significant proportion of trait variance.	High-throughput low-cost genotyping technologies, such as DNA microarrays, make it possible to capture allelic variation at millions of SNPs in large samples. These technologies fueled the rise of genome-wide association studies (GWASs) that systematically screen vast numbers of SNPs, testing each for a relationship with a trait of interest.	One GWAS design is a case or control study assessing contributions of common DNA variation to a disorder (or categorical trait). GWAS designs can also identify associations of SNPs with individual differences in quantitative traits. Because the effect size of one SNP may be tiny, cohorts of tens (even hundreds) of thousands of people are needed to give adequate power while adjusting for substantial multiple testing.	In a multicohort GWAS study of individual differences in quantitatively assessed reading- and language-related skills involving <34,000 participants, researchers could capture up to 26% of trait variability with common DNA variation (*179*).	Findings on genetic contributions to individual differences in language-related skills and/or neural infrastructure in living humans can be integrated with information about evolutionary signatures across the genome over a range of different timescales in primate and hominin history. For example, a UK Biobank study used this approach to uncover effects of human-gained regulatory elements on left-hemisphere brain regions related to speech, among other findings (*180*).
			Studies of genetic associations with language-related traits can be extended to individual differences in brain structure and function, assessed with neuroimaging. Effect sizes of individual SNPs are small even for traits measured with magnetic resonance imaging (*181*). With availability of neuroimaging and DNA data in large biobanking resources, it is now possible to carry out GWAS studies of neural circuits involved in language processing.	GWAS investigations of structural and functional connectivity in the brains of ~30,000 to 32,000 participants in the UK Biobank have given new insights into how genetic variants contribute to language-related circuits in the human brain [e.g. (*182*)].	

Downloaded from https://www.science.org at Radboud University Nijmegen on February 05, 2026

Identifying genes contributing to VPL across species allows researchers to use a transformative new data source to test hypotheses about language evolution: ancient DNA. In the past 15 years, it became possible to obtain high-quality sequence information from nuclear genomes of Neanderthals and Denisovans, extinct hominins that shared with modern humans a most recent common ancestor ~600 kyr ago (*41*). These archaic hominins existed until a few tens of thousands of years ago, temporarily overlapping with *H. sapiens* at sites across Eurasia (*42*). Analyzing ancient genomes enables detection of DNA variants that arose in modern humans after our split from Neanderthals and Denisovans (*43*). It also enables detection of variants that we share with archaic hominins, but that are distinct from those in extant nonhuman apes. These more ancient variants arose after our split from the common ancestor of chimpanzees and bonobos ~6 million years (Myr) ago but before the split between modern and archaic humans (*44*). Applied to *FOXP2*, this approach identified two amino acid changes in the protein that it encodes, both arising on the *Homo* lineage during the 6 Myr to 600 kyr time window (*45*). Researchers used genetic manipulations to introduce the hominin amino acid substitutions into mice, observing varied effects on vocal behaviors and basal ganglia functions (*46*–*49*) (Fig. 2E). Thus, by identifying evolutionary variants in genes implicated in facets of language and introducing them into nonhuman animals, we can investigate whether these variants affect brains and behavior in ways that might be relevant to language emergence. Despite this promise, we stress that no single genetic change is, by itself, sufficient to yield a vocal-learning brain (*50*).

Evolution acts not only through genetically specified changes to protein structure and function but also by modifying where and when

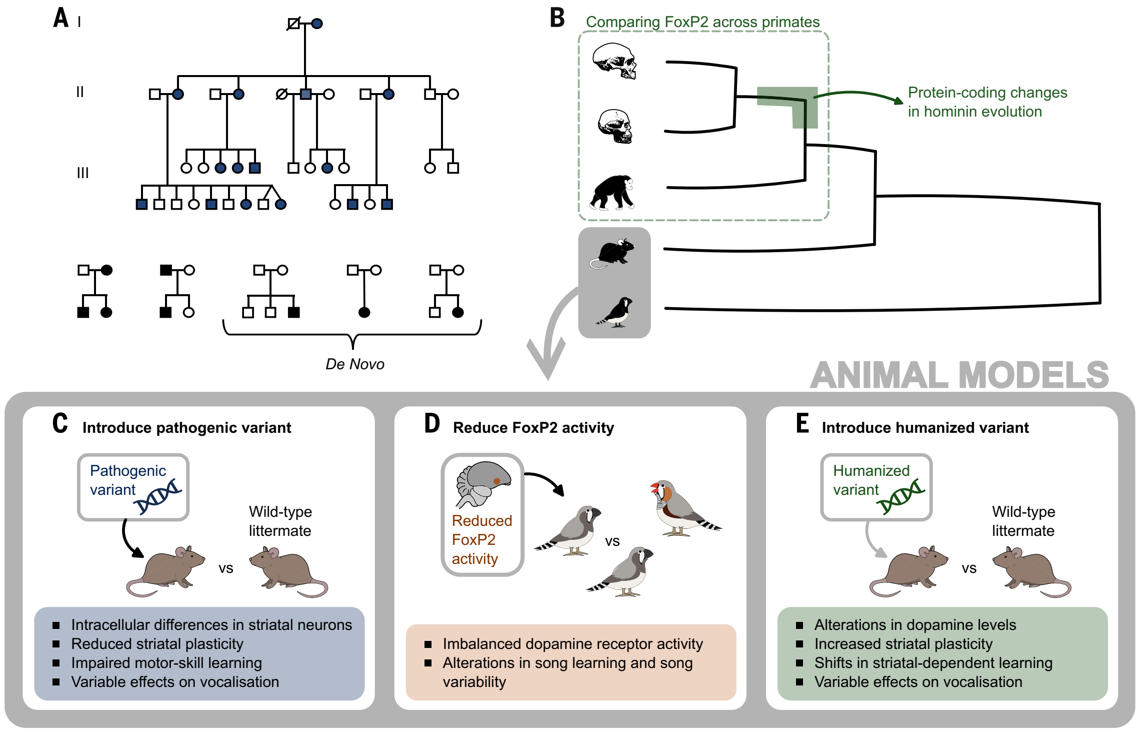


Fig. 2. Investigating evolution of vocal production learning with tools of molecular genetics: *FOXP2* as an example. (**A**) The starting point was a three-generation family, the KE family, in which half of the relatives (shaded symbols) were affected by a neurodevelopmental disorder primarily involving childhood apraxia of speech, accompanied by expressive and receptive language deficits (top). The affected relatives carried a change of one DNA letter (nucleotide) in the *FOXP2* gene (*29*). This small change in DNA alters the amino acid sequence and, hence, the shape of a key part of the regulatory protein that *FOXP2* encodes, stopping it from functioning in its normal way. Advances in DNA sequencing led to identification of >28 additional individuals (from 17 families) carrying different pathogenic single-nucleotide variants of *FOXP2*, with problems in speech development being the most common feature found in these cases (*30*). As shown in the bottom of the panel, although pathogenic variants were sometimes inherited from affected parents, in many of the cases, they arose de novo in children with unaffected parents. (**B**) Comparisons of DNA sequences across different species (comparative genomics) identified versions of *FOXP2* in distantly related vertebrates, including mammals, birds, reptiles, fish, and amphibians (*26*,*31*), showing that the gene has a deep evolutionary history. Against this background, integration of findings from extant apes and extinct archaic hominins revealed that changes in the amino acid sequence of the encoded protein occurred on the *Homo* lineage after splitting from the common ancestor of chimpanzees and bonobos (*45*). (**C**) Researchers engineered mouse models that carry the same pathogenic variant that causes speech problems in the KE family. Investigations of these mice reported motor skill learning deficits (*33*), reduced plasticity in the striatum (part of the basal ganglia) (*34*), disturbed intracellular “protein motors” in striatal neurons (*35*), and loss of neuronal homeostasis in deep-layer cortical neurons (*36*), among other findings. (**D**) Moving to songbirds, lentivirus-mediated RNA interference has been used to reduce activity of *FoxP2* (the avian equivalent of *FOXP2*) in Area X, a key nucleus in the basal ganglia of male zebra finches. Such studies uncovered effects of the gene on song learning and the control of song variability, potentially mediated by changes in dopaminergic signaling (*37*,*38*). (**E**) When researchers used genetic manipulations to introduce hominin amino acid substitutions of *FOXP2* into mice, they observed regional changes in dopamine levels and increased plasticity in the striatum (*46*). Motor skill learning and vocal behaviors of adult male mice were unaffected according to one study (*47*), but later investigations of female and male vocalizations in social contexts found that the partially “humanized” mice used higher frequencies and more complex syllable types (*48*). Another study of these mice uncovered different patterns of striatal-dependent stimulus-response association learning (*49*). Overall, this suite of human and animal model studies shows how genes involved in VPL can be empirically investigated across species to give new insights into evolutionary pathways.

Downloaded from https://www.science.org at Radboud University Nijmegen on February 05, 2026

genes or proteins are active in development and adulthood (*51*). These effects are mediated by a wide variety of regulatory elements in the genome. Many of the DNA variants distinguishing us from other extant apes and extinct hominins may lie within such elements. For example, among primates, *FOXP2* shows human-specific expression in microglia, the primary immune cells of the brain, although the regulatory elements responsible for this specificity are not yet described (*52*). Moreover, innovations in paleoepigenetics take advantage of degradation processes in ancient DNA to reconstruct patterns of methylation, a chemical modification that helps mediate changes in gene activity without changing DNA sequence itself. This approach revealed changes in gene regulation differentiating Neanderthals and/or Denisovans from modern humans (*53*). Several of the identified modern human-specific gene expression changes are associated with genes that affect the face and voice and may underlie characteristics that are specific to modern humans (*54*).

Additional insights into VPL evolution come from considering developmental processes. Take babbling, an early, self-initiated form of vocal production in infants that starts as simple and repetitive verbal “play” but gradually approaches a mature form. Babbling-like behaviors have been documented in humans, songbirds, parrots (*55*, *56*), and vocal-learning bats (*57*) but are not common in species lacking VPL. Manual “babbling” is seen in hearing and deaf human babies exposed to signed language from birth (*58*), illustrating both the multimodality of language and the role of babbling in language acquisition. Deaf babies also babble vocally, but this babbling does not progress normally when appropriate input models are inaccessible (*59*), demonstrating how biological preparedness and environmental input interact in language learning.

Babbling (termed “subsong” in birds) is self-generated and self-rewarding, occurring without immediate environmental triggers or exogenous rewards. Thus, part of the biological preparedness for VPL includes an endogenous reward system, making vocal play enjoyable to the young organism without feedback from parents or others. Although little is known about the underlying circuitry in humans, recent evidence implicates endogenous reward in songbird vocal behaviors. Avian song learning begins with a sensory learning period in which the bird stores auditory templates of exemplars of its species' song. This involves an endogenously rewarding listening process: Juveniles selectively attend to and memorize songs of their own species, indicating that hearing them is intrinsically rewarding (*60*). Endogenous reward is key during the subsequent sensory-motor learning period, when spontaneous subsong is gradually adjusted, without external feedback, to approach stored adult template(s) (*61*–*64*). Vocal practice correlates temporally with neural expression of opioid markers and increased activity in reward systems (*61*), and blocking dopamine receptors in the basal ganglia in young zebra finches impairs song copying (*63*). Later in development, both infant babbling and bird subsong are impacted by social reinforcement (see case study 3), but the early self-reinforcing stages are required to provide raw material for later, exogenously directed, learning. Evolution of VPL may therefore depend both on changes to neural circuits involved in learning and on those underlying endogenous reward.

Case study 2: The emergence of linguistic structure

Human language shows systematic structure at multiple levels and of multiple kinds. Elements can be combined in productive ways, with the meaning of larger units composed of the meanings of their parts (e.g., "cat," "cats," "big cats"). There is ongoing debate on how to define and quantify this systematicity. In this case study, we classify a behavior as systematic when it can be described more concisely as a set than as a collection of individual instances. "Grammars" in linguistics (in the most theory-neutral use of that term) refer to these shorter descriptions and are possible because language is systematic. For example, it is more concise to describe formation of the regular English plural using the rule "add -s to the singular form" than to list all plural forms. Although prevalent in human language, systematicity is rare in the vast majority of communication systems in nature. An extensive literature investigates neural correlates of systematic language structure [see (*23*, *65*)]; comparing those circuits across humans and nonhuman primates offers ways to study their evolution [e.g., (*66*–*68*)]. Though illuminating, current literature leaves open how linguistic structure first came about.

Over the past 25 years, various experimental and computational methods have been developed to study origins of systematic linguistic structure (*17*, *69*–*71*) and ask how that structure is shaped by cognitive and communicative pressures. Specifically, language must serve the communicative needs of interacting language users and be learnable by subsequent generations of language users. Because language is culturally transmitted (passed on by being repeatedly learned and used by multiple generations), its structure is impacted by the interplay of communicative and cognitive forces. To demonstrate, we here focus on one feature: combinatoriality, the fact that language has units that can be recombined, at multiple levels of linguistic analysis (*19*, *72*, *73*). For example, sounds can be combined into words, and words can be combined to form sentences.

How did this combinatoriality emerge? We have no access to or record of hominin communication systems preceding modern human languages. However, insights can come from real-world cases of emergence (*74*) and lab-based studies recreating evolutionary processes in miniature (*20*, *75*, *76*). Two real-world settings illuminate the pressures and biases impacting the emergence of linguistic structure in modern-day humans. One is homesign, gesture systems created by individuals whose hearing loss prevented them from accessing spoken language and who were not exposed to sign language (*77*). Another is emerging sign languages, where new signed languages develop in communities with a high proportion of deaf individuals, lacking access to an established signed language. An influential example is Nicaraguan Sign Language (NSL), which spontaneously emerged when homesigners were first brought together in the mid-1980s (*78*). Examining how linguistic structure in such systems changes over time demonstrates how individual learning and cultural transmission impact the emergence of structure.

Researchers have documented and compared linguistic structure in solo language creators (homesign), homesigners who came together and formed the first NSL cohort (NSL1), and subsequent cohorts of signers who entered the community after the language began and were thus exposed to a language model (NSL2, NSL3, etc.). Some linguistic structural properties are observed in solo creators, irrespective of the specific cultural environment they were raised in. Child homesigners in the US, China, Turkey, and Nicaragua use gestures to refer to objects, actions, and attributes and combine them into strings characterized by consistent word order. For example, gestures for the object of an action appear before gestures for the action, yielding "grape-eat" as opposed to "eat-grape" (*77*, *79*) (Fig. 3). These same properties appear across societies, without exposure to linguistic input, indicating shared human cognitive biases [e.g., hierarchical structure (*80*)]. Other properties emerge only after homesigners came together to form NSL1 [e.g., a stable lexicon (*79*)], highlighting the importance of communication with others in shaping linguistic structure. Still other linguistic properties are not produced by homesigners or in NSL1 but tend to appear only after the emerging sign languages are transmitted to new learners. For example, spatial modulations are rare in NSL1 but commonly used to indicate shared reference in NSL2 (*14*).

A well-studied feature of NSL, relevant to combinatoriality, involves how complex motion events are conveyed. Consider a ball bouncing down a hill. In early NSL cohorts, both the ball's path (downwards) and its movement manner (bouncing) are typically conveyed simultaneously [(*78*) but see (*81*)]. However, later cohorts typically segment path and manner into separate parts, yielding a more combinatorial flexible system. Thus, transmission from one cohort to another seems important for building and enhancing combinatoriality. In an experimental analog of the homesign situation, nonsigning participants are asked to convey meanings with gesture and no speech (*82*). When presented with complex motion events possessing manner and path components, participants prefer to convey both aspects simultaneously even if expressed separately in their spoken language (*83*). However, gestures and vocalizations beginning as holistic (e.g., simultaneously expressing motion and speed) become more segmented and linearized during ongoing dyadic communication (*84*, *85*).

Experiments can recreate processes of cultural evolution using a paradigm called iterated learning. In these studies, a participant learns from the output of a previous participant in the experiment, creating multiple simulated "generations" (*17*). When silent gestures get transmitted in this way, there is a learning-driven preference for segmented manner and path. This preference is amplified over generations [as in NSL2 (*78*)] so that more systematic and combinatorial behaviors emerge (*84, 85*). More generally, productive units emerge during dyadic and group communication paradigms (*20*, *86*, *87*), a process enhanced by learning and transmission (*88*). The combined findings demonstrate that systems starting as a collection of wholes are gradually segmented and analyzed into productive parts, consistent with evolutionary approaches that propose a holistic origin for language (*89, 90*). Further evidence for the role of whole-to-part learning comes from first language acquisition (*91*) and homesign creation (*92*), where learners discover parts from unanalyzed wholes [e.g., "Ididit" > "I did it" (*93*)] in ways that facilitate the mastery and emergence of systematic structure (Fig. 3) (*94*, *95*).

Real-world language emergence and lab-based studies necessarily involve humans with modern brains but are nevertheless informative. Linguistic features evident in homesign provide insights into products of biological evolution. Properties that homesigners fail to develop but

Downloaded from https://www.science.org at Radboud University Nijmegen on February 05, 2026

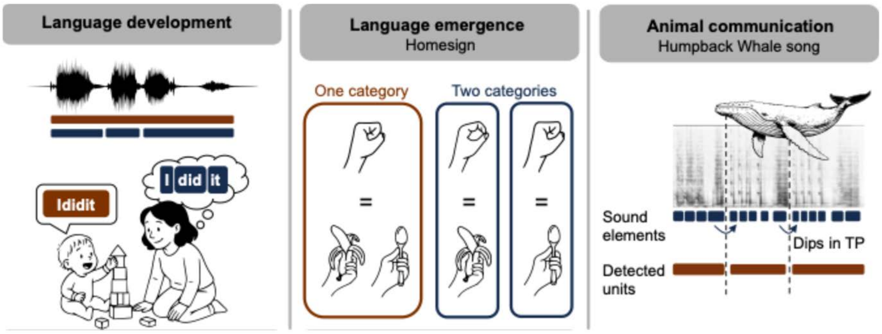


Fig. 3. Finding the right units. One of the challenges in studying communication in children and nonhuman animals is zeroing in on the right unit of analysis. This is challenging because the units we use to code data are influenced by hypotheses [explicit or implicit (*158*)], often based on our own categories. For example, when we describe early child language, we typically attribute individuated words to the child (left). But we might be wrong; a child might use a larger unit, treating several words as a single "chunk" (*91*,*159*). Infants extract single word units from the speech they hear, but they also extract larger units containing more than one lexical word (*91*,*93*). In fact, starting from larger units plays an important role in learning linguistic structure, particularly in learning grammatical relations between words (*91*, *94*, *95*), and in creating linguistic structure (*92*). One way to validate the categories we use is to find systematic patterns based on those categories, providing indirect evidence for the categories and also for their level of representation. For example, using semantic roles (patient, act, recipient, etc.) to categorize homesigners' gestures results in systematic orderings (patient-act, patient-recipient, and act-recipient), which validates coding at this level (*160*). But sometimes our coding system fails to produce systematic patterns. This may be the time to scrap the system and start again, coding at a level smaller than the one previously used (middle). For example, homesigners could vary thumb-to-finger distance so that the handshape in the gesture for banana grasping is distinct from that in the gesture for spoon grasping (as they are when these objects are actually grasped). Alternatively, homesigners could use the same handshape in both gestures, introducing one larger category for grasping objects <1 inch in diameter. To discover the homesigner's categories, we need to code in units that are smaller than the units on which those categories are based; otherwise, the categories may be created by us, not the child (*161*). When we seek the right units in nonhuman communication [e.g., gestures in great apes (*162*)], the challenge is greater because we have limited insight into the categories relevant to nonhuman animals (*163*) and must validate the categories in the animal itself [e.g., by using playback experiments, (*164*)]. Nonetheless, the approach of seeking out coherent patterns can also help reveal units in animal communication (right). For example, using transitional probabilities (TP) between syllables to segment humpback whale song [a cue used by human infants to segment speech (*113*)] uncovered statistically coherent subsequences whose frequency distribution followed a particular power law also found in all human languages (*165*). This points to a notable similarity between two evolutionarily distant species (whales and humans), united by having culturally transmitted communication systems. Debates about how to detect the appropriate units continue (*166*), with new perspectives coming from machine learning (*167*). In general, allowing for units at multiple levels of representation provides insight into structure in child language, homesign, and animal communication (*168*).

are found in emerging sign languages are good candidates for ones that require cultural evolution to emerge (*15*). With lab-based studies, researchers can manipulate communicative and cognitive pressures in ways that cannot be done in the real world to assess effects on emerging systems. Moreover, computational simulations, not yet discussed [but see Fig. 1 and (*96*, *97*)], allow investigation of learning biases that may differ from those of modern humans. Notably, it is still an open question of which, if any, capabilities underlying language structure are distinctly enhanced in humans. One component hypothesized as highly developed in humans and weak or absent in other species is "dendrophilia," a domain-general proclivity to infer tree structures from data whenever possible (*98*) (Fig. 4).

So far, we have demonstrated impacts of learning, communication, and cultural transmission in creating combinatorial structure in humans who are already biologically prepared for language. Roles of biological evolution can be investigated by using animal models. In this case, researchers control preparedness through the choice of species while experimentally manipulating social pressures and rewards. Two relevant models are songbirds and baboons, who differ from humans and each other in important ways. As introduced in case study 1, songbirds are vocal production learners with culturally transmitted song (*99*, *100*). Songbirds are also among the few nonhuman animals whose signaling exhibits combinatorial structure. Songs are constructed from individual elements that are repeated and recombined (*101*). Despite their "instinct to learn" a species-typical song (*102*), isolated songbirds deprived of appropriate input sing only a harsh, atonal "isolate song" (*101*). When isolate songs are transmitted to successive generations of learners through iterated learning, zebra finches converge within a few generations on a new (albeit well-formed and species-typical) song exhibiting combinatorial structure (*103*). Even if this developmental process is limited to one bird hearing its own songs played back after a delay, a more species-typical song results (*104*). Transmission over generations or iterations seems crucial for the emergence of elaborated species-typical structure, whether the starting point is isolate song in vocal-learning birds or homesign in humans.

Songbirds are biologically prepared to learn and transmit songs. By contrast, despite powerful learning abilities (*105*), nonhuman primates largely lack culturally transmitted communication systems, and the combinatoriality of their gestural signaling is not as productive or widespread as in humans or songbirds (*106*). Recent evidence indicates a degree of combinatoriality in vocalizations of some ape or primate species [e.g., (*107*, *108*)], though it is unclear whether they are used communicatively. Experiments with captive baboons suggest that systematically structured behavior can emerge in animals lacking it if cultural transmission is supported externally (*109*). In these studies, baboons are given exogenous rewards for reproducing randomly generated visual patterns. Cultural transmission is experimentally simulated by providing patterns from one baboon as input to another, creating an iterated learning design. Notably, systematically structured patterns emerge over iterations: although baboons lack biological preparedness for cultural transmission, when transmission is supported externally, structured patterns emerge. These experiments demonstrate that adding exogenous rewards for copying behavior (present endogenously in humans and songbirds) facilitates the emergence of systematicity.

Case study 3: Social underpinnings of language

Social interactions are key for first language acquisition, with individual learning typically occurring within interactive contexts (*110*, *111*). These interactions provide children with valuable linguistic input, facilitating learning in various ways (*112*). Although infants can learn in noncommunicative settings and do so in experimental contexts (*113*, *114*), many aspects of language learning are facilitated by social interaction (*112*). For example, contingent maternal responses yield more mature vocalizations in human infants (*115*). Similarly, learning of non-native phonetic sounds in infancy is enhanced by social interaction (*116*). Both homesign and emerging sign languages (case study 2) are motivated by the need and desire for social communication [although in homesign, the communications systems themselves are not shared with others (*117*)]. Later in development, there is evidence of bidirectional links between language abilities and aspects of

Downloaded from https://www.science.org at Radboud University Nijmegen on February 05, 2026

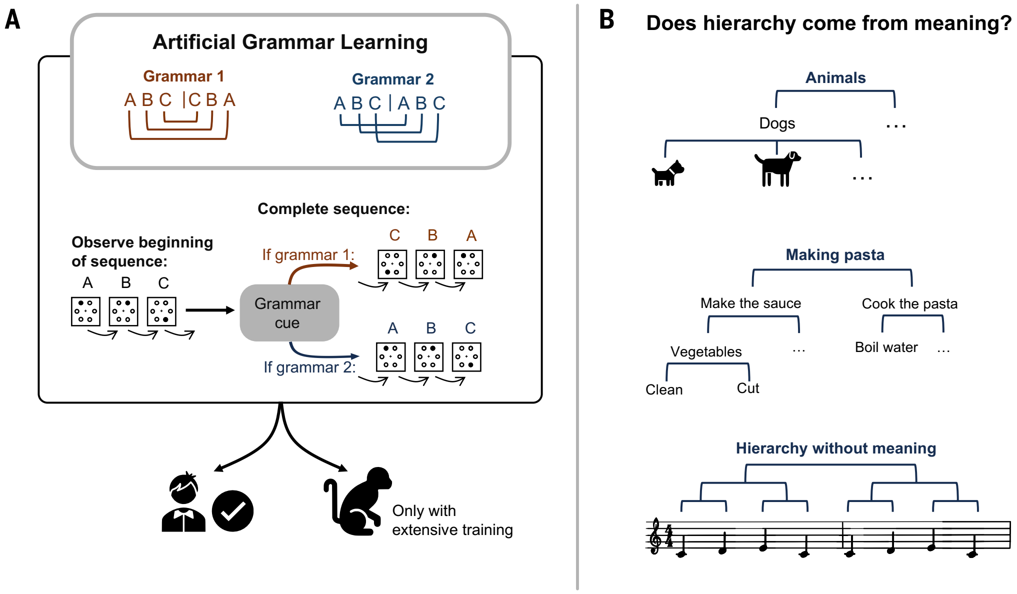


Fig. 4. The origins of hierarchical structure: Dendrophilia or semantics? An open question for the field concerns which, if any, capabilities underlying language are specifically enhanced in humans. One component hypothesized as highly developed in humans and weak or absent in other species is "dendrophilia," a domain-general proclivity to infer tree structures from data whenever possible (98). Dendrophilia combines a domain-general capacity to perceive hierarchical structures in stimuli with a strong preference to encode data into hierarchical structures. (**A**) This preference is often studied using Artificial Grammar Learning (AGL) experiments, where learners are exposed to sequences of stimuli whose appearance is governed by an underlying hierarchical grammar. If learners deduced the grammar, then they should be able to complete sequences in a way that conforms to it. Considerable experimental evidence from cross-species AGL research supports dendrophilia as being both highly developed and biologically canalized in humans and reduced or absent in other species studied to date (4,169,170). For example, a recent study found that, with adequate time and a consistent exogenous reward structure, macaque monkeys can learn hierarchical structures based on meaningless spatial or motor sequences, but learning required many months and tens of thousands of rewarded trials. By contrast, preschool children learn these same systems rapidly, in as few as six trials, with few or no errors (170). The presence of some hierarchical structure in homesign (case study 2) offers further evidence of biological preparedness for dendrophilia in our species (80). However, the finding that linguistic structure emerges gradually over generations indicates that cultural transmission is important for explaining hierarchical structure in fully developed languages (as for birdsong). Some precursor(s) of dendrophilia may be present in the motor and/or social domain in other primates, such as the perception and processing of complex dominance hierarchies, as shown in baboons and other socially complex species (171,172). (**B**) The problem of acquiring and using treelike structures may be greatly reduced in contexts involving signal or meaning pairs (as in human language). If semantics already possess hierarchical structure and signals are mapped onto this hierarchical meaning space, then it may strongly bias the learner to impose or perceive tree structure in the signals themselves. Notably, the existence of hierarchical structure in human music [e.g., (173)] or similar systems, such as bird or whale song [e.g., (174)], where signals do not map onto highly structured meanings, suggests that compositional semantic mappings are not necessary (or solely responsible) for hierarchical structure to emerge. Similarly, in AGL experiments, humans readily perceive hierarchical structure in meaningless visual strings (175,176). Better understanding of the neural mechanisms involved in structural learning and innovative new methods to "tweak" reward structures in animals can shed light on origins of hierarchical structure not just in language but also other domains, such as music and art.

social cognition or interaction (118–120). Gains in prosocial behaviors in early childhood (age 3 to 5 years) are significantly associated with later gains in verbal ability (age 5 to 11 years), and vice versa (121). Better language skills facilitate children's social–emotional competence, allowing formation of more prosocial, cooperative relationships (121–124). Conversely, language difficulties often associate with increased difficulty with social interactions (125, 126).

Social interaction contributes to individual learning in other species with culturally transmitted communication systems. Although we stress the importance of endogenous reward during early birdsong acquisition, exogenous reinforcement from social partners is crucial in later learning, maintenance, and modification of song (127, 128). Female cowbirds provide behavioral feedback (wing flaps) to courting males, influencing later use of particular syllables or syllable sequences (127). When parent zebra finches give behavioral feedback, juveniles develop more accurate copies of tutor song (128). Exogenous reward of vocal learning involves dopaminergic systems (129, 130). Socially tutored zebra finches show higher activity of dopaminergic neurons in the ventral tegmental area compared with that of birds who passively heard songs or untutored controls (129), and optogenetically blocking dopaminergic input to song circuits during social tutoring impairs song learning (64).

Beyond social underpinnings, which enable cultural transmission of language, humans have a strong, unparalleled internal drive to socially share information, including about inner states, emotions, and ideas using language ["Mitteilungsbedürfnis" (2)]. Sharing for sharing's sake is prevalent in humans but rare in nonhuman primates. Even language-trained apes, who master aspects of human sign language, show limited interest in using this to express things other than direct requests (131, 132). By contrast, our drive to share thoughts and feelings with others is so strong that humans create a communication system de novo even if one is not available (case study 2).

The social context together with our "mitteilungsbedürfnis" make language learning and use rewarding for humans. How might the evolution of reward mechanisms relate to emergence of communication systems? One evolutionary pathway of potential relevance is the process of domestication. The Bengalese finch, a domesticated variant of a wild songbird, the white-rumped munia (133–136), offers an example of relationships between changes in reward and communication systems. As in many domesticates, stress hormone levels are significantly lower in Bengalese finches compared with that in munias (135), and the former display less aggression (134) and explore new environments faster than their wild counterparts (136). Notably, Bengalese finches produce songs with greater phonological and syntactic complexity than those of munias (137). They are capable of learning munia songs, whereas munias struggle to master Bengalese songs (133). Domesticated Bengalese show higher concentrations of cerebral oxytocin than their wild ancestors (138, 139); oxytocin and dopaminergic reward systems are known to be closely interconnected (140, 141).

Did humans follow evolutionary pathways similar to those underlying animal domestication, where less aggressive individuals that were more prone to cooperatively interact had a greater likelihood to survive and/or reproduce? According to the human self-domestication hypothesis, such processes enhanced social learning and cultural transmission in humans (142–144). This could generate virtuous cycles at the

Downloaded from https://www.science.org at Radboud University Nijmegen on February 05, 2026

community level: Increased social reward for communication favors emergence of more advanced forms of communication, enabling larger in-groups and more interaction with nonkin, which results in even richer social interaction. Some support for links between greater communicative complexity and greater social complexity (operationalized by larger group sizes, more dense networks, etc.) comes from multispecies comparative research, from bats to primates (*145*, *146*). Human experimental data provide suggestive evidence: Artificial languages evolving in larger microsocieties of interacting participants show more systematic compositional structures, emerging faster and more consistently than in smaller groups (*20*).

Discussion

Our case studies include diverse data sources (behavioral, neural, genetic, and developmental) and adopt broad comparative perspectives, with particular focus on humans, primates, and songbirds. They demonstrate how facets involved in language emergence can be insightfully studied in nonhumans. No single method, tool, or model holds all the answers, and investigations of different facets may require different approaches. A common thread is that exaptation and recombination of abilities present in nonhumans, combined with intra- and intergenerational cultural transmission, can yield linguistic capacities in our own species. VPL, crucial for acquiring spoken language in humans, is a capacity that humans share with other species, appearing in diverse branches of the vertebrate evolutionary tree. Social underpinnings needed for human language transmission are documented in other species with culturally transmitted systems, but humans also demonstrate communicative tendencies rarely observed in nonhuman animals (e.g., our “mitteilungsbedürfnis”). Emergence of linguistic structure, a defining property of human language, involves a combination of biological, cognitive, and cultural conditions: Although some (or all) conditions are shared with various nonhuman species, the combination may be specific to humans.

The case studies demonstrate the value of explicit biocultural framing, showing how language emerges dynamically at three distinct but interacting levels: the individual (language acquisition and use), the community (cultural evolution and historical language change), and the species (biological evolution). The human ability to acquire and use language as well as languages themselves result from multiple interactions over time and among these levels, making all three important for understanding language emergence. Biological evolution generates the biological preparedness to acquire language shared by all human infants. Through individual learning in a social setting, the child acquires the language(s) of their community, which themselves develop through dynamic processes of cultural evolution. The ways in which these distinct levels interact, constrain, and structure one another can be nonintuitive. Understanding them requires combinations of data, models, and experiments.

A recurring theme and promising avenue for future research is the role of biological reward systems in language evolution. These systems include the motivation to communicate and both endogenous and exogenous reward for successful imitation and communication during language acquisition, use, and transmission. Although it is currently impossible to “insert” endogenous rewards for babbling into species that lack them, we can experimentally block such rewards, as shown for songbirds [case study 1 (*63*)]. Further, we can experimentally introduce exogenous rewards to trigger learning in species that lack endogenous reward systems for the learned behavior; potentially “unmasking” cognitive capabilities that were previously unexpressed in that species, as in baboon studies [case study 2 (*147*)]. These experiments can empirically circumvent the common criticism that investigating modern humans, who are already biologically prepared to acquire language, reveals nothing about how key facets evolved.

An open issue concerns modality. Like many researchers, we see language as inherently multimodal (*22*), and our case studies consider both spoken and signed languages. However, we have not discussed possibilities that the use of visual and auditory modalities may emphasize different types of structure. Because gestural capabilities of apes greatly exceed their vocal flexibility, some researchers have suggested that human language origins may be found in gesture and/or sign (“gestural protolanguage”) rather than speech (*148, 149*), whereas others vehemently deny this (*150, 151*). The multifaceted biocultural framework combined with data showing that, like humans, primates are multimodal communicators (*152, 153*) strongly suggest that gesture coexisted with vocal communication, and eventually language, all along. Furthermore, vocal-learning abilities might have already been enhanced in archaic hominins, such as Neanderthals (case study 1), although they likely lacked fully modern language. Thus, asking whether sign or speech came first is the wrong question. Productive future debate should center on how gesture and speech support one another and why language (unlike, say, birdsong) is flexible enough to be conveyed by radically different sensory systems.

REFERENCES AND NOTES

1. M. H. Christiansen, S. Kirby, “Language evolution: The Hardest Problem in Science?” in *Language Evolution*, M. H. Christiansen, S. Kirby, Eds. (OUP Oxford, 2003), pp. 1–15. doi: 10.1093/acprof:oso/9780199244843.003.0001
2. W. T. Fitch, *The Evolution of Language* (Cambridge University Press, 2010). doi: 10.1017/CBO9780511817779
3. M. D. Hauser *et al.*, The mystery of language evolution. *Front. Psychol.* **5**, 401 (2014). doi: 10.3389/fpsyg.2014.00401; pmid: 24847300
4. W. T. Fitch, Empirical approaches to the study of language evolution. *Psychon. Bull. Rev.* **24**, 3–33 (2017). doi: 10.3758/s13423-017-1236-5; pmid: 28150125
5. E. H. Lenneberg, *Biological Foundations of Language* (Wiley, 1967). doi: 10.1080/21548331.1967.11707799
6. S. J. Gould, E. S. Vrba, Exaptation-A Missing Term in the Science of Form. *Paleobiology* **8**, 4–15 (1982). doi: 10.1017/S0094837300004310
7. S. E. Fisher, G. F. Marcus, The eloquent ape: Genes, brains and the evolution of language. *Nat. Rev. Genet.* **7**, 9–20 (2006). doi: 10.1038/nrg1747; pmid: 16369568
8. R. G. Klein, *The Dawn of Human Culture* (Turner Publishing Company, 2007).
9. T. J. Crow, Schizophrenia as the price that homo sapiens pays for language: A resolution of the central paradox in the origin of the species. *Brain Res. Brain Res. Rev.* **31**, 118–129 (2000). doi: 10.1016/S0165-0173(99)00029-6; pmid: 10719140
10. A. Bergström, C. Stringer, M. Hajdinjak, E. M. L. Scerri, P. Skoglund, Origins of modern human ancestry. *Nature* **590**, 229–237 (2021). doi: 10.1038/s41586-021-03244-5; pmid: 33568824
11. K. Laland, A. Seed, Understanding Human Cognitive Uniqueness. *Annu. Rev. Psychol.* **72**, 689–716 (2021). doi: 10.1146/annurev-psych-062220-051256; pmid: 33400565
12. J. R. Hurford, *The Origins of Grammar: Language in the Light of Evolution II* (OUP Oxford, 2012).
13. P. Skoglund, I. Mathieson, Ancient Genomics of Modern Humans: The First Decade. *Annu. Rev. Genomics Hum. Genet.* **19**, 381–404 (2018). doi: 10.1146/annurev-genom-083117-021749; pmid: 29709204
14. A. Senghas, M. Coppola, Children creating language: How Nicaraguan sign language acquired a spatial grammar. *Psychol. Sci.* **12**, 323–328 (2001). doi: 10.1111/1467-9280.00359; pmid: 11476100
15. S. Goldin-Meadow, What the hands can tell us about language emergence. *Psychon. Bull. Rev.* **24**, 213–218 (2017).
16. L. Raviv, A. Meyer, S. Lev-Ari, Compositional structure can emerge without generational transmission. *Cognition* **182**, 151–164 (2019). doi: 10.1016/j.cognition.2018.09.010; pmid: 30267952
17. S. Kirby, H. Cornish, K. Smith, Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10681–10686 (2008). doi: 10.1073/pnas.0707835105; pmid: 18667697
18. Y. Motamedi, M. Schouwstra, K. Smith, J. Culbertson, S. Kirby, Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition* **192**, 103964 (2019). doi: 10.1016/j.cognition.2019.05.001; pmid: 31302362
19. S. Kirby, M. Tamariz, Cumulative cultural evolution, population structure and the origin of combinatoriality in human language. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **377**, 20200319 (2022). doi: 10.1098/rstb.2020.0319; pmid: 34894728
20. L. Raviv, A. Meyer, S. Lev-Ari, Larger communities create more systematic languages. *Proc. Biol. Sci.* **286**, 20191262 (2019). doi: 10.1098/rspb.2019.1262; pmid: 31311478
21. K. Smith, How Culture and Biology Interact to Shape Language and the Language Faculty, How Culture and Biology Interact to Shape Language and the Language Faculty. *Top. Cogn. Sci.* **12**, 690–712 (2020). doi: 10.1111/tops.12377; pmid: 30182526
22. P. Hagoort, A. Özyürek, Extending the Architecture of Language From a Multimodal Perspective. *Top. Cogn. Sci.* **17**, 877–887 (2025). doi: 10.1111/tops.12728; pmid: 38493475

Downloaded from https://www.science.org at Radboud University Nijmegen on February 05, 2026

23. P. Hagoort, The core and beyond in the language-ready brain. *Neurosci. Biobehav. Rev.* **81** (Pt B), 194–204 (2017). doi: 10.1016/j.neubiorev.2017.01.048; pmid: 28193452
24. S. C. Vernes, V. M. Janik, W. T. Fitch, P. J. B. Slater, Vocal learning in animals and humans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **376**, 20200234 (2021). doi: 10.1098/rstb.2020.0234; pmid: 34482718
25. W. T. Fitch, "The Biology and Evolution of Language: 'Deep Homology' and the Evolution of Innovation" in *Cogn. Neurosci. Michael Gazzaniga* (The MIT Press, 2009), pp. 873–883. doi: 10.7551/mitpress/8029.003.0075
26. C. Scharff, J. Petri, Evo-devo, deep homology and FoxP2: Implications for the evolution of speech and language. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2124–2140 (2011). doi: 10.1098/rstb.2011.0001; pmid: 21690130
27. N. Shubin, C. Tabin, S. Carroll, Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818–823 (2009). doi: 10.1038/nature07891; pmid: 19212399
28. J. den Hoed, S. E. Fisher, Genetic pathways involved in human speech disorders. *Curr. Opin. Genet. Dev.* **65**, 103–111 (2020). doi: 10.1016/j.gde.2020.05.012; pmid: 32622339
29. C. S. Lai, S. E. Fisher, J. A. Hurst, F. Vargha-Khadem, A. P. Monaco, A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519–523 (2001). doi: 10.1038/35097076; pmid: 11586359
30. L. D. Morison *et al.*, In-depth characterisation of a cohort of individuals with missense and loss-of-function variants disrupting *FOXP2*. *J. Med. Genet.* **60**, 597–607 (2023). doi: 10.1136/jmg-2022-108734; pmid: 36328423
31. S. E. Fisher, C. Scharff, FOXP2 as a molecular window into speech and language. *Trends Genet.* **25**, 166–177 (2009). doi: 10.1016/j.tig.2009.03.002; pmid: 19304338
32. S. E. Fisher, Evolution of language: Lessons from the genome. *Psychon. Bull. Rev.* **24**, 34–40 (2017). doi: 10.3758/s13423-016-1112-8; pmid: 27432000
33. M. Groszer *et al.*, Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Curr. Biol.* **18**, 354–362 (2008). doi: 10.1016/j.cub.2008.01.060; pmid: 18328704
34. C. A. French *et al.*, An aetiological Foxp2 mutation causes aberrant striatal activity and alters plasticity during skill learning. *Mol. Psychiatry* **17**, 1077–1085 (2012). doi: 10.1038/mp.2011.105; pmid: 21876543
35. H.-Y. Kuo *et al.*, Speech- and language-linked FOXP2 mutation targets protein motors in striatal neurons. *Brain* **146**, 3542–3557 (2023). doi: 10.1093/brain/awad090; pmid: 37137515
36. M. Druart, M. Groszer, C. Le Magueresse, An Etiological Foxp2 Mutation Impairs Neuronal Gain in Layer VI Cortico-Thalamic Cells through Increased $\mathrm{GABA_B}$/GIRK Signaling. *J. Neurosci.* **40**, 8543–8555 (2020). doi: 10.1523/JNEUROSCI.2615-19.2020; pmid: 33020214
37. P. Norton, P. Barschke, C. Scharff, E. Mendoza, Differential Song Deficits after Lentivirus-Mediated Knockdown of FoxP1, FoxP2, or FoxP4 in Area X of Juvenile Zebra Finches. *J. Neurosci.* **39**, 9782–9796 (2019). doi: 10.1523/JNEUROSCI.1250-19.2019; pmid: 31641053
38. L. Xiao *et al.*, Expression of FoxP2 in the basal ganglia regulates vocal motor sequences in the adult songbird. *Nat. Commun.* **12**, 2617 (2021). doi: 10.1038/s41467-021-22918-2; pmid: 33976169
39. V. M. Janik, M. Knörnschild, Vocal production learning in mammals revisited. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **376**, 20200244 (2021). doi: 10.1098/rstb.2020.0244; pmid: 34482736
40. M. E. Wirthlin *et al.*, Vocal learning-associated convergent evolution in mammalian proteins and regulatory elements. *Science* **383**, eabn3263 (2024). doi: 10.1126/science.abn3263; pmid: 38422184
41. V. Slon *et al.*, The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature* **561**, 113–116 (2018). doi: 10.1038/s41586-018-0455-x; pmid: 30135579
42. T. Higham *et al.*, The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* **512**, 306–309 (2014). doi: 10.1038/nature13621; pmid: 25143113
43. M. Kuhlwilm, C. Boeckx, A catalog of single nucleotide changes distinguishing modern humans from archaic hominins. *Sci. Rep.* **9**, 8463 (2019). doi: 10.1038/s41598-019-44877-x; pmid: 31186485
44. Z. N. Kronenberg *et al.*, High-resolution comparative analysis of great ape genomes. *Science* **360**, eaar6343 (2018). doi: 10.1126/science.aar6343; pmid: 29880660
45. J. Krause *et al.*, The derived FOXP2 variant of modern humans was shared with Neandertals. *Curr. Biol.* **17**, 1908–1912 (2007). doi: 10.1016/j.cub.2007.10.008; pmid: 17949978
46. W. Enard, The Molecular Basis of Human Brain Evolution. *Curr. Biol.* **26**, R1109–R1117 (2016). doi: 10.1016/j.cub.2016.09.030; pmid: 27780052
47. K. Hammerschmidt *et al.*, A humanized version of Foxp2 does not affect ultrasonic vocalization in adult mice. *Genes Brain Behav.* **14**, 583–590 (2015). doi: 10.1111/gbb.12237; pmid: 26250064
48. S. von Merten, C. Pfeifle, S. Künzel, S. Hoier, D. Tautz, A humanized version of Foxp2 affects ultrasonic vocalization in adult female and male mice. *Genes Brain Behav.* **20**, e12764 (2021). doi: 10.1111/gbb.12764; pmid: 34342113
49. C. Schreiweis *et al.*, Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14253–14258 (2014). doi: 10.1073/pnas.1414542111; pmid: 25225386
50. S. E. Fisher, Human Genetics: The Evolving Story of FOXP2. *Curr. Biol.* **29**, R65–R67 (2019). doi: 10.1016/j.cub.2018.11.047; pmid: 30668952
51. S. B. Carroll, Evolution at two levels: On genes and form. *PLOS Biol.* **3**, e245 (2005). doi: 10.1371/journal.pbio.0030245; pmid: 16000021
52. S. Ma *et al.*, Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science* **377**, eabo7257 (2022). doi: 10.1126/science.abo7257; pmid: 36007006
53. D. Gokhman *et al.*, Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science* **344**, 523–527 (2014). doi: 10.1126/science.1250368; pmid: 24786081
54. D. Gokhman *et al.*, Differential DNA methylation of vocal and facial anatomy genes in modern humans. *Nat. Commun.* **11**, 1189 (2020). doi: 10.1038/s41467-020-15020-6; pmid: 32132541
55. P. Marler, Birdsong and speech development: Could there be parallels? *Am. Sci.* **58**, 669–673 (1970). pmid: 5480089
56. I. Pepperberg, K. J. Brese, B. J. Harris, Solitary sound play during acquisition of English vocalizations by an African Grey parrot (Psittacus erithacus): Possible parallels with children's monologue speech. *Appl. Psycholinguist.* **12**, 151–178 (1991). doi: 10.1017/S0142716400009127
57. A. A. Fernandez, L. S. Burchardt, M. Nagy, M. Knörnschild, Babbling in a vocal learning bat resembles human infant babbling. *Science* **373**, 923–926 (2021). doi: 10.1126/science.abf9279; pmid: 34413237
58. L. A. Petitto, S. Holowka, L. E. Sergio, B. Levy, D. J. Ostry, Baby hands that move to the rhythm of language: Hearing babies acquiring sign languages babble silently on the hands. *Cognition* **93**, 43–73 (2004). doi: 10.1016/j.cognition.2003.10.007; pmid: 15110725
59. D. K. Oller, R. E. Eilers, D. H. Bull, A. E. Carney, Prespeech vocalizations of a deaf infant: A comparison with normal metaphonological development. *J. Speech Hear. Res.* **28**, 47–63 (1985). doi: 10.1044/jshr.2801.47; pmid: 3981997
60. P. Adret, Operant conditioning, song learning and imprinting to taped song in the zebra finch. *Anim. Behav.* **46**, 149–159 (1993). doi: 10.1006/anbe.1993.1170
61. L. V. Riters, J. A. Spool, D. P. Merullo, A. H. Hahn, Song practice as a rewarding form of play in songbirds. *Behav. Processes* **163**, 91–98 (2019). doi: 10.1016/j.beproc.2017.10.002; pmid: 29031813
62. Y. Kim, S. Kwon, R. Rajan, C. Mori, S. Kojima, Intrinsic motivation for singing in songbirds is enhanced by temporary singing suppression and regulated by dopamine. *Sci. Rep.* **11**, 20350 (2021). doi: 10.1038/s41598-021-99456-w; pmid: 34645903
63. E. Hisey, M. G. Kearney, R. Mooney, A common neural circuit mechanism for internally guided and externally reinforced forms of motor learning. *Nat. Neurosci.* **21**, 589–597 (2018). doi: 10.1038/s41593-018-0092-6; pmid: 29483664
64. M. Tanaka, F. Sun, Y. Li, R. Mooney, A mesocortical dopamine circuit enables the cultural transmission of vocal behaviour. *Nature* **563**, 117–120 (2018). doi: 10.1038/s41586-018-0636-7; pmid: 30333629
65. E. Fedorenko, A. A. Ivanova, T. I. Regev, The language network as a natural kind within the broader landscape of the human brain. *Nat. Rev. Neurosci.* **25**, 289–312 (2024). doi: 10.1038/s41583-024-00802-4; pmid: 38609551
66. Y. Becker *et al.*, Long arcuate fascicle in wild and captive chimpanzees as a potential structural precursor of the language network. *Nat. Commun.* **16**, 4485 (2025). doi: 10.1038/s41467-025-59254-8; pmid: 40374618
67. J. Sierpowska *et al.*, Comparing human and chimpanzee temporal lobe neuroanatomy reveals modifications to human language hubs beyond the frontotemporal arcuate fasciculus. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2118295119 (2022). doi: 10.1073/pnas.2118295119; pmid: 35787056
68. G. Gallardo *et al.*, Morphological evolution of language-relevant brain areas. *PLOS Biol.* **21**, e3002266 (2023). doi: 10.1371/journal.pbio.3002266; pmid: 37656748
69. S. Kirby, Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Trans. Evol. Comput.* **5**, 102–110 (2001). doi: 10.1109/4235.918430
70. C. Kemp, Y. Xu, T. Regier, Semantic Typology and Efficient Communication. *Annu. Rev. Linguist.* **4**, 109–128 (2018). doi: 10.1146/annurev-linguistics-011817-045406
71. E. Gibson *et al.*, How Efficiency Shapes Human Language. *Trends Cogn. Sci.* **23**, 389–407 (2019). doi: 10.1016/j.tics.2019.02.003; pmid: 31006626
72. B. de Boer, W. Sandler, S. Kirby, New perspectives on duality of patterning: Introduction to the special issue. *Lang. Cogn.* **4**, 251–259 (2012). doi: 10.1515/langcog-2012-0014; pmid: 24587833
73. H. Little, K. Eryılmaz, B. de Boer, Signal dimensionality and the emergence of combinatorial structure. *Cognition* **168**, 1–15 (2017). doi: 10.1016/j.cognition.2017.06.011; pmid: 28641078
74. W. Sandler, C. Padden, M. Aronoff, Emerging Sign Languages. *Languages (Basel)* **7**, 284 (2022). doi: 10.3390/languages7040284
75. T. C. Scott-Phillips, S. Kirby, Language evolution in the laboratory. *Trends Cogn. Sci.* **14**, 411–417 (2010). doi: 10.1016/j.tics.2010.06.006; pmid: 20675183
76. M. Tamariz, Experimental Studies on the Cultural Evolution of Language. *Annu. Rev. Linguist.* **3**, 389–407 (2017). doi: 10.1146/annurev-linguistics-011516-033807
77. S. Goldin-Meadow, C. Mylander, Spontaneous sign systems created by deaf children in two cultures. *Nature* **391**, 279–281 (1998). doi: 10.1038/34646; pmid: 9440690
78. A. Senghas, S. Kita, A. Ozyürek, Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science* **305**, 1779–1782 (2004). doi: 10.1126/science.1100199; pmid: 15375269
79. S. Goldin-Meadow, D. Brentari, M. Coppola, L. Horton, A. Senghas, Watching language grow in the manual modality: Nominals, predicates, and handshapes. *Cognition* **136**, 381–395 (2015). doi: 10.1016/j.cognition.2014.11.029; pmid: 25546342

Downloaded from https://www.science.org at Radboud University Nijmegen on February 05, 2026

80. D. Hunsicker, S. Goldin-Meadow, Hierarchical structure in a self-created communication system: Building nominal constituents in homesign. *Language* **88**, 732–763 (2012). doi: 10.1353/lan.2012.0092; pmid: 23626381
81. A. Senghas, A. Özyürek, S. Goldin-Meadow, "Homesign as a way-station between co-speech gesture and sign language: The evolution of segmenting and sequencing" in *Evol. Emergence Lang. Evid. Inference*, R. Botha, M. Everaert, Eds. (OUP Oxford, 2013), pp. 62–76. doi: 10.1093/acprof:oso/9780199654840.003.0004
82. S. Goldin-Meadow, W. C. So, A. Ozyürek, C. Mylander, The natural order of events: How speakers of different languages represent events nonverbally. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9163–9168 (2008). doi: 10.1073/pnas.0710060105; pmid: 18599445
83. Ş. Özçalışkan, C. Lucero, S. Goldin-Meadow, Does language shape silent gesture? *Cognition* **148**, 10–18 (2016). doi: 10.1016/j.cognition.2015.12.001; pmid: 26707427
84. Y. Motamedi, M. Schouwstra, S. Kirby, An evolutionary approach to sign language emergence: From *state* to *process*. *Behav. Brain Sci.* **40**, e65 (2017). doi: 10.1017/S0140525X15003003; pmid: 29342521
85. S. Kirby, Y. Motamedi, L. Wolters, K. Smith, Compositionality Arises From Iterated Learning Despite a Preference for Holistic Signals: An Experimental Model of Sign Language Emergence. *Proceedings of the Joint Conference on Language Evolution (JCoLE)* (2022), pp. 402–404.
86. J. Nölle, M. Staib, R. Fusaroli, K. Tylén, The emergence of systematicity: How environmental and communicative factors shape a novel communication system. *Cognition* **181**, 93–104 (2018). doi: 10.1016/j.cognition.2018.08.014; pmid: 30173106
87. R. Selten, M. Warglien, The emergence of simple languages in an experimental coordination game. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7361–7366 (2007). doi: 10.1073/pnas.0702077104; pmid: 17449635
88. S. Kirby, M. Tamariz, H. Cornish, K. Smith, Compression and communication in the cultural evolution of linguistic structure. *Cognition* **141**, 87–102 (2015). doi: 10.1016/j.cognition.2015.03.016; pmid: 25966840
89. A. Wray, Protolanguage as a holistic system for social interaction. *Lang. Commun.* **18**, 47–67 (1998). doi: 10.1016/S0271-5309(97)00033-5
90. S. Kirby, "Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners" in *The Evolutionary Emergence of Language* (Cambridge Univ. Press, 2000), pp. 303–323. doi: 10.1017/CBO9780511606441.019
91. I. Arnon, The Starting Big approach to language learning. *J. Child Lang.* **48**, 937–958 (2021). doi: 10.1017/S0305000921000386; pmid: 34219627
92. S. Goldin-Meadow, I. Arnon, Whole-to-part development in language creation. *Trends Cogn. Sci.* **29**, 12–14 (2025). doi: 10.1016/j.tics.2024.09.015; pmid: 39472231
93. B. Skarabela, M. Ota, R. O'Connor, I. Arnon, 'Clap your hands' or 'take your hands'? One-year-olds distinguish between frequent and infrequent multiword phrases. *Cognition* **211**, 104612 (2021). doi: 10.1016/j.cognition.2021.104612; pmid: 33578095
94. I. Arnon, M. Ramscar, Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition* **122**, 292–305 (2012). doi: 10.1016/j.cognition.2011.10.009; pmid: 22169657
95. N. Siegelman, I. Arnon, The advantage of starting big: Learning from unsegmented input facilitates mastery of grammatical gender in an artificial language. *J. Mem. Lang.* **85**, 60–75 (2015). doi: 10.1016/j.jml.2015.07.003
96. S. Kirby, T. Griffiths, K. Smith, Iterated learning and the evolution of language. *Curr. Opin. Neurobiol.* **28**, 108–114 (2014). doi: 10.1016/j.conb.2014.07.014; pmid: 25062470
97. A. Cangelosi, D. Parisi, *Simulating the Evolution of Language* (Springer Science & Business Media, 2012).
98. W. T. Fitch, Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Phys. Life Rev.* **11**, 329–364 (2014). doi: 10.1016/j.plrev.2014.04.005; pmid: 24969660
99. J. Hyland Bruno, E. D. Jarvis, M. Liberman, O. Tchernichovski, Birdsong Learning and Culture: Analogies with Human Spoken Language. *Annu. Rev. Linguist.* **7**, 449–472 (2021). doi: 10.1146/annurev-linguistics-090420-121034
100. H. Williams, R. F. Lachlan, Evidence for cumulative cultural evolution in bird song. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **377**, 20200322 (2022). doi: 10.1098/rstb.2020.0322; pmid: 34894731
101. W. H. Thorpe, *Bird-song: the biology of vocal communication and expression in birds* (Cambridge Univ. Press, 1961).
102. P. Marler, "The instinct to learn" in *Epigenesis Mind Essays Biol. Cogn* (Psychology Press, 2014), pp. 37–66.
103. O. Fehér, H. Wang, S. Saar, P. P. Mitra, O. Tchernichovski, De novo establishment of wild-type song culture in the zebra finch. *Nature* **459**, 564–568 (2009). doi: 10.1038/nature07994; pmid: 19412161
104. O. Fehér, I. Ljubičić, K. Suzuki, K. Okanoya, O. Tchernichovski, Statistical learning in songbirds: From self-tutoring to song culture. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**, 20160053 (2017). doi: 10.1098/rstb.2016.0053; pmid: 27872371
105. C. I. Petkov, C. Ten Cate, Structured Sequence Learning: Animal Abilities, Cognitive Operations, and Language Evolution. *Top. Cogn. Sci.* **12**, 828–842 (2020). doi: 10.1111/tops.12444; pmid: 31359600
106. K. Zuberbühler, Combinatorial capacities in primates. *Curr. Opin. Behav. Sci.* **21**, 161–169 (2018). doi: 10.1016/j.cobeha.2018.03.015
107. C. Girard-Buttoz *et al.*, Versatile use of chimpanzee call combinations promotes meaning expansion. *Sci. Adv.* **11**, eadq2879 (2025). doi: 10.1126/sciadv.adq2879; pmid: 40344055
108. Q. Gallot, C. Depriester, S. Moran, K. Zuberbühler, A nonhuman primate combinatorial system for long-distance communication. *iScience* **27**, 111172 (2024). doi: 10.1016/j.isci.2024.111172; pmid: 39559758
109. N. Claidière, K. Smith, S. Kirby, J. Fagot, Cultural evolution of systematically structured behaviour in a nonhuman primate. *Proceedings of the Royal Society B: Biological Sciences* **281**, 2014.1541 (2014).
110. E. V. Clark, "Language is Acquired in Interaction" in *Algebraic Structures in Natural Language* (CRC Press, 2022). doi: 10.1201/9781003205388-4
111. M. Tomasello, The social bases of language acquisition. *Soc. Dev.* **1**, 67–87 (1992). doi: 10.1111/j.1467-9507.1992.tb00135.x
112. S. R. Lytle, P. K. Kuhl, "Social Interaction and Language Acquisition" in *The Handbook of Psycholinguistics* (John Wiley & Sons, Ltd, 2017), pp. 615–634, doi: 10.1002/9781118829516.ch27
113. J. R. Saffran, R. N. Aslin, E. L. Newport, Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928 (1996). doi: 10.1126/science.274.5294.1926; pmid: 8943209
114. M. Babineau, R. Shi, A. Christophe, 14-month-olds exploit verbs' syntactic contexts to build expectations about novel words. *Infancy* **25**, 719–733 (2020). doi: 10.1111/infa.12354; pmid: 32857439
115. M. H. Goldstein, A. P. King, M. J. West, Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8030–8035 (2003). doi: 10.1073/pnas.1332441100; pmid: 12808137
116. P. K. Kuhl, F.-M. Tsao, H.-M. Liu, Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9096–9101 (2003). doi: 10.1073/pnas.1532872100; pmid: 12861072
117. S. Goldin-Meadow, *Thinking with Your Hands: The Surprising Science Behind How Gestures Shape Our Thoughts* (Hachette, 2023).
118. E. V. Clark, Conversational Repair and the Acquisition of Language. *Discourse Process.* **57**, 441–459 (2020). doi: 10.1080/0163853X.2020.1719795
119. W. T. Fitch, L. Huber, T. Bugnyar, Social cognition and the evolution of language: Constructing cognitive phylogenies. *Neuron* **65**, 795–814 (2010). doi: 10.1016/j.neuron.2010.03.011; pmid: 20346756
120. S. C. Levinson, "Interactional Foundations of Language: The Interaction Engine Hypothesis" in *Human Language*, P. Hagoort, Ed. (The MIT Press, 2019), pp. 189–200. doi: 10.7551/mitpress/10841.003.0018
121. D. I. Tsomokos, L. Raviv, A bidirectional association between language development and prosocial behavior in childhood: Evidence from a longitudinal birth cohort in the United Kingdom. *Dev. Psychol.* (2024). doi: 10.1037/dev0001875; pmid: 39531699
122. R. Ensor, D. Spencer, C. Hughes, 'You Feel Sad?' Emotion Understanding Mediates Effects of Verbal Ability and Mother–Child Mutuality on Prosocial Behaviors: Findings from 2 Years to 4 Years. *Soc. Dev.* **20**, 93–110 (2011). doi: 10.1111/j.1467-9507.2009.00572.x
123. E. Longobardi, P. Spataro, C. Rossi-Arnaud, Direct and Indirect Associations of Empathy, Theory of Mind, and Language with Prosocial Behavior: Gender Differences in Primary School Children. *J. Genet. Psychol.* **180**, 266–279 (2019). doi: 10.1080/00221325.2019.1653817; pmid: 31456504
124. V. Ornaghi, A. Pepe, I. Grazzani, False-Belief Understanding and Language Ability Mediate the Relationship between Emotion Comprehension and Prosocial Orientation in Preschoolers. *Front. Psychol.* **7**, 1534 (2016). doi: 10.3389/fpsyg.2016.01534; pmid: 27774075
125. N. P. van den Bedem, J. E. Dockrell, P. M. van Alphen, S. V. Kalicharan, C. Rieffe, Victimization, Bullying, and Emotional Competence: Longitudinal Associations in (Pre)Adolescents With and Without Developmental Language Disorder. *J. Speech Lang. Hear. Res.* **61**, 2028–2044 (2018). doi: 10.1044/2018_JSLHR-L-17-0429; pmid: 29998317
126. C. L. Forrest, J. L. Gibson, S. L. Halligan, M. C. St Clair, A longitudinal analysis of early language difficulty and peer problems on later emotional difficulties in adolescence: Evidence from the Millennium Cohort Study. *Autism Dev. Lang. Impair.* **3**, 2396941518795392 (2018). doi: 10.1177/2396941518795392
127. M. J. West, A. P. King, Female visual displays affect the development of male song in the cowbird. *Nature* **334**, 244–246 (1988). doi: 10.1038/334244a0; pmid: 3398921
128. S. Carouso-Peck, O. Menyhart, T. J. DeVoogd, M. H. Goldstein, Contingent parental responses are naturally associated with zebra finch song learning. *Anim. Behav.* **165**, 123–132 (2020). doi: 10.1016/j.anbehav.2020.04.019
129. Y. Chen, L. E. Matheson, J. T. Sakata, Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6641–6646 (2016). doi: 10.1073/pnas.1522306113; pmid: 27247385
130. S. Yanagihara, Y. Yazaki-Sugiyama, Social interaction with a tutor modulates responsiveness of specific auditory neurons in juvenile zebra finches. *Behav. Processes* **163**, 32–36 (2019). doi: 10.1016/j.beproc.2018.04.003; pmid: 29656094
131. P. M. Greenfield, E. S. Savage-Rumbaugh, Comparing communicative competence in child and chimp: The pragmatics of repetition. *J. Child Lang.* **20**, 1–26 (1993). doi: 10.1017/S0305000900009090; pmid: 7681067
132. C. Yang, Ontogeny and phylogeny of language. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6324–6327 (2013). doi: 10.1073/pnas.1216803110; pmid: 23576720
133. M. Takahasi, K. Okanoya, Song Learning in Wild and Domesticated Strains of White-Rumped Munia, Lonchura striata, Compared by Cross-Fostering Procedures: Domestication Increases Song Variability by Decreasing Strain-Specific Bias. *Ethology* **116**, 396–405 (2010). doi: 10.1111/j.1439-0310.2010.01761.x

Downloaded from https://www.science.org at Radboud University Nijmegen on February 05, 2026

134. K. Suzuki, K. Okanoya, Domestication effects on aggressiveness: Comparison of biting motivation and bite force between wild and domesticated finches. *Behav. Processes* **193**, 104503 (2021). doi: 10.1016/j.beproc.2021.104503; pmid: 34537267
135. K. Suzuki, H. Yamada, T. Kobayashi, K. Okanoya, Decreased fecal corticosterone levels due to domestication: A comparison between the white-backed Munia (Lonchura striata) and its domesticated strain, the Bengalese finch (Lonchura striata var. domestica) with a suggestion for complex song evolution. *J. Exp. Zool. Part A. Ecol. Genet. Physiol.* **317**, 561–570 (2012). doi: 10.1002/jez.1748; pmid: 22927235
136. K. Suzuki, M. Ikebuchi, H. Kagawa, T. Koike, K. Okanoya, Effects of domestication on neophobia: A comparison between the domesticated Bengalese finch (Lonchura striata var. domestica) and its wild ancestor, the white-rumped munia (Lonchura striata). *Behav. Processes* **193**, 104502 (2021). doi: 10.1016/j.beproc.2021.104502; pmid: 34530107
137. E. Honda, K. Okanoya, Acoustical and Syntactical Comparisons between Songs of the White-backed Munia (Lonchura striata) and Its Domesticated Strain, the Bengalese Finch (Lonchura striata var. domestica). *Zool. Sci.* **16**, 319–326 (1999). doi: 10.2108/zsj.16.319
138. Y. Tobari *et al.*, Oxytocin variation and brain region-specific gene expression in a domesticated avian species. *Genes Brain Behav.* **21**, e12780 (2022). doi: 10.1111/gbb.12780; pmid: 34854547
139. T. O'Rourke *et al.*, Capturing the Effects of Domestication on Vocal Learning Complexity. *Trends Cogn. Sci.* **25**, 462–474 (2021). doi: 10.1016/j.tics.2021.03.007; pmid: 33810982
140. K. M. Kendrick, The neurobiology of social bonds. *J. Neuroendocrinol.* **16**, 1007–1008 (2004). doi: 10.1111/j.1365-2826.2004.01262.x; pmid: 15667456
141. T. M. Love, Oxytocin, motivation and the role of dopamine. *Pharmacol. Biochem. Behav.* **119**, 49–60 (2014). doi: 10.1016/j.pbb.2013.06.011; pmid: 23850525
142. B. Hare, Survival of the Friendliest: Homo sapiens Evolved via Selection for Prosociality. *Annu. Rev. Psychol.* **68**, 155–186 (2017). doi: 10.1146/annurev-psych-010416-044201; pmid: 27732802
143. L. Progovac, A. Benítez-Burraco, From Physical Aggression to Verbal Behavior: Language Evolution and Self-Domestication Feedback Loop. *Front. Psychol.* **10**, 2807 (2019). doi: 10.3389/fpsyg.2019.02807; pmid: 31920850
144. J. Thomas, S. Kirby, Self domestication and the evolution of language. *Biol. Philos.* **33**, 9 (2018). doi: 10.1007/s10539-018-9612-8; pmid: 29606782
145. T. M. Freeberg, R. I. M. Dunbar, T. J. Ord, Social complexity as a proximate and ultimate factor in communicative complexity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 1785–1801 (2012). doi: 10.1098/rstb.2011.0213; pmid: 22641818
146. L. Raviv, L. R. Peckre, C. Boeckx, What is simple is actually quite complex: A critical note on terminology in the domain of language and communication. *J. Comp. Psychol.* **136**, 215–220 (2022). doi: 10.1037/com0000328; pmid: 36222620
147. A. Formaux, D. Sperber, J. Fagot, N. Claidière, Guinea baboons are strategic cooperators. *Sci. Adv.* **9**, eadi5282 (2023). doi: 10.1126/sciadv.adi5282; pmid: 37889969
148. M. C. Corballis, From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behav. Brain Sci.* **26**, 199–208, discussion 208–260 (2003). doi: 10.1017/S0140525X03000062; pmid: 14621511
149. M. A. Arbib, From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behav. Brain Sci.* **28**, 105–124, discussion 125–167 (2005). doi: 10.1017/S0140525X05000038; pmid: 16201457
150. D. McNeill, *How Language Began: Gesture and Speech in Human Evolution* (Cambridge University Press, 2012). doi: 10.1017/CBO9781139108669
151. P. F. MacNeilage, B. L. Davis, The Frame/Content theory of evolution of speech: A comparison with a gestural-origins alternative. *Interact. Stud.* **6**, 173–199 (2005). doi: 10.1075/is.6.2.03mac
152. M. Fröhlich, C. Sievers, S. W. Townsend, T. Gruber, C. P. van Schaik, Multimodal communication and language origins: integrating gestures and vocalizations. *Biol. Rev. Camb. Philos. Soc.* **94**, 1809–1829 (2019). doi: 10.1111/brv.1253; pmid: 31250542
153. M. A. Arbib, K. Liebal, S. Pika, Primate vocalization, gesture, and the evolution of human language. *Curr. Anthropol.* **49**, 1053–1063, discussion 1063–1076 (2008). doi: 10.1086/593015; pmid: 19391445
154. P. J. Richerson, R. Boyd, A dual inheritance model of the human evolutionary process I: Basic postulates and a simple model. *J. Soc. Biol. Struct.* **1**, 127–154 (1978). doi: 10.1016/S0140-1750(78)80002-5
155. S. Kirby, M. Dowman, T. L. Griffiths, Innateness and culture in the evolution of language. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5241–5245 (2007). doi: 10.1073/pnas.0608222104; pmid: 17360393
156. B. Thompson, S. Kirby, K. Smith, Culture shapes the evolution of cognition. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4530–4535 (2016). doi: 10.1073/pnas.1523631113; pmid: 27044094
157. G. E. Hinton, S. J. Nowlan, How Learning Can Guide Evolution. *Complex Syst.* **1**, 495–502 (1987).
158. E. Ochs, "Transcription as theory" in *Dev. Pragmat*, E. Ochs, B. Schieffelin, Eds. (Academic Press, 1979), pp. 43–72.
159. A. M. Peters, *The Units of Language Acquisition* (Cambridge Univ. Press, 1983).
160. S. Goldin-Meadow, *The resilience of language: What gesture creation in deaf children can tell us about how all children learn language* (Psychology Press, 2003).
161. S. Goldin-Meadow, C. Mylander, C. Butcher, The resilience of combinatorial structure at the word level: Morphology in self-styled gesture systems. *Cognition* **56**, 195–262 (1995). doi: 10.1016/0010-0277(95)00662-I; pmid: 7554795
162. C. Grund, G. Badihi, K. E. Graham, A. Safryghin, C. Hobaiter, GesturalOrigins: A bottom-up framework for establishing systematic gesture data across ape species. *Behav. Res. Methods* **56**, 986–1001 (2024). doi: 10.3758/s13428-023-02082-9; pmid: 36922450
163. Y. Prat, Animals Have No Language, and Humans Are Animals Too. *Perspect. Psychol. Sci.* **14**, 885–893 (2019). doi: 10.1177/1745691619858402; pmid: 31398084
164. D. L. Cheney, R. M. Seyfarth, *Baboon Metaphysics: The Evolution of a Social Mind* (Univ. of Chicago Press, 2008). doi: 10.7208/9780226102429
165. I. Arnon *et al.*, Whale song shows language-like statistical structure. *Science* **387**, 649–653 (2025). doi: 10.1126/science.adq7055; pmid: 39913578
166. Y. Yovel, O. Rechavi, AI and the Doctor Dolittle challenge. *Curr. Biol.* **33**, R783–R787 (2023). doi: 10.1016/j.cub.2023.06.063; pmid: 37552937
167. C. Rutz *et al.*, Using machine learning to decode animal communication. *Science* **381**, 152–155 (2023). doi: 10.1126/science.adg7314; pmid: 37440653
168. C. Hobaiter, K. E. Graham, R. W. Byrne, Are ape gestures like words? Outstanding issues in detecting similarities and differences between human language and ape gesture. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **377**, 20210301 (2022). doi: 10.1098/rstb.2021.0301; pmid: 35934962
169. S. Ferrigno, S. J. Cheyette, S. T. Piantadosi, J. F. Cantlon, Recursive sequence generation in monkeys, children, U.S. adults, and native Amazonians. *Sci. Adv.* **6**, eaaz1002 (2020). doi: 10.1126/sciadv.aaz1002; pmid: 32637593
170. X. Jiang *et al.*, Production of Supra-regular Spatial Sequences by Macaque Monkeys. *Curr. Biol.* **28**, 1851–1859.e4 (2018). doi: 10.1016/j.cub.2018.04.047; pmid: 29887304
171. T. J. Bergman, J. C. Beehner, D. L. Cheney, R. M. Seyfarth, Hierarchical classification by rank and kinship in baboons. *Science* **302**, 1234–1236 (2003). doi: 10.1126/science.1087513; pmid: 14615544
172. O. Mascaro *et al.*, Human and animal dominance hierarchies show a pyramidal structure guiding adult and infant social inferences. *Nat. Hum. Behav.* **7**, 1294–1306 (2023). doi: 10.1038/s41562-023-01634-5; pmid: 37386104
173. A. D. Patel, Language, music, syntax and the brain. *Nat. Neurosci.* **6**, 674–681 (2003). doi: 10.1038/nn1082; pmid: 12830158
174. E. C. Garland, P. K. McGregor, Cultural Transmission, Evolution, and Revolution in Vocal Displays: Insights From Bird and Whale Song. *Front. Psychol.* **11**, 544929 (2020). doi: 10.3389/fpsyg.2020.544929; pmid: 33132953
175. Mde. J. Martins, Z. Muršič, J. Oh, W. T. Fitch, Representing visual recursion does not require verbal or motor resources. *Cogn. Psychol.* **77**, 20–41 (2015). doi: 10.1016/j.cogpsych.2015.01.004; pmid: 25743443
176. G. Westphal-Fitch, B. Giustolisi, C. Cecchetto, J. S. Martin, W. T. Fitch, Artificial Grammar Learning Capabilities in an Abstract Visual Task Match Requirements for Linguistic Syntax. *Front. Psychol.* **9**, 1210 (2018). doi: 10.3389/fpsyg.2018.01210; pmid: 30087630
177. E. G. Atkinson *et al.*, No Evidence for Recent Selection at FOXP2 among Diverse Human Populations. *Cell* **174**, 1424–1435.e15 (2018). doi: 10.1016/j.cell.2018.06.048; pmid: 30078708
178. E. Eising *et al.*, A set of regulatory genes co-expressed in embryonic human brain is implicated in disrupted speech development. *Mol. Psychiatry* **24**, 1065–1078 (2019). doi: 10.1038/s41380-018-0020-x; pmid: 29463886
179. E. Eising *et al.*, Genome-wide analyses of individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2202764119 (2022). doi: 10.1073/pnas.2202764119; pmid: 35998220
180. G. Alagöz *et al.*, Using neuroimaging genomics to investigate the evolution of human brain structure. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2200638119 (2022). doi: 10.1073/pnas.2200638119; pmid: 36161899
181. J. Uddén *et al.*, Toward Robust Functional Neuroimaging Genetics of Cognition. *J. Neurosci.* **39**, 8778–8787 (2019). doi: 10.1523/JNEUROSCI.0888-19.2019; pmid: 31570534
182. J. S. Amelink *et al.*, Imaging genetics of language network functional connectivity reveals links with language-related abilities, dyslexia and handedness. *Commun. Biol.* **7**, 1209 (2024). doi: 10.1038/s42003-024-06890-3; pmid: 39342056

ACKNOWLEDGMENTS

We thank the Israel Institute for Advanced Studies for providing the funds that allowed us to convene this research group and for creating a truly unique space for deep and intense intellectual exploration. We thank the participants of the 2023 workshop on "The Origins of human language: a multidisciplinary perspective" for stimulating and insightful discussions. (https://iias.huji.ac.il/event/origins-language-multidisciplinary-perspective-research-group-conference). All remaining mistakes are our own. **Funding:** Israel Institute for Advanced Studies research group fellowship (all authors); Israel Science Foundation grant 445/20 (I.A.); Institute for Language Communication and the Brain (ILCB) through the Agence Nationale de la Recherche grant ANR-16-CONV-0002 (N.C.); ANR TECHNICION grant ANR-21-CE28-0023-02 (N.C.); SCALUP grant ANR-21-CE28-0016-02 (N.C.); Japan Society for the Promotion of Sciences and Ministry of Education grant JP23H05428 (K.O.); Ministry of Education, Culture, Sports, Science and Technology grant #4903 (K.O.); Max Planck Society (L.R., S.E.F.); NIDCD grant R01 DC00491 (S.G.-M.). **Author contributions:** Conceptualization: All authors; Visualization: L.W., S.K., N.C., L.C., S.E.F.; Funding acquisition: I.A. and L.C. secured the funds for the research fellowship that led to the development of the paper; Project administration: I.A., S.E.F.; Writing – original draft: I.A. and S.E.F. created the first draft of the paper structure and led the introduction. Case studies were drafted by S.E.F. and W.T.F. (vocal production learning); S.K. and I.A. (emergence of structure); and L.R., K.O., and N.C. (prosociality). W.T.F. led the writing of the discussion. Figures were drafted by L.W. with S.K. (Fig. 1), S.E.F. (Fig. 2), S.G.-M. (Fig. 3), and W.T.F. (Fig. 4). The table was drafted by S.E.F. Writing – review & editing: Led by I.A. and S.E.F.; all authors contributed. **Competing interests:** The authors declare that they have no competing interests. **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 5 July 2024; resubmitted 12 February 2025; accepted 23 September 2025

10.1126/science.adq8303

Downloaded from https://www.science.org at Radboud University Nijmegen on February 05, 2026

What enables human language? A biocultural framework

Inbal Arnon, Liran Carmel, Nicolas Claidière, W. Tecumseh Fitch, Susan Goldin-Meadow, Simon Kirby, Kazuo Okanoya, Limor Raviv, Lucie Wolters, and Simon E. Fisher

Science **390** (6775), eadq8303. DOI: 10.1126/science.adq8303

Editor's summary

For nearly as long as humans have had an understanding of language, we have been interested in understanding its origins. Although human language is unique in many ways, recent research has made clear that language per se is not unique to humans. Arnon *et al*. describe a framework for understanding language evolution that incorporates both culture and biological preparedness. The authors then demonstrate the value of this framework in case studies of three aspects of language evolution. —Sacha Vignieri

View the article online
https://www.science.org/doi/10.1126/science.adq8303
Permissions
https://www.science.org/help/reprints-and-permissions

Downloaded from https://www.science.org at Radboud University Nijmegen on February 05, 2026

Use of this article is subject to the Terms of service

Science (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 © 2025 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works