From vocal homophily to vocal repertoire flexibility: Unravelling the socioecological drivers of language evolution

Sabine Stoll[a,*], Carel P. van Schaik[a,b,c,*], Judith M. Burkart[a,b], Konatsu Ono[b], Nikhil Phaniraj[b], Zitan Song[c], and Balthasar Bickel[a]

[a]Institute for the Interdisciplinary Study of Language Evolution, University of Zurich, Affolternstrasse 56, 8050 Zurich, Switzerland
[b]Institute for Evolutionary Anthropology, University of Zurich,Winterthurerstrasse 190, 8057 Zurich, Switzerland
[c]Department for the Ecology of Animal Societies, Max Planck Institute of Animal Behavior, Bücklestrasse 5, 78467 Konstanz, Germany

June 2025

Abstract

Since diverging from the last common ancestor with bonobos and chimpanzees, the communication system of the hominin lineage underwent a radical transformation. Vocal production learning - the ability to produce novel vocalizations based on experience - is the necessary precondition for changing an ape-like communication system into one that is infinitely flexible, extensively learned, and culturally transmitted. However, the evolutionary conditions favoring this shift remain unclear. Drawing on archeological evidence of intensified cooperation for social defense, hunting, and shared child-care among savanna-dwelling hominins, we suggest that the move to open savanna environments around two million years ago created pressures toward communicative capacities beyond the ancestral vocal repertoire. Specifically, we propose that vocal production learning arose from the increased use of preexisting vocal accommodation, where individuals align vocal features as a sign of homophily. We therefore predicted that the function of vocal homophily more closely aligns with the functions of human language than with those of vocal production learning in other species. A systematic comparison of the functions of vocal production learning and vocal accommodation across birds and mammals confirmed this prediction. It also revealed that vocal homophily often cooccurs with key features of human language, such as babbling, tutoring, turn-taking, and high vocal activity, especially in species with shared offspring care. We conclude that increased vocal homophily facilitated the evolution of vocal production learning in the hominin lineage and ultimately language.

Keywords: language evolution, vocal production learning, homophily, hominins, vocal accomodation, interdependence

Human language is among our most distinctive adaptations. Its evolution is a major puzzle because language is so vastly different from the communication system of our closest living relatives, chimpanzees and bonobos, with whom we share a common ancestor as recently as about 8 million years ago [1]. Great ape vocal communication relies on a limited set of innate calls, which are used relatively infrequently [2] and occur predominantly as single calls [3]. In humans, by contrast, very little of the communication system is innate, with thousands of languages spoken today, each continually evolving and changing [4].

The evolution of this system involved significant changes in vocal anatomy and brain regions responsible for vocal control since the divergence from other apes [5–8]. Major components of modern-day language originated in the period before 1.25 million years ago, as known from the genetics of learning mechanisms [7, 9–13] and fossil endocasts [14]. Current findings suggest that the transition evolved not in a single step but by combining different traits, each with its own evolutionary trajectory [5, 15–17]. Many of these traits were already present to a limited degree in the hominoid ancestors, such as some aspects of theory of mind [18], pragmatic inference [19], some amount of imitation and transmission skills [20–22], some degrees of compositionality [23, 24], the ability to process non-adjacent dependencies [25], and occasional communicative creativity [26–30]. What stands out, however, is the extent to which language must be learned by children [31]. This suggests the limiting factor was the absence of full-fledged vocal production learning in other hominoids, more than any other trait.

Current research on vocal production learning has greatly enriched our understanding of convergent genetic and neuroanatomical mechanisms. However, the conditions favoring its evolution, i.e. its social and communicative functions, have remained unclear. Here, we propose that the key first step toward vocal production learning was an increased need for vocal communication triggered by the enhanced need for cooperation due to the move onto the open savanna around 2 million years ago [32]. This increased interdependence led to greater vocal convergence among group members to signal social proximity and bonding, a form of vocal homophily. The underlying process, known as vocal accommodation [33, 34], involves the acoustic modification of existing vocalizations and results in vocal convergence. This intradyadic imitation gradually broadened into full-fledged vocal production learning. Intensified vocal homophily thus marked the transition from an innate communication system into a fully dynamic communication system entirely reliant on learning.

Since vocal production learning is a key requirement for language as we know it but was absent in our ancestors, we need a comparative approach to better understand its evolution in our lineage. We first compare the uses of vocal production learning and vocal accommodation across species. Our hypothesis is supported to the extent that both mechanisms have predominantly a homophily function in humans but not in other species where it has been observed. We then probe correlates of vocal production learning in human language, as for instance, babbling. If the correlates are again predominantly associated with vocal homophily, this would strengthen our hypothesis that vocal production learning in humans evolved from vocal homophily. Finally, we place these developments in the paleanthropological context of the evolution of Homo that evidences an increase of cooperation 2 My ago.

The vocal learning spectrum and the origin of speech

Interest in vocal learning has greatly increased since Janik & Slater [35] broadly defined it in 1997 as the ability to modify the structure of vocalizations conditioned by the experience with those of other individuals. Comparative work indicates different degrees of production learning that suggest a continuum of abilities ranging from vocal non-learners like chickens to high vocal learners like parrots and humans [36–40] or even a multi-dimensional spectrum [41, 42]. However, a basic distinction is generally made between vocal accommodation and the production and faithful imitation of entirely novel vocalizations, i.e., vocal production learning in the narrow sense (henceforth VPL). VPL is the necessary underlying mechanism for extending the innate repertoire. In the following, we subsume both forms under the umbrella term vocal learning in a wider sense.

Since VPL is essential for the mastery of human speech, studies of language evolution have mostly focused on species engaging in VPL [5, 36]. VPL has only been documented in a limited number of vertebrate species [42–44]: in birds, among at least three orders; in mammals, among bats, elephants, cetaceans, and pinnipeds. VPL evolved convergently in these lineages [43, 45], perhaps capitalizing on deep homologies in the neural development of motor learning [46, 47]. Vocal accommodation/alignment, in contrast, is reported in many species. However, as we show in what follows, this contrast seems exaggerated.

The distinction between VPL and vocal accommodation is traditionally made with the following criteria: (1) the absence of vocal development in young animals that are deaf or receive no vocal input; (2) the spread of new kinds of vocalizations in natural populations; or (3) the ability to imitate novel sounds outside the species' vocal repertoire. Criterion 1 is rarely used today and criterion 2 is in practice only relevant where new songs (rather than calls) spread in birds or cetaceans [48].

Closer examination of criterion 3 suggests that VPL may be more widespread than is currently known because this third criterion may produce many false negatives. For instance, immature bats and pinnipeds give isolation calls that are adjusted due to vocal input and thus would merely provide evidence for vocal accommodation. However, because they can also reproduce experimentally produced artificial calls [49, 50], they are also capable of VPL. The absence of evidence for such imitation skill leaves numerous potential candidate species as non-VPL. Likewise, various cases of VPL are based on mammals (elephants, cetaceans, pinnipeds) or parrots mimicking the voice or utterances of human caretakers, which are generally well outside the range of species-specific calls [38, 40, 42]. In fact, for some of the species in which VPL is documented, this is only known from captivity, even in well-studied Asian elephants or most parrots [51]. Excluding species for which only accommodation is documented may therefore unduly limit the scope of attempts to reconstruct the emergence of VPL in hominins and may thwart attempts to identify its early functions (rather than mechanisms).

There are additional arguments for including vocal accommodation. First, humans are primates, and the ancestral state in the hominin lineage almost certainly was non-VPL [52]. Yet, evidence for vocal accommodation is found scattered among primates [53]. Chimpanzees, for instance, show both short-term [54] and long-term accommodation in captivity [55] and the wild [56–58]. Great apes thus possess abilities that could have provided a stepping stone toward VPL and so released likely constraints on language evolution in our hominin ancestors. Second, developmental plasticity in vocal learning is actually documented in some primates [59–61]. Third, while volitional control over vocalizations is a precondition for VPL, species with vocal accommodation may also be capable of volitional control [62]: it is well-documented throughout life in common marmosets [63] and among juvenile rhesus monkeys [64]. Finally, a focus on great apes in particular is warranted by the presence of similarities in cytoarchitecture and within-cortex connections between the Broca's area involved in human speech production, and the homologous structures in non-human primate brains [65]. This suggests that learning mechanisms may be more widely shared than the ability to control sound production as in VPL, or even the need to use it in routine intraspecific communication.

We therefore compiled data on taxa with evidence for VPL or for vocal accommodation to

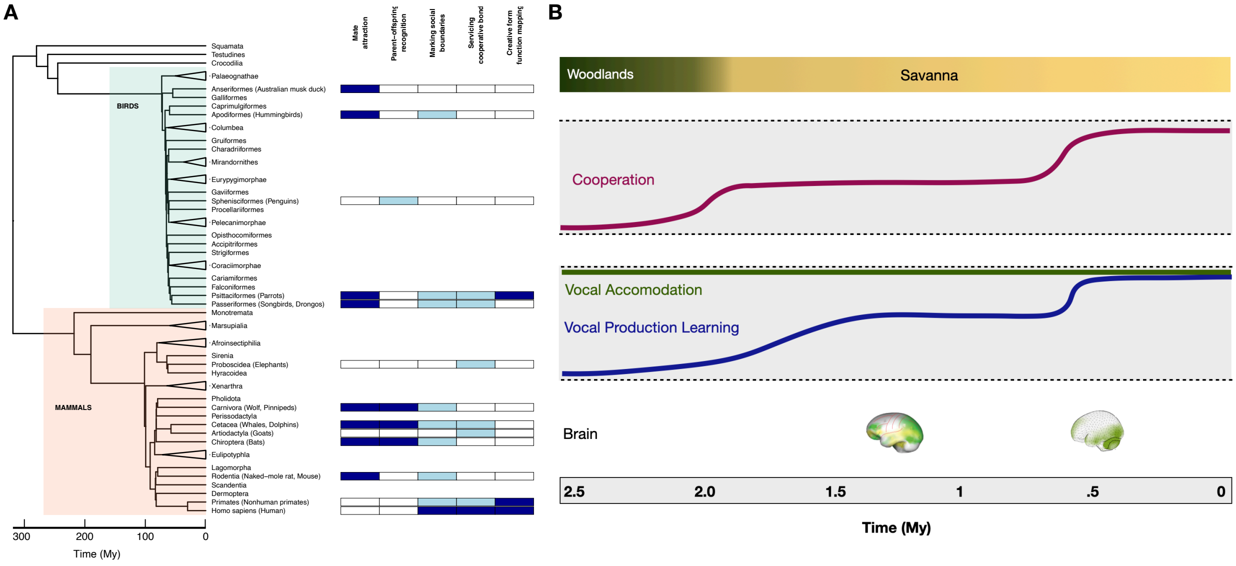

Figure 1: Distribution and hypothesized evolution of vocal learning. **A**: The known taxonomic distribution of two main kinds of vocal learning (vocal production learning, in dark blue, and vocal accommodation, in light blue) by function across amniotic vertebrates. Tip taxa are orders, but where vocal learning is recorded for isolated taxa within orders, these are mentioned after the order (for details, see the Supporting Information). Note that in primates, there is no evidence for the first two functions. **B:** Proposed evolution in hominins. The expansion into the savanna intensified cooperation, selecting for new brain structures [14]; learning skills improved dramatically again with new neurogenetic changes [67, 68] after the split from Neanderthals and other lineages.

identify the functional contexts in which VPL is found, as well as some of their correlates: features that reliably accompany vocal learning and are also found in human language learning.

Functions of vocal learning

VPL and vocal accommodation have a variety of functions [53, 66] (Figure 1A). Human language can serve all the functions listed in this table, but here the goal is to identify which of them was the primary context for the evolution of VPL in our lineage. We start with undisputed cases of VPL and then consider cases of vocal accommodation. We identified five distinct functions.

The first functional context of VPL requires sexual selection, as it serves to attract mates or defend territories. This function is widespread in birds, where one or both sexes produce songs that may be complex and variable in time and space [69]. The scattered cases outside passerines and parrots, namely in hummingbirds and Australian musk ducks [44], refer to lekking species, in which a male's songs and vocalizations also serve to attract females. The same context is common in whales [70] and harbor seals [71] . The learning process may involve a phase of learning through imperfect imitation (known as subsong in birds) and auditory feedback, sometimes accompanied by adult tutoring, or through direct imitation, as in mimics [72]. Learning is often possible only during a sensitive or even critical period [73].

Humans sing (often accompanied by dance and music), which is a cultural universal [74], and among its functions is mate attraction, but also strengthening of social cohesion. However, song lacks many of the informational features essential for language, such as semantic compositionality.

Second, language is not primarily used in attracting or competing over mates. Third, and most importantly, language is more plausibly derived from calls carrying information about the external world than from song. Indeed, the primary uses of language are to think, to inform or to instruct (for both technical and normative knowledge), coordinate the execution of joint action, and gossip (evaluate the reliability of cooperation partners). This is not to deny that the learning process shares many similarities with song learning, including a long phase of imperfect imitation and adult tutoring [75], but it marks a critical difference for the nature of the selective drivers.

A second context of VPL involves vocal recognition between parent and offspring, particularly in contexts where numerous dependent young are clustered together and must reunite with their parents after separation. This is most common in breeding colonies, especially where young are mobile, and therefore in sea birds, pinnipeds, or bats, where parents can be gone for a long time. In such cases, failure to reconnect is likely to lead to the death of the young. Thus, we may expect selection on modifying existing calls to enhance individually identifying elements of the call and so maximize mutual recognition. This pressure selects for vocal learning but also for creativity in vocal production in general to maximize the acoustic space. Many examples of pinnipeds and bats fit in this functional domain, but we hypothesize that the call exchanges of parents and chicks in many colony-breeding birds, such as penguins, would also show this.

The case of individualized signature whistles of dolphins may also fall into this category. In dolphins, mother-calf association is strong, but especially somewhat older calves increasingly undertake excursions, and because visual or gestalt recognition in water is difficult at larger distances, especially in murky water, relatives have come to rely on vocal recognition. By actively modifying the specific calls away from the modal pattern, pairs may enlarge the vocal space and thus avoid errors and improve mutual recognition.

Like mate attraction, this context is also unlikely to be the foundation for vocal learning in hominins. During early infancy, primate young are generally close to the mother and mutual recognition is generally guaranteed by a combination of vocal and olfactory cues, individual profiles in movement and gesture, and presumably above all visual cues, as documented in numerous experiments on individual recognition from photographs or videos [76]. Furthermore, by the time language is thought to have begun to emerge, hominins were no longer living in dense habitats, and children were highly unlikely to be far from other group members. Above all, true long-distance calls are facilitated by air sacs, which actually impair speech production [6], so simultaneous selection on long-distance calls and the kind of speech found in humans is physically impossible.

A third function is to mark social boundaries. This function is used to differentiate friend from foe when not in visual contact because pairs or groups periodically split up or spread out in habitats with low visibility. This social marking can happen at the level of dyads or groups. First, contact calls may make coordinated travel easier; and where kin or bonded partners are more likely to respond than others, vocal convergence may facilitate vocal recognition (as in the function of parent-offspring recognition) and thus be favored by selection. Second, selection on vocal convergence may be favored where fissioning groups are surrounded by hostile neighbors, which makes it important not to simply approach any conspecific contact call. We predict that these two conditions may explain the vocal convergence of group members in many primates, including chimpanzees, but also probably elephants and cetaceans, or in pair-living species that may forage apart like hornbills, large parrots, or marmosets. The predominant mechanism here is vocal accommodation, since the modification always occurs within the same call type, although in species capable of VPL these calls may be modified more drastically. However, to distinguish calls by other group members from those of other groups, vocal accommodation can produce not only vocal convergence within social units, but also vocal divergence between them, as suggested by differences between communities in chimpanzees [56].

Humans are extreme accommodators in exactly this context. Accommodation is used not only to create in-group similarity but also to emphasize differences with other groups. This is known as schismogenesis [77], which is amply documented [78–80] and considered a major driver of language change [4]. This context is therefore likely to have been crucial in the emergence of language.

Fourth, vocal accommodation serves to strengthen social bonds and enhance collaboration. This function has been identified in African penguins [81], chimpanzees, and various monkeys [53], most extensively in callitrichids, in which breeders have especially strong social bonds, and newly formed pairs vocally converge as their social bond gets stronger [82, 83]. Strong convergence is also found in cases where the recipient's identity is encoded in the contact call, as found in spectacled parrotlets [84], dolphins [85], African elephants [86] and marmosets [87]. These cases reflect situations in which an individual label for both the sender and the intended recipient is adaptive [51], because benefits to social contact are dyad-specific.

In both group-level and dyadic contexts, vocal convergence, probably due to vocal accommodation, is a form of vocal mimicry, similar to the more general mimicry of positional and gestural similarities in humans and other primates that serve the homophilic function of establishing trust and strengthening social bonds [88]. In the species involved, the two functions for vocal convergence (individual recognition and bond servicing) blend into each other, given that they both serve to establish or service social bonds or to mark membership of the same group. We call this phenomenon vocal homophily. These two contexts appear particularly relevant to language evolution.

A final functional context is creative form-function mapping to transmit information. While this is obviously ubiquitous in humans, the only possible evidence for it in animals is in parrots, in which males and females form stable, long-term pair bonds. Parrots of both sexes are known to produce novel, idiosyncratic calls for (non-sexual) social functions, such as recruitment of the long-term partner to defend nest sites and informing each other about toxic or armored foods [51]. However, so far, no detailed follow-up studies have ensued of this exciting parallel to the key function of human language.

In conclusion, the contexts most relevant to the emergence of human language are those in which vocal learning is mostly used for social recognition and bonding functions (Figure 1A). These are the core functions associated with increased group cooperation, consistent with our hypothesis.

Correlates of vocal learning

Language as we use it today is a package of multimodal traits that involve active assistance during learning. This manifests itself in such traits as babbling, volubility, tutoring, and turn taking. In what follows, we examine whether these traits are also found in nonhuman species, in contexts in which either accommodation or VPL is found. If they are found in species with accommodation, but not in species with VPL, this supports our hypothesis that accommodation in the hominin lineage was the ancestral state and turned into vocal production learning due to intensified needs of coordinated cooperation.

Babbling Human language production in early ontogeny starts out with early innate vocalizations turning into babbling. Initially, infant vocalizations start out with vowels followed by repetitive consonant-vowel combinations. Around 6-8 months of age, infants try out sounds that are potential candidates for phonemes (functionally distinctive sounds) in specific consonant-vowel combinations [89]. This coincides with anatomical changes in the vocal tract that enable speech-like sounds [90, 91]. During this phase, the sounds become increasingly similar to those of the native language [92].

Table 1: Taxa in which babbling has been recorded

Taxon	Rearing system	VPL
Confirmed cases		
Songbirds (subsong)	(mostly) biparental	yes
Green-rumped parrotlet	biparental	yes
Giant otter	cooperative	no
Greater sac-winged bat	mother-only	yes
Pale spear-nosed bat	allomaternal	yes
Gray mouse lemur	allomaternal	no
Common marmoset	cooperative	no
Pygmy marmoset	cooperative	no
Suggested cases		
Bottlenose dolphin	allomaternal	yes
Beluga	allomaternal	?
Naked mole rat	cooperative	no?
Degu	biparental	?
Norway rat	allomaternal	no

Babbling in immature animals is hard to define, given that each species' babbling may contain a somewhat different set of features [93, 94]. Nonetheless, its core is the use of rhythmic repetitions of adult vocalizations out of context in often long and seemingly random sequences [95]. Unfortunately, the taxonomic distribution of animal babbling is incompletely known. Among species with documented VPL, it is only known for humans, bats, and songbirds (where it is called subsong) and various other birds, but, perhaps surprisingly, not for any of the other species with VPL [94]. This may imply that the song-learning mechanism of VPL may be fundamentally different from that used for learning individually distinct contact or lost calls serving in vocal recognition.

Against this background, it is remarkable that babbling has also been described for a number of species not known to have VPL. These cases show a striking pattern (Table 1): except for one of the bat species, all recorded species share extensive allomaternal care (i.e. by non-mothers of dependent offspring, called cooperative breeding where care is provided by both male and female parents as well as others). This suggests that babbling may primarily function to elicit the attention of parents or caretakers [96, 97]. Given that mother-only care is the mode in mammals, this concentration among species with allomaternal care is striking, even if the lack of solid evidence on absences prevents formal statistical evaluation. Although the vocal-practice and care-elicitation functions are quite compatible, the presence in many mammals without variable songs suggests that the care-elicitation function may be ancestral, linked to attracting the attention of potentially selective caretakers.

We therefore propose that prelinguistic babbling in humans is functionally similar to what we find in prosocial species, especially cooperative breeders such as marmosets. Linguistic babbling, by contrast, is an active exercise to acquire the native language, like subsong in birds. This ontogenetic sequence is consistent with the hypothesis that babbling in hominins initially arose for the function of soliciting care and was later exapted for the additional function of language learning, without losing its original function, especially in its prelinguistic form.

Volubility There is large variation in how much species communicate, but this is poorly documented. Compared to humans, great apes are remarkably silent [2], whereas cooperative breeders are often described as highly voluble [96, 97], and in our experience far more voluble than any great ape. If future quantitative work confirms this, it points to the greater need for communication about coordinated cooperation in species with allomaternal care. Such greater communicative needs may also be expressed in the size of the vocal repertoire. Although interspecific comparisons are notoriously difficult, detailed comparisons find larger vocal repertoires in cooperative breeders among herpestids [98] and birds [99]. Comparisons suggest that in these cases the differences are largely related to warning and coordination, in other words cooperative (prosocial) activities. A similar trend may be observed in primates, as the four species with the largest vocal repertoires in the most recent compilation [100] include not only chimpanzees and bonobos, but also two species of cooperatively breeding callitrichids.

Immature-directed input All vocal learning requires vocal input. This can be either surrounding input that is not immediately directed at the learner, or specifically directed input, as in infant-directed communication in humans or tutoring in songbirds [69, 101]. In the great apes in contrast to humans, there is very little immature directed input [2]. In humans, infant-directed communication can be either interactive (contingent utterances to child vocalizations) or non-interactive. Recent research suggests that the interactive form is the best predictor of learning [102–104], presumably because it funnels the infant's attention. Unfortunately, in animals, this interactive tutoring has only been studied in a few species: marmosets, cowbirds and zebra finches [66, 105]. Perhaps surprisingly, in common marmosets, parents and caretakers even provide acoustic feedback to incorrect vocalizations of the immature [63, 106], facilitating the infants' vocal development [107, 108]. There is no information on other primate species, but since in most cases, tutoring and vocal feedback coincide, the lack of documenting tutoring implies that vocal feedback is also likely to be rare [2, 101, 109]. Thus, common marmosets or callitrichids more generally may stand out among primates in providing contingent tutoring.

Turn taking Turn taking is another intrinsic part of human language use [110]. Turns generally characterize cooperative exchanges in animals, but competitors may also benefit from having an undisturbed reception of the calls of their rivals. Among group-living animals, the most common context for turn-taking is contact calling between animals that are not in visual contact. Although this can be interpreted as cooperative behavior [111, 112], it is not necessarily so, as it does not always lead to coordinated movement or group fusion. However, in common marmosets, where signaling is predominantly prosocial, pronounced turn-taking during long bouts of antiphonal calling, reminiscent of human conversation, has been described [113, 114]. More generally, the coordination underlying turn-taking is greatly facilitated by the absence of gaze aversion. Among primates, callitrichids are known to readily engage in mutual gaze, in contrast to most other species, which typically avoid it [115, 116]. Thus, extensive turn-taking, while not unique among primates, is very pronounced in callitrichid monkeys.

In sum, correlates of language are not an automatic byproduct of VPL. They are at least as likely, if not more so, among species showing only vocal accommodation. Among the latter, most correlates are found in callitrichid cooperative breeders, but apart from babbling, the evidence from other primates is too scattered to definitively establish the link between the presence of these features and allomaternal care. Nonetheless, at the current state of knowledge, the most parsimonious conclusion is that features that also accompany human language are scarce among species without allomaternal care, regardless of the presence or absence of VPL. This is consistent with our hypothesis that increased vocal learning is associated with increased demands for cooperative behavior.

The emergence of language in Homo

Vocal production learning (VPL) is a critical foundation of the massive increase in acoustic diversity that evolved in humans. VPL requires both the ability to create novel sounds and vocalizations as well as the ability to learn them from others. With its emergence, the early hominin communication system with a call repertoire of at most a few dozen call types gradually turned into language: complex group-specific communication systems that require extensive learning.

Studies of animal species with VPL have yielded important insights into the convergent evolution of the relevant mechanisms, including the genetic basis of true VPL. However, our survey revealed a critical difference in the functions of VPL between humans and other species. The most extensively studied functional contexts in animals, such as sexual selection and parent-offspring recognition, do not align with the contexts in which language has evolved in our hominin ancestors. In our lineage, the main function is rich information exchange in a variety of contexts, enabled by the ability to produce arbitrary form-meaning pairings. This mismatch of underlying functions suggests that the animal species typically studied for VPL may offer little insight into reconstructing the conditions favoring language evolution.

We therefore suggest that a more plausible ancestral foundation for human VPL was vocal accommodation due to homophily (Figure 1B). In other words, its primary function is social alignment, providing the foundation for coordinated joint action and successful negotiation and servicing social bonds [117]. This aligns with the idea that language evolved primarily for close-range, often dyadic interactions.

Vocal accommodation is ubiquitous in primates, making it the most likely ancestral state in early hominins. The transition from accommodation to full-fledged VPL may not have posed a significant evolutionary hurdle. First, the frequent evolution of VPL in other lineages indicates that the genetic changes for its emergence in hominins can readily evolve. Second, we noted that the creativity that is necessary for VPL was already present in the earliest hominins. Third, recent work has revealed evidence for cortical control of vocalizations in enculturated apes [59–61]. If the same neurological connections also evolved in hominins in response to the need for enhanced cooperative communication, this greatly eased selection on strengthening cortical control to make vocalizations more volitional.

We further propose that the key functional context for human language evolution is the expanded use of vocal communication, triggered by a more consistent presence on the open savanna around 2 million years ago [118]. This new environment demanded increased cooperation, which in turn fostered greater reliance on vocal signals for coordination. We thus suggest that vocal accommodation gradually expanded into vocal production learning as social demands increased. Importantly, accommodation as a process or mechanism did not vanish; it persists within language in its original function.

Paleoanthropological evidence supports the scenario illustrated in Figure 1B. Unlike other great apes, *H. erectus* gradually developed a unique hunter-gatherer niche, shaped by the increasing demands of social transmission when complex technology accumulated [119] and social organization increased interdependence [120, 121]. Indeed, *H. erectus* clearly engaged in high levels of cooperation based on trust and social bonding [122–125], consistent with various cases of support for the injured as shown by evidence of healing after debilitating injuries [126, 127]. Finally, *H. erectus* also quite possibly already engaged in some level of cooperative breeding [128], as suggested by a brain size outside the demographically viable range of great apes [129] and the ability to survive in harsher habitats than their ancestors [130, 131].

This scenario of intense cooperation is supported by features of communication among the cooperatively breeding callitrichids. They share parental care and food, both of which involve extensive communication, and among primates, they show the most pronounced vocal accommodation. With their high volubility, vocal flexibility on short time scales and following pair formation, babbling by infants, turn-taking in calling, tutoring, and some level of voluntary control over vocalizations [97] they clearly share many functions of human language [110]. As we saw, many of these elements are missing in the species with VPL. We suggest that these features were already present in our hominin ancestors before language emergence and could therefore easily be recruited for language functions.

Overall, these findings support the suggestion that language evolved to improve the efficiency of cooperation in both foraging and child rearing [15, 63, 132–134]. No other primates, even highly cooperative ones like the callitrichids, made the transition to open-ended and locally unique vocal repertoires, let alone language. Two non-exclusive explanations are plausible. First, it took the large brains of hominins, with their strong memory and association areas, to turn a mere increase in vocal repertoires into language. Second, hominin cooperation became more complex. Language evolved to manage collective foraging movements, extensive food sharing, and allomaternal care that had evolved to a level well beyond that found in callitrichids (or for that matter other intensely cooperating species such as African wild dogs, which hunt and rear offspring together) [135]. The latter is compatible with the archaeological evidence revealing the incipient hunting and gathering lifeway [120, 128, 136]. Social transmission of complex technological skills, foraging tactics and social negotiation skills linked to a multilevel social system [137] was at least one element, increasingly critical considering the slow rate of acquisition [138]. Experiments in modern humans suggest that such transmission is more efficient when linguistic instruction is allowed [139]. The massive importance of social transmission in humans is compatible with the evolution of pedagogy [140], which extends the active teaching found in many cooperative breeders [141] to the communication system [101] and explains why human language could become an adaptation that relies to an unprecedented extent on learning.

Thus, increasing cooperation drove accommodation into VPL in *H. erectus*. But VPL abilities increased yet again after the split from the Neanderthals and Denisovans [67, 142–144] (Figure 1B). This was likely an evolutionary response to a new challenge in cooperation posed by the quickly increasing population size in *Homo sapiens* [68, 145]: How to identify partners for cooperation that one can trust? The solution to the challenge was the introduction of ethnic markers that would distinguish in-group strangers, who one could trust because any violation of trust would eventually return through gossip in the group, from out-group strangers, with whom cooperation would be more risky [68]. The most ubiquitous ethnic marker is through communication [78, 146] and this triggered the relentless schismogenetic diversification of the human communication system into thousands of dialects and languages [4]. This selected for increased levels of VPL abilities.

Language began to emerge in a hominin with strong dyadic and group-level bonds of trust, who had in addition more information to share than other such species. Once a richer communication system involving gestures, sounds, and noises was in place, selection could favor the evolution of full VPL: vocal communication is more efficient and does not require the use of hands or external props, which then can be used to support meaning transmission; it can also be produced at low volume and in the dark. These characteristics beyond the fixed and innate vocalizations of the great apes provide a clear selective advantage in the interdependent lifestyle of the hominin lineage, paving the way for language and concomitant cultural evolution.

Our findings suggest that research on the conditions favoring language origins should not focus on identifying VPL in other species, but rather on identifying novel form-function mappings in cooperative contexts that parallel the roots of language (see Figure 1). Both parrots and orangutans show promising, highly suggestive evidence of novel form-function mappings that turn into local traditions [26, 51]. Future work will profit from focusing on species with a need for communicating in a cooperative context.

References

[1] KE Langergraber, et al., Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl. Acad. Sci.* **109**, 15716–15721 (2012).

[2] F Wegdell, et al., The evolution of infant-directed communication: Comparing vocal input across all great apes. *Sci. Adv.* (to appear).

[3] M Leroux, B Chandia, AB Bosshard, K Zuberbühler, SW Townsend, Call combinations in chimpanzees: a social tool? *Behav. Ecol.* **33**, 1036–1043 (2022).

[4] B Bickel, AL Giraud, K Zuberbühler, CP Van Schaik, Language follows a distinct mode of extra-genomic evolution. *Phys. Life Rev.* **50**, 211–225 (2024).

[5] W Fitch, The evolution of speech: a comparative review. *Trends Cogn. Sci.* **4**, 258–267 (2000).

[6] T Nishimura, et al., Evolutionary loss of complexity in human vocal anatomy as an adaptation for speech. *Science* **377**, 760–763 (2022).

[7] N Staes, et al., FOXP2 variation in great ape populations offers insight into the evolution of communication skills. *Sci. Reports* **7**, 16866 (2017).

[8] RG Klein, Language and human evolution. *J. Neurolinguistics* **43**, 204–221 (2017).

[9] J Krause, et al., The derived FOXP2 variant of modern humans was shared with Neandertals. *Curr. Biol.* **17**, 1908–1912 (2007).

[10] C Schreiweis, et al., Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proc. Natl. Acad. Sci.* **111**, 14253–14258 (2014).

[11] D Dediu, SC Levinson, Neanderthal language revisited: not only us. *Curr. Opin. Behav. Sci.* **21**, 49–55 (2018).

[12] SE Fisher, Human genetics: The evolving story of FOXP2. *Curr. Biol.* **29**, R65 – R67 (2019).

[13] LG Casten, et al., Rapidly evolved genomic regions shape individual language abilities in present-day humans (2025).

[14] MS Ponce De León, et al., The primitive brain of early *Homo*. *Science* **372**, 165–171 (2021).

[15] M Tomasello, *The Origins of Human Communication.* (MIT Press), (2008).

[16] SC Levinson, On the human "interaction engine" in *Roots of human sociality: culture, cognition and interaction*, eds. NJ Enfield, SC Levinson. (Cambridge University Press, Cambridge), (2006).

[17] CP Van Schaik, C.P., *The Primate Origins of Human Nature.* (John Wiley & Sons, Hoboken, NJ), (2016).

[18] C Krupenye, J Call, Theory of mind in animals: Current and future directions. *Wiley Interdiscip. Rev. Cogn. Sci.* **10**, e1503 (2019).

[19] RM Seyfarth, DL Cheney, Precursors to language: Social cognition and pragmatic inference in primates. *Psychon. Bull. & Rev.* **24**, 79–84 (2017).

[20] A Whiten, N McGuigan, S Marshall-Pescini, LM Hopper, Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philos. Transactions Royal Soc. B: Biol. Sci.* **364**, 2417–2428 (2009).

[21] EJC Van Leeuwen, KA Cronin, DBM Haun, A group-specific arbitrary tradition in chimpanzees (Pan troglodytes). *Animal Cogn.* **17**, 1421–1425 (2014).

[22] EJC Van Leeuwen, et al., Getting to the bottom of social learning: Chimpanzees copy arbitrary behavior from conspecifics (2024).

[23] M Leroux, et al., Call combinations and compositional processing in wild chimpanzees. *Nat. Commun.* **14**, 2225 (2023).

[24] M Berthet, M Surbeck, SW Townsend, Extensive compositionality in the vocal system of bonobos. *Science* **388**, 104–108 (2025).

[25] SK Watson, et al., Nonadjacent dependency processing in monkeys, apes, and humans. *Sci. Adv.* **6** (2020).

[26] SA Wich, et al., Call cultures in orang-utans? *PLOS ONE* **7**, e36180 (2012).

[27] DP Watts, Production of grooming-associated sounds by chimpanzees (Pan troglodytes) at Ngogo: variation, social learning, and possible functions. *Primates* **57**, 61–72 (2016).

[28] MM Robbins, et al., Behavioral variation in gorillas: Evidence of potential cultural traits. *PLOS ONE* **11**, e0160483 (2016).

[29] CP Van Schaik, et al., Orangutan cultures and the evolution of material culture. *Science* **299**, 102–105 (2003).

[30] M Fröhlich, CP Van Schaik, Must all signals be evolved? A proposal for a new classification of communicative acts. *WIREs Cogn. Sci.* **11**, e1527 (2020).

[31] N Chomsky, *Aspects of the theory of syntax.* (MIT Press, Cambridge, Mass.), (1965).

[32] D Bickerton, *More than nature needs: Language, mind, and evolution.* (Harvard University Press, Cambridge, MA), (2014).

[33] H Giles, N Coupland, J Coupland, Accommodation theory: Communication, context, and consequence in *Contexts of Accommodation*, eds. H Giles, J Coupland, N Coupland. (Cambridge University Press), pp. 1–68 (1991).

[34] JS Pardo, E Pellegrino, V Dellwo, B Möbius, Special issue: Vocal accommodation in speech communication. *J. Phonetics* **95**, 101196 (2022).

[35] VM Janik, PJ Slater, Vocal Learning in mammals in *Advances in the Study of Behavior.* (Elsevier) Vol. 26, pp. 59–99 (1997).

[36] C Petkov, E Jarvis, Birds, primates and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* **4**, 12 (2012).

[37] G Arriaga, ED Jarvis, Mouse vocal communication system: Are ultrasounds learned or innate? *Brain Lang.* **124**, 96–116 (2013).

[38] VM Janik, M Knörnschild, Vocal production learning in mammals revisited. *Philos. Transactions Royal Soc. B: Biol. Sci.* **376**, 20200244 (2021).

[39] J Fischer, F Wegdell, F Trede, F Dal Pesco, K Hammerschmidt, Vocal convergence in a multi-level primate society: insights into the evolution of vocal learning. *Proc. Royal Soc. B: Biol. Sci.* **287**, 20202531 (2020).

[40] PL Tyack, A taxonomy for vocal learning. *Philos. Transactions Royal Soc. B: Biol. Sci.* **375**, 20180406 (2020).

[41] PT Martins, C Boeckx, Vocal learning: Beyond the continuum. *PLOS Biol.* **18**, e3000672 (2020).

[42] SC Vernes, et al., The multi-dimensional nature of vocal learning. *Philos. Transactions Royal Soc. B: Biol. Sci.* **376**, 20200236 (2021).

[43] ED Jarvis, Evolution of vocal learning and spoken language. *Science* **366**, 50–54 (2019).

[44] C Ten Cate, Re-evaluating vocal production learning in non-oscine birds. *Philos. Transactions Royal Soc. B: Biol. Sci.* **376**, 20200249 (2021).

[45] ME Wirthlin, et al., Vocal learning–associated convergent evolution in mammalian proteins and regulatory elements. *Science* **383**, eabn3263 (2024).

[46] C Scharff, J Petri, Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. *Philos. Transactions Royal Soc. B: Biol. Sci.* **366**, 2124–2140 (2011).

[47] SE Fisher, Evolution of language: Lessons from the genome. *Psychon. Bull. & Rev.* **24**, 34–40 (2017).

[48] EC Garland, PK McGregor, Cultural transmission, evolution, and revolution in vocal displays: Insights from bird and whale song. *Front. Psychol.* **11** (2020).

[49] M Knörnschild, Vocal production learning in bats. *Curr. Opin. Neurobiol.* **28**, 80–85 (2014).

[50] RJ Schusterman, Vocal learning in mammals with special emphasis on pinnipeds in *The evolution of communicative flexibility: complexity, creativity and adaptability in human and animal communication*, eds. DK Oller, U Gribel. (MIT Press, Cambridge, MA), pp. 41–70 (2008).

[51] JW Bradbury, TJS Balsby, The functions of vocal learning in parrots. *Behav. Ecol. Sociobiol.* **70**, 293–312 (2016).

[52] K Hammerschmidt, J Fischer, Constraints in primate vocal production in *Evolution of Communicative Flexibility*, eds. DK Oller, U Griebel. (The MIT Press), pp. 92–119 (2008).

[53] H Ruch, Y Zürcher, JM Burkart, The function and mechanism of vocal accommodation in humans and other primates. *Biol. Rev.* **93**, 996–1013 (2018).

[54] J Mitani, J Gros-Louis, Chorusing and call convergence in chimpanzees: Tests of three hypotheses. *Behaviour* **135**, 1041–1064 (1998).

[55] AJ Marshall, RW Wrangham, AC Arcadi, Does learning affect the structure of vocalizations in chimpanzees? *Animal Behav.* **58**, 825–830 (1999).

[56] C Crockford, I Herbinger, L Vigilant, C Boesch, Wild chimpanzees produce group-specific calls: A case for vocal learning? *Ethology* **110**, 221–243 (2004).

[57] KE Slocombe, K Zuberbühler, Chimpanzees modify recruitment screams as a function of audience composition. *Proc. Natl. Acad. Sci.* **104**, 17228–17233 (2007).

[58] SW Townsend, T Deschner, K Zuberbühler, Female chimpanzees use copulation calls flexibly to prevent social competition. *PLoS ONE* **3**, e2431 (2008).

[59] AR Lameira, Bidding evidence for primate vocal learning and the cultural substrates for speech evolution. *Neurosci. & Biobehav. Rev.* **83**, 429–439 (2017).

[60] WD Hopkins, JP Taglialatela, DA Leavens, Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behav.* **73**, 281–286 (2007).

[61] AG Ekström, C Gannon, J Edlund, S Moran, AR Lameira, Chimpanzee utterances refute purported missing links for novel vocalizations and syllabic speech. *Sci. Reports* **14**, 17135 (2024).

[62] SB Hrdy, JM Burkart, How reliance on allomaternal care shapes primate development with special reference to the genus homo in *Evolutionary Perspectives on Infancy*, eds. SL Hart, DF Bjorklund. (Springer International Publishing, Cham), pp. 161–188 (2022).

[63] AA Ghazanfar, DA Liao, DY Takahashi, Volition and learning in primate vocal behaviour. *Animal Behav.* **151**, 239–247 (2019).

[64] SR Hage, A Nieder, Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nat. Commun.* **4**, 2409 (2013).

[65] SR Hage, A Nieder, Dual neural network model for the evolution of speech and language. *Trends Neurosci.* **39**, 813–829 (2016).

[66] S Carouso-Peck, MH Goldstein, WT Fitch, The many functions of vocal learning. *Philos. Transactions Royal Soc. B: Biol. Sci.* **376**, 20200235 (2021).

[67] P Gunz, et al., Neandertal introgression sheds light on modern human endocranial globularity. *Curr. Biol.* **29**, 120–127.e5 (2019).

[68] P Cerrito, et al., Pleistocene origins of cultural and linguistic diversification: How Homo sapiens and Neanderthals differed. *EcoEvoRxiv* (2025).

[69] P Marler, Some ethological implications for neuroethology: the ontogeny of birdsong in *Advances in Vertebrate Neuroethology.* (Plenum Press, New York), pp. 21–52 (1983).

[70] LM Herman, et al., Humpback whale song: who sings? *Behav. Ecol. Sociobiol.* **67**, 1653–1663 (2013).

[71] D Duengen, A Ravignani, The paradox of learned song in a semi-solitary mammal. *Ethology* **129**, 445–453 (2023).

[72] AH Dalziell, JA Welbergen, B Igic, RD Magrath, Avian vocal mimicry: a unified conceptual framework. *Biol. Rev.* **90**, 643–668 (2015).

[73] J Sakata, S Woolley, R Fay, A Popper, *The Neuroethology of Birdsong.* (Springer International Publications, Cham), (2020).

[74] D Brown, *Human Universals.* (McGraw-Hill, New York), (1991).

[75] L Wilbrecht, F Nottebohm, Vocal learning in birds and humans. *Mental Retard. Dev. Disabil. Res. Rev.* **9**, 135–148 (2003).

[76] LA Parr, FBM De Waal, Visual kin recognition in chimpanzees. *Nature* **399**, 647–648 (1999).

[77] G Bateson, Culture Contact and Schismogenesis. *Man* **35**, 178 (1935).

[78] E Cohen, The evolution of tag-based cooperation in humans: The case for accent. *Curr. Anthropol.* **53**, 588–616 (2012).

[79] QD Atkinson, A Meade, C Venditti, SJ Greenhill, M Pagel, Languages evolve in punctuational bursts. *Science* **319**, 588–588 (2008).

[80] J Mansfield, H Leslie-O'Neill, H Li, Dialect differences and linguistic divergence: A crosslinguistic survey of grammatical variation. *Lang. Dyn. Chang.* **13**, 232–276 (2023).

[81] L Baciadonna, et al., Vocal accommodation in penguins (*Spheniscus demersus*) as a result of social environment. *Proc. Royal Soc. B: Biol. Sci.* **289**, 20220626 (2022).

[82] Y Zürcher, EP Willems, JM Burkart, Trade-offs between vocal accommodation and individual recognisability in common marmoset vocalizations. *Sci. Reports* **11**, 15683 (2021).

[83] N Phaniraj, K Wierucka, JM Burkart, Dynamic vocal learning in adult marmoset monkeys (2023).

[84] R Wanker, Y Sugama, S Prinage, Vocal labelling of family members in spectacled parrotlets, Forpus conspicillatus. *Animal Behav.* **70**, 111–118 (2005).

[85] VM Janik, LS Sayigh, RS Wells, Signature whistle shape conveys identity information to bottlenose dolphins. *Proc. Natl. Acad. Sci.* **103**, 8293–8297 (2006).

[86] MA Pardo, et al., African elephants address one another with individually specific name-like calls. *Nat. Ecol. & Evol.* **8**, 1353–1364 (2024).

[87] G Oren, et al., Vocal labeling of others by nonhuman primates. *Science* **385**, 996–1003 (2024).

[88] TL Chartrand, JL Lakin, The antecedents and consequences of human behavioral mimicry. *Annu. Rev. Psychol.* **64**, 285–308 (2013).

[89] DK Oller, *The Emergence of the Speech Capacity.* (Psychology Press), (2000).

[90] CT Sasaki, PA Levine, JT Laitman, ES Crelin, Postnatal descent of the epiglottis in man: A preliminary report. *Arch. Otolaryngol. - Head Neck Surg.* **103**, 169–171 (1977).

[91] HK Vorperian, RD Kent, Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *J. Speech, Lang. Hear. Res.* **50**, 1510–1545 (2007).

[92] BD Boysson-Bardies, L Sagart, C Durand, Discernible differences in the babbling of infants according to target language. *J. Child Lang.* **11**, 1–15 (1984).

[93] A Langehennig-Peristenidou, D Romero-Mujalli, T Bergmann, M Scheumann, Features of animal babbling in the vocal ontogeny of the gray mouse lemur (Microcebus murinus). *Sci. Reports* **13**, 21384 (2023).

[94] SM Ter Haar, et al., Cross-species parallels in babbling: animals and algorithms. *Philos. Transactions Royal Soc. B: Biol. Sci.* **376**, 20200239 (2021).

[95] AA Fernandez, LS Burchardt, M Nagy, M Knörnschild, Babbling in a vocal learning bat resembles human infant babbling. *Science* **373**, 923–926 (2021).

[96] C Snowdon, AM Elowson, 'Babbling' in Pygmy marmosets: Development after infancy. *Behaviour* **138**, 1235–1248 (2001).

[97] JM Burkart, et al., A convergent interaction engine: vocal communication among marmoset monkeys. *Philos. Transactions Royal Soc. B: Biol. Sci.* **377**, 20210098 (2022).

[98] MB Manser, et al., Vocal complexity in meerkats and other mongoose species in *Advances in the Study of Behavior*. (Elsevier) Vol. 46, pp. 281–310 (2014).

[99] GM Leighton, Cooperative breeding influences the number and type of vocalizations in avian lineages. *Proc. Royal Soc. B: Biol. Sci.* **284**, 20171508 (2017).

[100] JC Dunn, JB Smaers, Neural Correlates of Vocal Repertoire in Primates. *Front. Neurosci.* **12**, 534 (2018).

[101] J Schick, et al., The function and evolution of child-directed communication. *PLOS Biol.* **20**, e3001630 (2022).

[102] J Schick, S Stoll, Children learn best from their peers: The crucial role of input from other children in language development. *Open Mind* **9**, 665–676 (2025).

[103] L Feurstein, R Furrer, S Stoll, Turn-taking predicts vocabulary acquisition at 18 months: A study of daylong recordings in *BUCLD: Proceedings of the Boston University Conference on Language Development.* (Cascadilla Press), pp. 255–267 (2022).

[104] S Carouso-Peck, O Menyhart, TJ DeVoogd, MH Goldstein, Contingent parental responses are naturally associated with zebra finch song learning. *Animal Behav.* **165**, 123–132 (2020).

[105] DY Takahashi, DA Liao, AA Ghazanfar, Vocal learning via social reinforcement by infant marmoset monkeys. *Curr. Biol.* **27**, 1844–1852.e6 (2017).

[106] CP Chow, JF Mitchell, CT Miller, Vocal turn-taking in a non-human primate is learned during ontogeny. *Proc. Royal Soc. B: Biol. Sci.* **282**, 20150069 (2015).

[107] DY Takahashi, AR Fenley, AA Ghazanfar, Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Philos. Transactions Royal Soc. B: Biol. Sci.* **371**, 20150370 (2016).

[108] YB Gultekin, SR Hage, Limiting parental interaction during vocal development affects acoustic call structure in marmoset monkeys. *Sci. Adv.* **4**, eaar4012 (2018).

[109] J Fischer, K Hammerschmidt, Towards a new taxonomy of primate vocal production learning. *Philos. Transactions Royal Soc. B: Biol. Sci.* **375**, 20190045 (2019).

[110] SC Levinson, Turn-taking in human communication – origins and implications for language processing. *Trends Cogn. Sci.* **20**, 6–14 (2016).

[111] A Ravignani, L Verga, MD Greenfield, Interactive rhythms across species: the evolutionary biology of animal chorusing and turn-taking. *Annals New York Acad. Sci.* **1453**, 12–21 (2019).

[112] F Abreu, S Pika, Turn-taking skills in mammals: A systematic review into development and acquisition. *Front. Ecol. Evol.* **10** (2022).

[113] R Landman, et al., Close-range vocal interaction in the common marmoset (Callithrix jacchus). *PLOS ONE* **15**, e0227392 (2020).

[114] CT Miller, X Wang, Sensory-motor interactions modulate a primate vocal behavior: antiphonal calling in common marmosets. *J. Comp. Physiol. A* **192**, 27–38 (2006).

[115] SU Nummela, MJ Jutras, JT Wixted, EA Buffalo, CT Miller, Recognition memory in marmoset and macaque monkeys: A comparison of active vision. *J. Cogn. Neurosci.* **31**, 1318–1328 (2019).

[116] FM Miss, JEC Adriaense, JM Burkart, Towards integrating joint action research: Developmental and evolutionary perspectives on co-representation. *Neurosci. & Biobehav. Rev.* **143**, 104924 (2022).

[117] RIM Dunbar, The social brain hypothesis. *Evol. Anthropol. Issues, News, Rev.* **6**, 178–190 (1998).

[118] BL Pobiner, The zooarchaeology and paleoecology of early hominin scavenging. *Evol. Anthropol. Issues, News, Rev.* **29**, 68–82 (2020).

[119] AB Migliano, L Vinicius, The origins of human cumulative culture: from the foraging niche to collective intelligence. *Philos. Transactions Royal Soc. B: Biol. Sci.* **377**, 20200317 (2022).

[120] M Tomasello, AP Melis, C Tennie, E Wyman, E Herrmann, Two key steps in the evolution of human cooperation: The Interdependence Hypothesis. *Curr. Anthropol.* **53**, 673–692 (2012).

[121] EJ Ringen, P Duda, AV Jaeggi, The evolution of daily food sharing: A Bayesian phylogenetic analysis. *Evol. Hum. Behav.* **40**, 375–384 (2019).

[122] LC Aiello, C Key, Energetic consequences of being a *Homo erectus* female. *Am. J. Hum. Biol.* **14**, 551–565 (2002).

[123] EP Willems, CP Van Schaik, The social organization of Homo ergaster: Inferences from anti-predator responses in extant primates. *J. Hum. Evol.* **109**, 11–21 (2017).

[124] J Yravedra, S Rubio-Jara, LA Courtenay, JA Martos, Mammal butchery by homo erectus at the lower pleistocene acheulean site of Juma's korongo 2 (JK2), bed III, Olduvai Gorge, Tanzania. *Quat. Sci. Rev.* **249**, 106612 (2020).

[125] B Pobiner, Meat-Eating Among the Earliest Humans. *Am. Sci.* **104**, 110 (2016).

[126] P Spikins, H Rutherford, A Needham, From homininity to humanity: Compassion from the earliest archaics to modern humans. *Time Mind* **3**, 303–325 (2010).

[127] M Haeusler, Spinal pathologies in fossil hominins in *Spinal Evolution*, eds. E Been, A Gómez-Olivencia, P Ann Kramer. (Springer International Publishing, Cham), pp. 213–245 (2019).

[128] SB Hrdy, *Mothers and Others: The Evolutionary Origins of Mutual Understanding.* (Harvard University Press, Cambrdige, MA), (2009).

[129] K Isler, CP Van Schaik, How our ancestors broke through the gray ceiling: comparative evidence for cooperative breeding in early *homo*. *Curr. Anthropol.* **53**, S453–S465 (2012).

[130] D Lukas, T Clutton-Brock, Climate and the distribution of cooperative breeding in mammals. *Royal Soc. Open Sci.* **4**, 160897 (2017).

[131] CK Cornwallis, et al., Cooperation facilitates the colonization of harsh environments. *Nat. Ecol. & Evol.* **1**, 0057 (2017).

[132] JM Burkart, SB Hrdy, CP Van Schaik, Cooperative breeding and human cognitive evolution. *Evol. Anthropol. Issues, News, Rev.* **18**, 175–186 (2009).

[133] EA Smith, Communication and collective action: language and the evolution of human cooperation. *Evol. Hum. Behav.* **31**, 231–245 (2010).

[134] K Zuberbühler, Cooperative breeding and the evolution of vocal flexibility in *The Oxford handbook of language evolution*, eds. M Tallerman, KR Gibson. (Oxford University Press, Oxford), pp. 72–81 (2012).

[135] S Creel, NM Creel, *The African wild dog: behavior, ecology, and conservation*, Monographs in behavior and ecology. (Princeton University Press, Princeton, N.J), (2002).

[136] JM Burkart, S Hrdy, CP Van Schaik, Cooperative breeding and human cognitive evolution. *Evol. Anthropol.* **18**, 175–186 (2009).

[137] AB Migliano, L Vinicius, The origins of human cumulative culture: from the foraging niche to collective intelligence. *Philos. Transactions Royal Soc. B: Biol. Sci.* **377**, 20200317 (2021).

[138] EJM Meulman, AM Seed, J Mann, If at first you don't succeed… Studies of ontogeny shed light on the cognitive demands of habitual tool use. *Philos. Transactions Royal Soc. B: Biol. Sci.* **368**, 20130050 (2013).

[139] CA Caldwell, E Renner, M Atkinson, Human teaching and cumulative cultural evolution. *Rev. Philos. Psychol.* **9**, 751–770 (2018).

[140] G Csibra, G Gergely, Natural pedagogy as evolutionary adaptation. *Philos. Transactions Royal Soc. B: Biol. Sci.* **366**, 1149–1157 (2011).

[141] MA Kline, How to learn about teaching: An evolutionary framework for the study of teaching behavior in humans and other animals. *Behav. Brain Sci.* **38**, e31 (2015).

[142] Y Tajima, et al., A humanized NOVA1 splicing factor alters mouse vocal communications. *Nat. Commun.* **16**, 1542 (2025).

[143] D Gokhman, et al., Differential DNA methylation of vocal and facial anatomy genes in modern humans. *Nat. Commun.* **11**, 1189 (2020).

[144] R Buisan, J Moriano, A Andirkó, C Boeckx, A brain region-specific expression profile for genes within large introgression deserts and under positive selection in homo sapiens. *Front. Cell Dev. Biol.* **10**, 824740 (2022).

[145] T Cousins, A Scally, R Durbin, A structured coalescent model reveals deep ancestral structure shared by all modern humans (2024).

[146] MW Moffett, Human Identity and the Evolution of Societies. *Hum. Nat.* **24**, 219–267 (2013).

Supporting Information for
From vocal homophily to vocal repertoire flexibility: Unravelling the socioecological drivers of language evolution

Sabine Stoll, Carel P. van Schaik, Judith M. Burkart, Konatsu Ono, Nikhil Phaniraj, Zitan Song & Balthasar Bickel

Table S1: Evidence for functions across taxa

Function	Mechanism	Taxa [references]
Mate attraction	VPL	Australian musk duck [1] Hummingbirds [1] Songbirds [2–4] Whales [5–9] Pinnipeds [10–13] Bats [14, 15] Mice [16, 17]
Parent-offspring recognition	VPL Accomodation	Penguins [18, 19] Pinnipeds [13, 20] Dolphins [21] Bats [22–25]
Marking social boundaries	Accommodation	Hummingbirds [26] Parrots [27, 28] Songbirds [29–32] Wolves [33] Whales [34, 35] Bats [36, 37] Naked-mole rat [38] Nonhuman-primates [39–42] Homo sapiens
Servicing cooperative bonds	VPL Accommodation	Parrots [27, 43, 44] Songbirds [32, 45] Bats [22, 46, 47] Elephants [48, 49] Dolphins [50, 51] Whales [52] Goats [53] Non-human primates [54, 55] Homo sapiens
Creative form/meaning mapping	VPL	Parrots [27, 56] Orangutans [57] Homo sapiens

References

[1] C Ten Cate, Re-evaluating vocal production learning in non-oscine birds. *Philosophical Transactions of the Royal Society B: Biological Sciences* **376**, 20200249 (2021).

[2] C Catchpole, PJB Slater, C Catchpole, *Bird Song: Biological Themes and Variations.* (Cambridge Univ. Press, Cambridge), (2003).

[3] P Marler, Some ethological implications for neuroethology: the ontogeny of birdsong in *Advances in Vertebrate Neuroethology.* (Plenum Press, New York), pp. 21–52 (1983).

[4] DE Kroodsma, BE Byers, The Function(s) of Bird Song. *American Zoologist* **31**, 318–328 (1991).

[5] KM Stafford, Singing behavior in the bowhead whale in *Ethology and Behavioral Ecology of Mysticetes*, eds. CW Clark, EC Garland. (Springer International Publishing, Cham), pp. 277–295 (2022).

[6] LM Herman, The multiple functions of male song within the humpback whale (M*egaptera novaeangliae*) mating system: review, evaluation, and synthesis. *Biological Reviews* **92**, 1795–1818 (2017).

[7] E Oleson, et al., Behavioral context of call production by eastern North Pacific blue whales. *Marine Ecology Progress Series* **330**, 269–284 (2007).

[8] EC Garland, PK McGregor, Cultural transmission, evolution, and revolution in vocal displays: Insights from bird and whale song. *Frontiers in Psychology* **11** (2020).

[9] EC Garland, et al., Population structure of humpback whales in the western and central South Pacific Ocean as determined by vocal exchange among populations. *Conservation Biology* **29**, 1198–1207 (2015).

[10] RJ Schusterman, Vocal learning in mammals with special emphasis on pinnipeds in *The evolution of communicative flexibility: complexity, creativity and adaptability in human and animal communication*, eds. DK Oller, U Gribel. (MIT Press, Cambridge, MA), pp. 41–70 (2008).

[11] B Sjare, I Stirling, The breeding behavior of Atlantic walruses, *Odobenus rosmarus rosmarus* , in the Canadian High Arctic. *Canadian Journal of Zoology* **74**, 897–911 (1996).

[12] D Duengen, A Ravignani, The paradox of learned song in a semi-solitary mammal. *Ethology* **129**, 445–453 (2023).

[13] S Insley, AV Phillips, I Charrier, A review of social recognition in pinnipeds. *Aquatic Mammals* **29**, 181–201 (2003).

[14] M Knörnschild, S Blüml, P Steidl, M Eckenweber, M Nagy, Bat songs as acoustic beacons - male territorial songs attract dispersing females. *Scientific Reports* **7**, 13918 (2017).

[15] AA Fernandez, LS Burchardt, M Nagy, M Knörnschild, Babbling in a vocal learning bat resembles human infant babbling. *Science* **373**, 923–926 (2021).

[16] G Arriaga, EP Zhou, ED Jarvis, Of mice, birds, and men: The mouse ultrasonic song system has some features similar to humans and song-Learning birds. *PLOS ONE* **7**, e46610 (2012).

[17] G Arriaga, ED Jarvis, Mouse vocal communication system: Are ultrasounds learned or innate? *Brain and Language* **124**, 96–116 (2013).

[18] P Jouventin, T Aubin, T Lengagne, Finding a parent in a king penguin colony: the acoustic system of individual recognition. *Animal Behaviour* **57**, 1175–1183 (1999).

[19] L Baciadonna, et al., Vocal accommodation in penguins (*Spheniscus demersus*) as a result of social environment. *Proceedings of the Royal Society B: Biological Sciences* **289**, 20220626 (2022).

[20] I Charrier, Mother–offspring vocal recognition and social system in pinnipeds in *Coding Strategies in Vertebrate Acoustic Communication*, eds. T Aubin, N Mathevon. (Springer International Publishing, Cham) Vol. 7, pp. 231–246 (2020).

[21] LS Sayigh, et al., Bottlenose dolphin mothers modify signature whistles in the presence of their own calves. (year?).

[22] K Esser, U Schmidt, Mother-infant communication in the lesser spear-nosed bat *phyllostomus discolor* (chiroptera, phyllostomidae) — evidence for acoustic learning. *Ethology* **82**, 156–168 (1989).

[23] EZ Lattenkamp, SC Vernes, L Wiegrebe, Vocal production learning in the pale spear-nosed bat, *Phyllostomus discolor*. *Biology Letters* **16**, 20190928 (2020).

[24] JP Balcombe, Vocal recognition of pups by mother Mexican free-tailed bats, Tadarida brasiliensis mexicana. *Animal Behaviour* **39**, 960–966 (1990).

[25] M Knörnschild, M Feifel, EK Kalko, Mother–offspring recognition in the bat Carollia perspicillata. *Animal Behaviour* **86**, 941–948 (2013).

[26] M Araya-Salas, et al., Social group signatures in hummingbird displays provide evidence of co-occurrence of vocal and visual learning. *Proceedings of the Royal Society B: Biological Sciences* **286**, 20190666 (2019).

[27] JW Bradbury, TJS Balsby, The functions of vocal learning in parrots. *Behavioral Ecology and Sociobiology* **70**, 293–312 (2016).

[28] TF Wright, CR Dahlin, Vocal dialects in parrots: patterns and processes of cultural evolution. *Emu - Austral Ornithology* **118**, 50–66 (2018).

[29] C Walcott, JN Mager, W Piper, Changing territories, changing tunes: male loons, Gavia immer, change their vocalizations when they change territories. *Animal Behaviour* **71**, 673–683 (2006).

[30] H Loning, L Verkade, SC Griffith, M Naguib, The social role of song in wild zebra finches. *Current Biology* **33**, 372–380.e3 (2023).

[31] J Podos, PS Warren, The evolution of geographic variation in birdsong in *Advances in the Study of Behavior*. (Elsevier) Vol. 37, pp. 403–458 (2007).

[32] IG Moran, et al., Vocal convergence and social proximity shape the calls of the most basal Passeriformes, New Zealand Wrens. *Communications Biology* **7**, 575 (2024).

[33] M Zaccaroni, et al., Group specific vocal signature in free-ranging wolf packs. *Ethology Ecology & Evolution* **24**, 322–331 (2012).

[34] V Deecke, J Ford, P Spong, Dialect change in resident killer whales: implications for vocal learning and cultural transmission. *Animal Behaviour* **60**, 629–638 (2000).

[35] LE Rendell, H Whitehead, Vocal clans in sperm whales (*Physeter macrocephalus*). *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, 225–231 (2003).

[36] M Knörnschild, M Nagy, M Metz, F Mayer, O von Helversen, Learned vocal group signatures in the polygynous bat *Saccopteryx bilineata*. *Animal Behaviour* **84**, 761–769 (2012).

[37] JW Boughman, Greater spear-nosed bats give group-distinctive calls. *Behavioral Ecology and Sociobiology* **40**, 61–70 (1997).

[38] AJ Barker, et al., Cultural transmission of vocal dialect in the naked mole-rat. *Science* **371**, 503–507 (2021).

[39] N Mathevon, S Keenan, JMG Stevens, K Zuberbühler, F Levréro, Vocal accommodation in bonobos. *Animal Behaviour* **219**, 123014 (2025).

[40] C Crockford, I Herbinger, L Vigilant, C Boesch, Wild chimpanzees produce group-specific calls: A case for vocal learning? *Ethology* **110**, 221–243 (2004).

[41] AJ Marshall, RW Wrangham, AC Arcadi, Does learning affect the structure of vocalizations in chimpanzees? *Animal Behaviour* **58**, 825–830 (1999).

[42] N Masataka, K Fujita, Vocal learning of Japanese and rhesus monkeys. *Behaviour* **109**, 191–199 (1989).

[43] KS Berg, S Delgado, KA Cortopassi, SR Beissinger, JW Bradbury, Vertical transmission of learned signatures in a wild parrot. *Proceedings of the Royal Society B: Biological Sciences* **279**, 585–591 (2012).

[44] R Wanker, Y Sugama, S Prinage, Vocal labelling of family members in spectacled parrotlets, Forpus conspicillatus. *Animal Behaviour* **70**, 111–118 (2005).

[45] JM Trainer, DB McDonald, WA Learn, The development of coordinated singing in cooperatively displaying long-tailed manakins. *Behavioral Ecology* **13**, 65–69 (2002).

[46] Y Prat, M Taub, Y Yovel, Vocal learning in a social mammal: Demonstrated by isolation and playback experiments in bats. *Science Advances* **1**, e1500019 (2015).

[47] JW Boughman, Vocal learning by greater spear–nosed bats. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **265**, 227–233 (1998).

[48] JH Poole, PL Tyack, AS Stoeger-Horwath, S Watwood, Elephants are capable of vocal learning. *Nature* **434**, 455–456 (2005).

[49] AS Stoeger, et al., An Asian elephant imitates human speech. *Current Biology* **22**, 2144–2148 (2012).

[50] SL King, PK McGregor, Vocal matching: the what, the why and the how. *Biology Letters* **12**, 20160666 (2016).

[51] LS Sayigh, et al., Bottlenose dolphin mothers modify signature whistles in the presence of their own calves. *Proceedings of the National Academy of Sciences* **120**, e2300262120 (2023).

[52] EM Panova, AV Agafonov, A beluga whale socialized with bottlenose dolphins imitates their whistles. *Animal Cognition* **20**, 1153–1160 (2017).

[53] EF Briefer, AG McElligott, Social effects on vocal ontogeny in an ungulate, the goat, *Capra hircus*. *Animal Behaviour* **83**, 991–1000 (2012).

[54] H Ruch, Y Zürcher, JM Burkart, The function and mechanism of vocal accommodation in humans and other primates. *Biological Reviews* **93**, 996–1013 (2018).

[55] J Fischer, F Wegdell, F Trede, F Dal Pesco, K Hammerschmidt, Vocal convergence in a multi-level primate society: insights into the evolution of vocal learning. *Proceedings of the Royal Society B: Biological Sciences* **287**, 20202531 (2020).

[56] IM Pepperberg, Cognitive and communicative abilities of Grey parrots. *Applied Animal Behaviour Science* **100**, 77–86 (2006).

[57] SA Wich, et al., Call cultures in orang-utans? *PLOS ONE* **7**, e36180 (2012).