Minimal viable sound systems for language evolution

Adriano R. Lameira[1*], Steven Moran[2,3,4*]

[1]Department of Psychology, University of Warwick, Coventry, UK
[2]Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland
[3]Department of Anthropology, University of Miami, Coral Gables, FL, USA
[4]Linguistic Research Infrastructure (LiRI), University of Zurich, Zurich, Switzerland
*Corresponding authors. Email: adriano.lameira@warwick.ac.uk (A.R.L.); steven.moran@unine.ch (S.M.)

Abstract

Human vocal gamut covers 3000 unique speech sounds comprising the world's languages, with each new speaker having to learn the sounds of its new language. Since large, expandable repertoires are facilitated by vocal learning, this capacity has long been considered a prerequisite for speech and language evolution. The postulation of a vocal learning and repertoire size ceiling has, however, never been tested, though such inquiry is needed: the prevalent vocal learning hypothesis rejects great apes as vocal learners, against evolutionary principles of shared ancestry and descent with modification, while failing to explain of how vocal learning would have otherwise emerged in the human lineage in the first place. Although collectively diverse, individual languages typically use two-digit repertoires of consonants and vowels. These repertoire sizes are tantamount to great ape repertoires, also composed by consonant-like and vowel-like calls. Furthermore, while new speakers must learn new vocalic and consonantal sounds, these do not depend on laryngeal (i.e., voice) control, as posited by the vocal learning hypothesis, but rather on supralaryngeal control. The tongue, lips and mandible must shape the oral cavity in novel ways to produce new formants and constrictions upon which vowel and consonant recognition depend, respectively. Novel acquisition of both call categories is now also well-documented in great apes. It appears the first language(s) never required vocal learning capacities nor repertoires larger than greats apes' for full functionality. Instead, an increase in the ability to recombine existing sounds in novel ways seems to have been far more pivotal in the evolution of speech and language.

1. Introduction

Humans are vocal virtuosi, with more than 3000 unique speech sounds (i.e., contrastive linguistic segments) comprising the world's (spoken) languages (1). Over the last thirty years (2), the basic evolutionary assumption has been that human ancestors must have had the capacity to learn new vocal behaviours and expand their call repertoires before rich sound systems capable of expressing language could evolve in the hominid lineage (2–8). This expansion in vocal learning capacities would have allegedly started with an ape-like ancestor with zero vocal learning to capacity levels equal to far related species championed for their vocal learning skills, such as birds, bats, elephants, walruses, dolphins, and whales (2–8). If each of the world's languages must be universally acquired through vocal learning by new speakers, the premise that vocal learning was necessary for language evolution appears to be an axiomatic truth (4), though an incomplete one.

The vocal learning hypothesis openly rejects the possibility that great apes may be vocal learners, therefore, discarding the role of homology, shared ancestry and descent with modification for language evolution. These has two critical consequences. First, the vocal learning hypothesis

becomes inherently unable to otherwise inform how vocal learning would have evolved in the human lineage in the first place: No testable predictions can be made about the precursor biotope, body, brain and behaviour of ancestral hominids that could have led to language and the sound systems required for its expression. Such equivocation and lack of predictivenss defeats the purpose of any evolutionary hypothesis. Second, if an ape-like ancestor allegedly evolved the necessary sound systems for language from zero vocal learning, why haven't lineages campaigned for their vocal learning capacities developed language in the same period?

Here, we bring to scrutiny the accepted but untested premise that the evolution of language and the sound repertoires needed for its expression was impossible without advanced levels of vocal learning.

2. A question of size

2.1. How many sounds do modern languages require? Not many

Despite an excess of 3000 sounds worldwide, each language only deploys a *very* modest subset of our species' possible speech sound repertoire. Across all the world's languages, infinite messages are generated by sound repertoires composed by an average number of 24.8 consonants and 10.6 vowels (9) (Fig. 1).

Some languages use as few as 3 vowels, such as Haida and Arrente (spoken off the coast of British Columbia and in Northern Australia, respectively). Others use as few as 6 or 8 consonants, such as in Rotokas and Nasioi (off the coast of Eastern New Guinea) (1, 10) (Fig. 1). In Africa, languages can operate with as few as 2 vowels, as in Zulgo, Cuvok and Buwal (Northen Cameron) (1, 10) and as few as 11 or 13 consonants, as in Klao and Waama (Liberia and Benin, respectively).

If one would create "chimera" languages by pulling together nominal consonant and vowel repertoires across the world, these languages would be composed by 9 to 13 sounds. In fact, there are real languages that operate with a number of sounds that compare to these fictional repertoires. Languages like Pirahã (Amazonia, Brazil) and Rotokas (Papua New Guinea) exhibit a grand total of 11 sounds (1, 10) (Fig. 1). In Africa, languages can exhibit repertoires with total counts of 20 or 21 sounds, as in Efif (Nigeria), Nubian and Jomang (Sudan) (1, 10) (Fig. 1).

When averaging across the entire world, the mean number of sounds per language is 36.7, ranging between 22.4 and 42.7 depending on the language family (9). Even if our species' innate vocalizations (crying, screams, etc., which total 8 types; (11)) are added to these languages' repertoires, their total sizes fall in the same order of magnitude of the rest of the extant members of the Hominid family. Furthermore, sound diversity across modern languages today is known to represent an *over*-estimation of language's ancestral form due to human demographic growth over time (12). That is, older languages in the human lineage are known to have functioned with less sounds that observed today.

The size and composition of the repertoire of fully operational individual languages is, thus, starkly different from copious collections of sounds that the vocal learning hypothesis presumes necessary among human ancestors. Modern languages reject by and large premise that the first language(s) could have not operated on modest small-sized repertoires with less than a total of 20 (consonantal and vocalic) sounds.

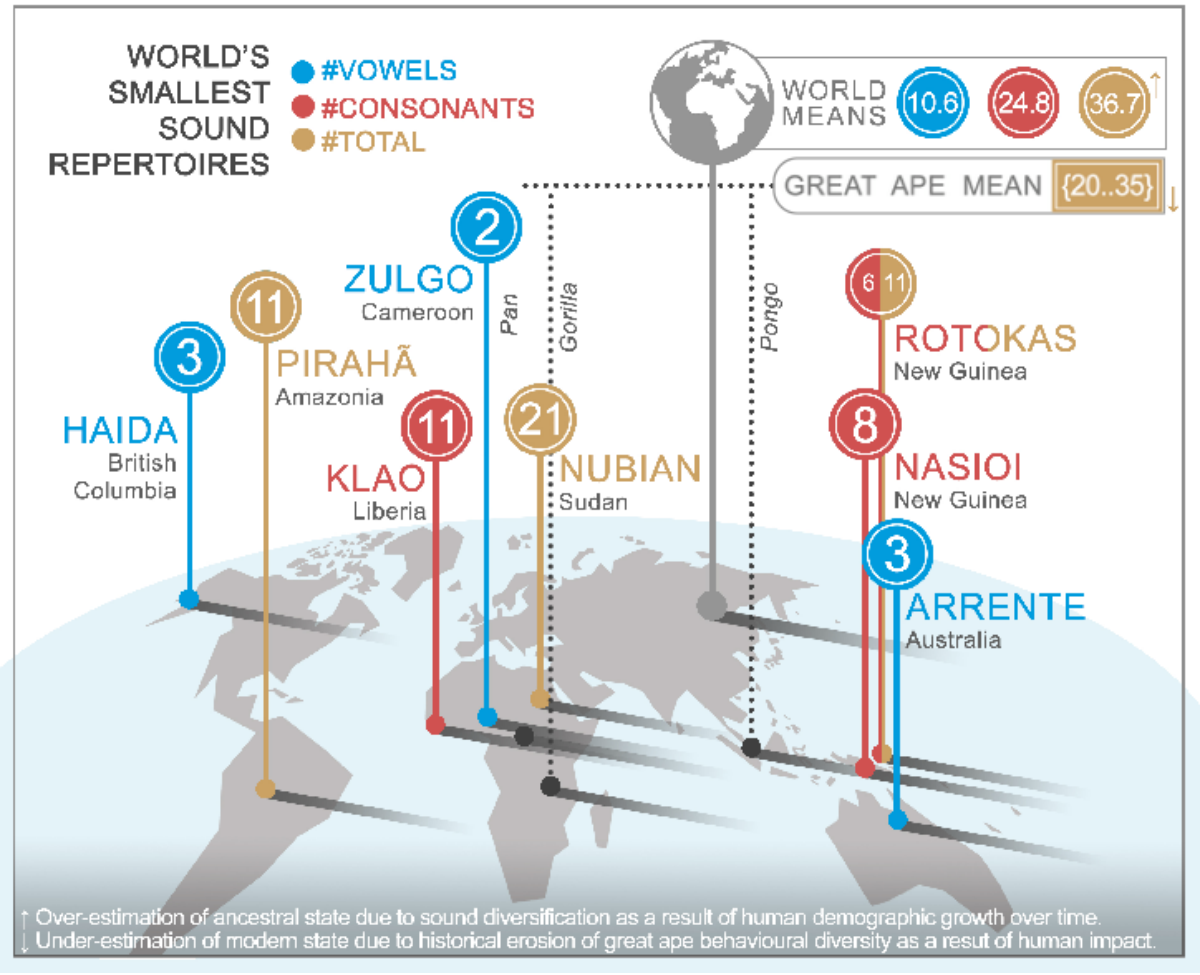

Fig. 1. World's smallest sound repertoires. The minimal sound system required for language around the world is fundamentally indistinguishable in magnitude and composition from great ape call repertoires.

2.2. How many sounds do great apes use? About as many

The reported call repertoire size in gorillas is 17 (13), which is assumed to be universal across populations, but this repertoire can be supplemented with at least 1 novel call tradition in the wild (14) and at least 10 novel sounds in captivity (15, 16). Gorilla vocal behaviour across populations in the wild and captivity remains, however, the least studied among great apes. The lower limit of the gorilla repertoire size is, thus, larger than that of several world languages. The gorilla repertoire upper size limit is liable to miscalculation and needs reassessment but given the demonstrated capacity for novel call acquisition in captivity, it is reasonable to expect this number to lay in the 20s.

The repertoires of chimpanzees and bonobos (*Pan*) have been described in 1977 (17) and in 1999 (18) as being composed by 14 and 15 call types, respectively. Although there has not been a concerted effort to re-assess these repertoires across sites (19), it is now known that some of the described calls are in fact compositions made up of different acoustically distinct calls (e.g., chimpanzee male pant hoots), that several calls under the same general name-banner have graded variations that are perceptually distinct to other chimpanzees (20–23) and that populations can develop local-specific call traditions, both in the wild (24, 25) and in captivity (26–31). A confident lower bound for repertoire size in *Pan* lies, therefore, in the mid 20s, but new repertoire inventories

and analyses are required to determine with greater precision the *Pan* repertoire upper size limit. Once again, chimpanzee repertoire sizes surpass those of many human languages.

The repertoire of orangutans (*Pongo*) has been investigated across populations and is composed by a foundation of 25 calls presumed to be universal across sites (32). As in the case of chimpanzee pant-hoots, orangutan long calls are compositions of different call types that should be counted as at least 6 different types instead of 1 (33). At least 6 additional local-specific call traditions have been identified so far in wild orangutans (32, 34, 35), with newly surveyed populations typically yielding new call types never heard before. In captivity, at least 3 additional sounds have emerged in different orangutan populations (36–40), with other sounds being known to exist but never having been officially described. This puts the *lower* bound for repertoire size in *Pongo* around the mid 30s and very close to the global average of language repertoire sizes.

Regrettably, human impact has driven great ape populations to the brink of extinction (41) and eroded their behavioural repertoires (42, 43), with several local call traditions assumed to have gone already extinct (44). Current great ape repertoire sizes are impoverished and thus *under*-estimated, reducing the gap observed today between modern great ape call repertoires sizes and modern language average repertoire sizes (which are in turn, as above mentioned, known to be overestimated). Both repertoires sit within the same range and magnitude order, falsifying the premise that language was unviable without large repertoires made greatly expandable by advanced vocal learning.

2.3. How many sounds did an ancestral language demand? Less than alleged

In sum, typical languages around the world make do with sound repertoires no larger than great apes'. This bodes ill with the belief that language stemmed from ancestral hominids who must have had the vocal learning skills of a dolphin or a songbird in order to be capable of expanding their repertoires with new calls. The notion of a vast ancestral sound repertoire finds no support in living hominids: language and great ape sound repertoires are neither copious nor are radically distinct from one another in size.

3. A matter of scenario

3.1. Weighing up two evolutionary trajectories

The disconnect between the actual sound repertoire sizes used for language and the projections of the vocal learning hypothesis invites a reassessment of presumed evolutionary scenarios (Fig. 2). Two scenarios come into view. The first, based on a premise that vocal learning was a prerequisite for language evolution, sees sound repertoires in ancestral apes having started small and then enlarging significantly well beyond their contemporary great ape counterparts. Which precursors and pressures would have driven such increase is, however, left undetermined – though this is the information that qualifies an evolutionary hypothesis as such. Subsequently, ancestral sound repertoires for language evolution would have regressed from a size like dolphin and songbird repertoires and attained sizes used today for language around the world, at similar levels to great apes. Which precursors and pressures would have caused such decrease is left, once again, undetermined (Fig. 2). The second scenario, based on observed human and great ape sound repertoire size, sees no positive or negative spurts in the human ancestral line for the emergence of the first language(s) (Fig. 2). Parsimony clearly supports the view that vocal learning or larger repertoires, per se, were not necessary for language evolution among Hominids.

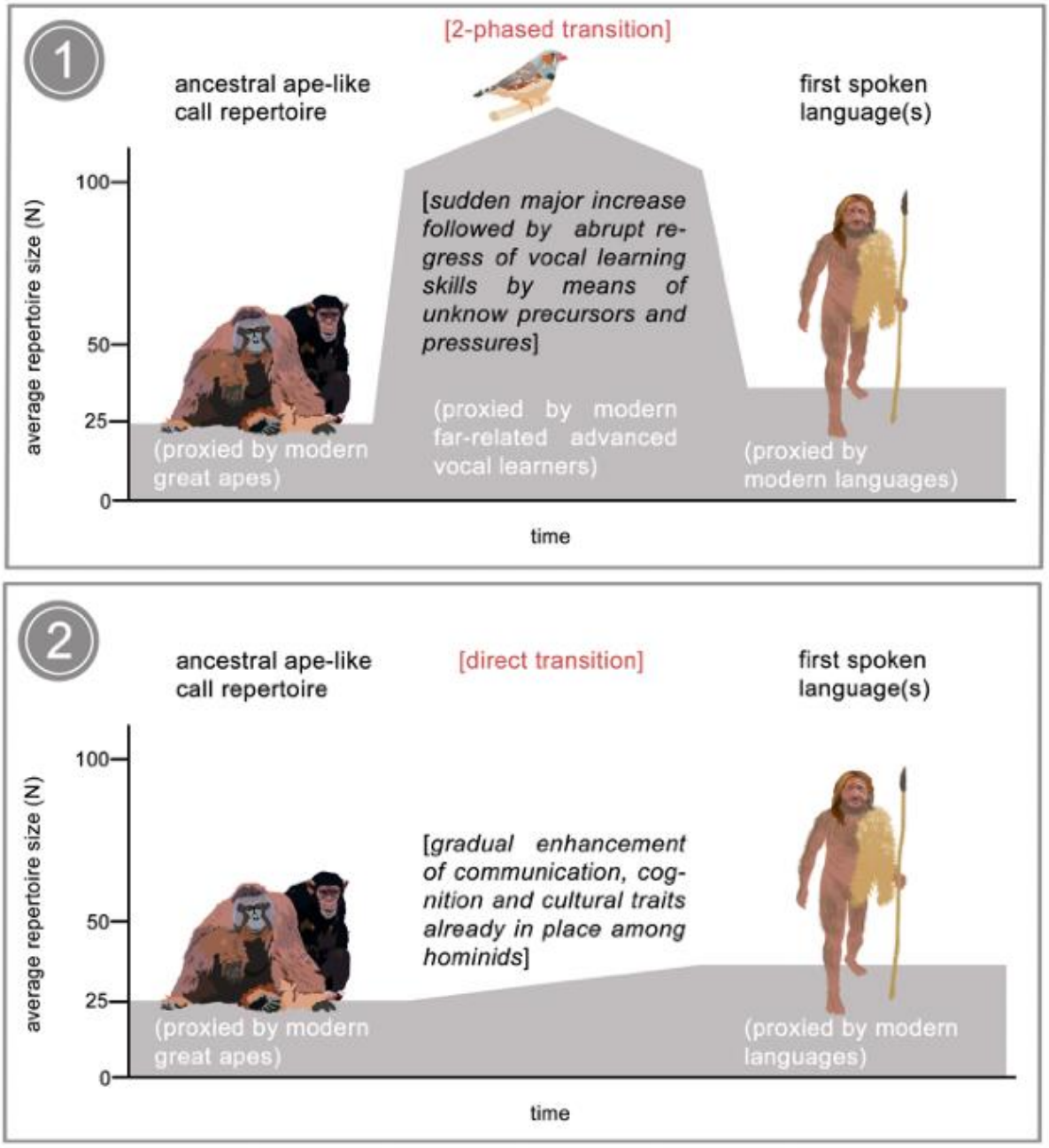

Fig. 2. Alternative evolutionary scenarios for language origin and evolution. Scenario 1 is based on vocal (production) learning hypothesis and is founded on unknown behaviour hominid precursors and unknown selective pressures that would have realistically operated in hominids. Scenario 1 supposes two major evolutionary transitions with opposite selective directions happened close one to another in a relatively short (but unspecified) window of time. Scenario 2 is based on present evidence from great ape and the world's languages sound repertoires. It presents a gradual and subtle evolution between ancestral and modern states of vocal communication within the hominid family.

3.2. Vocal learning checkpoint

The breath of sounds that humans and vocal learning species can learn today appears to have been a red herring for predicting sound repertoire requirements for language and its evolution from an

ancestral hominid call repertoire. Support for a vocal learning hypothesis has been primarily drawn from analogy and speculation. A re-evaluation of how vocal learning capacities in taxa far related from humans, and the assumption of how these have seemingly been acritically transposed to putative ancestral hominids is, therefore, warranted.

3.2.1. Voiced production is not more valuable than voiceless

Vocal learners across taxa use wildly different vocal anatomies for sound production, structures that are analogous between lineages, and thus, also analogous to primate vocal folds (45). For comparative purposes, the vocal learning hypothesis has put a functional premium on the structures that across taxa produce "voice" – a term used loosely for the acoustic output of species regardless of how it is actually achieved. This has led the vocal learning hypothesis to downplay other possible means of vocal production, including those found in modern languages and great apes. Critically, *voiceless* sounds produced with the mouth, via lip, tongue, jaw and/or airflow control, have been ignored or grossly dismissed as irrelevant to gauge primate vocal capacities and their putative precursor role for the evolution of speech sounds (2–4, 6, 7, 46–49).

Human voiceless utterances virtually always take the form of consonants, whereas human vowels are virtually always voiced (50). The combination consonant(s)+vowel(s) is one of the few uncontroversial universal traits of human languages, and one of the first to emerge in human infants in the form of canonical babbling (51, 52). Moreover, consonants are the most common type of sounds across the worlds' languages (53–55) and the only universal type of consonants are plosives/stops (like the speech sounds [p, t, k]) are prototypically voiceless. Trying to explain language evolution without recognizing the role of voicelessness is, hence, clearly a misdirected effort.

Indeed, (voiceless) consonants are especially numerous in languages with some of the largest sound repertoires in the world, such as !Xóõ (Botswana and Namibia) and Soghpo Tibetan (Tibet), with a total 161 and 133 sounds. Of these, 130 and 100 are consonants, respectively (1, 10). Indeed, this trend applies to all languages in the world, as well as those spoken in Eurasia and Africa, that is, the only two world regions where repertoires with >100 sounds have emerged. In humans, the larger a language's sound repertoire, the greater the proportion of voiceless sounds in that language (1, 10)(Fig. 3). It is critical to note here that it is exactly in languages with very large repertoires that the vocal learning hypothesis would predict that the learning of voiced calls would be paramount, while the opposite actually happens (Fig. 3). Once again, the premises of the vocal learning hypothesis find, thus, no empirical support in the real world.

Vocal learning deployed by children during language acquisition also contradicts the basic premise that voiced production is more important and more difficult than voiceless production. In English, for example, the [p] consonant sound is perfected until age 3, [ch, sh, z, j] consonant sounds are honed from age 4 to 7, and it can take 8 years until the [th, zh] consonant sounds are articulated in an adult-like fashion (56). This process can be even lengthier in other languages. Indeed, in English and many other languages, some consonant sounds are *never* mastered by speakers throughout their lifetimes. For example, a life-long difficulty to pronounce [r] consonant sounds is a phenomenon known as rhotacism, oftentimes referred as a "speech impediment" in medical contexts. The sounds that the vocal learning hypothesis have dismissed – sounds produced in the mouth – are exactly those that require some of the most dedicated and sustained learning effort from children (and adult speakers) around the world.

Dolphins and songbirds do not have lips or other soft pliant structures positioned at the end of their vocal tracts. Naturally, these animals cannot be reasonably expected to produce labial sounds or to filter sounds with lips they do not have (45). Anachronistically, however, the vocal learning hypothesis has accepted the opposite as true: that lip-produced and other voiceless sounds in great apes and human are not relevant for language origin because they are not observed in other vocal learners.

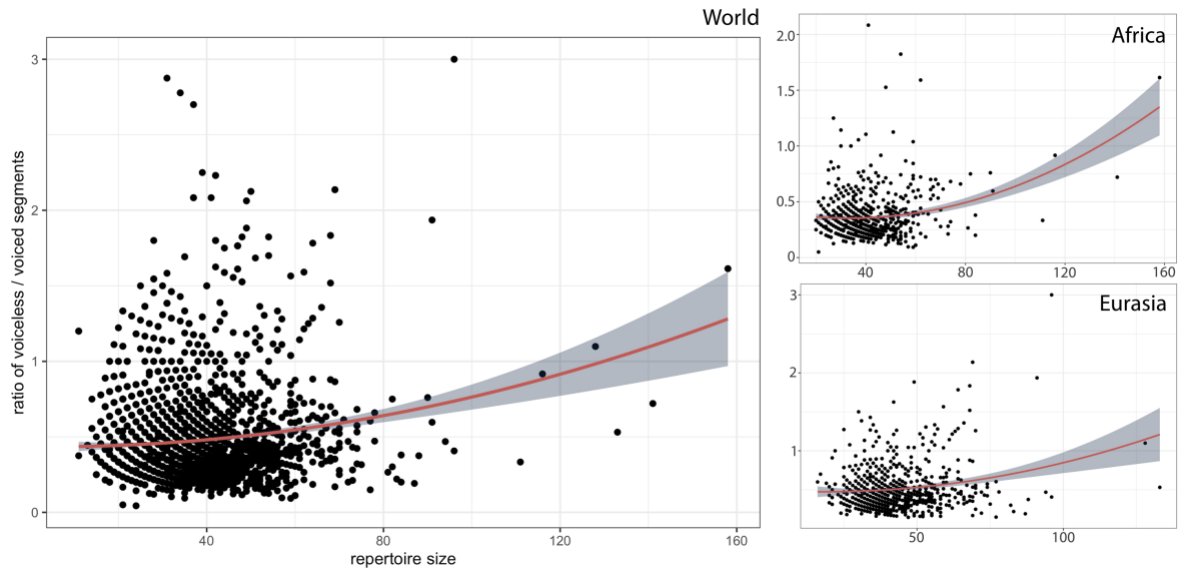

Fig. 3. Graphic representation of ratio of voiceless:voiced sounds in the world's languages plotted against language sound repertoire size. The more sounds a language employs, the larger the proportion and importance of voiceless sounds becomes in that language, contrarily to what is predicted by the vocal learning hypothesis.

3.2.2. Different vowel qualities are not created at the vocal folds

The emphasis put on voice/vocal fold/laryngeal control by the vocal learning hypothesis should seemingly help explain at least vowel evolution (57–59), since human vowels are virtually always voiced across languages. However, the production of different vowels does not sit at the vocal folds – it sits at the lips, tongue and lower jaw.

Vowel recognition is based on formants; bands of enhanced acoustic power that are "molded" by the shape and volume of the supralaryngeal vocal tract, which reflect where the lips, tongue and jaw are positioned or how they move. Vowel recognition is not based on voice pitch, which correlates with vocal fold oscillatory action. If this were the case, children, women and men would have different vowel-systems (60) and be unintelligible to one another. Their different voice pitches (e.g. high in children vs. low in men) would change their vowels. The observation that vowels derive from vocal tract configurations, and not from vocal fold action, has been made by linguists for decades (61–63), as well as proponents of the vocal learning hypothesis more than 20 years ago (64), but this fact has surprisingly never been factored in the hypothesis itself. Evolutionarily, increased voice control in our human ancestors could have facilitated the production of different voice frequencies (65). This is something that could have proven advantageous, for instance, in the production of different musical tones/notes with the voice. Increased voice control would have played no such role, however, in extending vowel space or range, though this has been the central assumption of the vocal learning hypothesis. Expansion of vowel space or range did not mechanically depend on voice control – it was brought by increased range in the controlled manoeuvring of the lips, tongue and lower jaw (54).

3.2.3. Vocal learning for linguistic expansion

The vocal learning hypothesis specifically stresses that control over *what* to produce is evolutionary more pertinent than control of *when* to produce, reflected in the original distinction between vocal *production* learning vs. vocal *contextual* learning, respectively. The latter form of vocal learning "affects the behavioral context or serial position of a signal" (46), enabling an individual to produce a signal in different contexts, timings and/or in combinations, but not to produce new signals or

alter the frequency of available ones. How does this distinction play out in relation to the emerging picture that the vocal learning has misrepresented hominid (incl. human) sound production capacities?

If our first linguistic ancestor already had at her disposal a consonant and vowel rootstock similar to the sound repertoires of great apes and typical modern languages, then the most critical step to produce syllables and word-like combinations and progress into new linguistic echelons would be to bring existing sounds together into new combinations in new contexts. Conversely, learning new sounds via vocal production learning would not necessarily translate into the formation of new combinations or linguistic permutations. All else being the same, the prediction is that vocal learning would in fact defer the emergence of new signal combinations because an individual could simply create a new call for every new required use, context or meaning, instead of combining existing calls to meet her new communicative needs. To increase her proto-linguistic expressiveness, our last pre-linguistic ancestor had to become generously skilled at vocal *contextual* learning, but not necessarily be as skilled at vocal production learning.

4. Ape-human vocal-verbal homologies

The similarities between modern languages and great ape sound repertoires go beyond the aspect of repertoire size (66, 67). Similarly to each and every language, great apes exhibit repertoires composed by consonant- and vowel-like calls (50, 53, 54, 68) that are shaped by social and cultural forces across historical time. For example, variation between great ape populations (69) can resemble human linguistic accents (where the same call is pronounced differently between locations, as in "tomato" pronounced by American vs. British English speakers) (23, 25, 50, 70) or human linguistic dialects (where distinct "synonym" calls are used in the same context and function between locations, as in "pants" vs. "trousers", whereas other populations may have no call for the same occasion) (35, 71). The number of social peers available for individuals to interact with across wild populations predicts how vocally innovative or confirmative those individuals are (44).

Social effects on great ape vocal production can also operate within seconds. For example, great apes can modulate vocal production depending on who they are interacting with (22, 72), their social peers' state of knowledge (73, 74) or their resources' perceived value (21). Other unique observations in the wild include the simultaneous production of consonant-like and vowel-like sounds through the joint engagement of supralaryngeal and laryngeal action, respectively, a feat otherwise only found in human beatboxing (75).

Under controlled settings in captivity, experimental work has confirmed that great apes exert real-time fine control over the lips (36, 39), tongue (76, 77), vocal fold action (i.e. voice) (37, 38) and vocal tract airflow (15, 36, 38, 77) for the purpose of sound production. This allows captive great apes, for instance, to learn human sounds directly from (unaware) caretakers (36, 39, 78), learn to mimic human speech-rhythm (40) [see also (79)], learn to imitate human words (31) and socially transmit horizontally (i.e. between peers) or vertically (i.e. down generations) atypical calls designed to gather the attention of human caretakers (28, 29, 36).

This motoric basis serves then as a canvas for advanced cognitive capacities, such as communication about past events via strings of consonant-vowel-like (and thus, human syllable-like) call combinations (80), communication about future locations (81), local traditions consisting of knowing how to acoustically deceive potential predators (34, 82, 83), and the construction of vocal motifs with hierarchical organization based on recursive operations (84).

Besides comparable sizes and homologous building blocks, the sound repertoires of great apes and humans are, thus, underpinned by motor and cognitive skills that only seem to differ within the hominid family in degree, not kind.

5. A hominid story

Minimalistic, but fully operational languages in humans and modest, but homologous vocal learning skills in great apes show that a minimal sound system powered by marginal vocal (production) learning capacities would not have prevented the emergence of language in our human ancestors. Structural resemblance in size and composition, and homologous neuro-behaviour between great ape and modern language sound repertoires hint at a direct evolutionary bridge across the vocal-verbal continuum.

Unquestionably, much can be learned from analogy and convergent evolution across a gamut of species. However, analogy is not a replacement to homology; convergent evolution is not a rebuttal to shared ancestry, as the vocal learning hypothesis has thus far presumed and promoted (2–4, 6, 7, 46–49). Literature reviews by non-great ape researchers recurrently make puzzlingly statements, forwarding that great ape vocal behaviour is of little empirical value, or of no heuristic consequence for how language took effect (only) in our clade (2–4, 6, 7, 46–49). The traditional view of language origin and evolution has come to expect that great apes ought to behave like humpback whales, walruses or parrots. It is clear, however, that the vocal capacities of one species should not serve as a benchmark to another, or as a basis for injunctions across distant taxa. Take the following examples: vocal learning in belugas is accomplished via motor control of the vestibular air sac (85), in elephants via control of the trunk (86) and in birds via control of the syrinx, a vocal organ that can contain up to three sound sources (87) – each one functionally analogous to a set of vocal folds in typical terrestrial mammals. How can this diversity of structures and behaviours be used to arbiter how any hominid (human or non-human) creates and learns new sounds when they do so by completely different means?

Analogy and convergent evolution allow us to infer selective forces that may have been experienced by distinct lineages, leading to related evolutionary outcomes, but they are mute about precursor forms. If great apes are not recognized for their own capacities, and as living proxies of ancestral hominid communication, cognition and culture (88), how will certain convergent selective pressures ever be understood *within the human clade*?

5.1. Speech versus song evolution

The primary premisses of the vocal learning hypothesis have been drawn from taxa far related to humans, namely, from their singing behaviour, such as birdsong and whale song, as well as taxa-specific behaviours that find no natural equivalent in humans, such as dolphins and bat sonar. These premisses have been freely transposed to speech and language evolution but, with dire heuristic consequences and evolutionary confusion, as flagged and explained here. To draw parallels, and hopefully gain evolutionary insight from analogy, between humans and birds, whales and other species with independent evolutionary histories and distinct biomes, biology, bodies, brains and behaviours, the signal systems under comparison should be kept the same. In point, presumptions about vocal learning requirements in singing species cannot be arbitrarily shifted to the evolution of speech and language in the human lineage, but they can and should within an evolutionary theory of human song. While some co-evolutionary interactions between song and speech have been believed to have played out at the base of the hominid clade since Darwin, the nature and causal directionality of possible interactions and their timing are still obscure and uncertain. To gain grip on this determinant phase in human evolution, it remains thus critically important to recognise the prerequisites of each system separately – those for the evolution of the first spoken languages versus those for the first forms of song – so that their interaction in the human deep past can be disentangled, reconstructed and understood. While song evolution can be accepted to have relied on laryngeal (i.e., vocal fold action) control to produce different frequency tones, therefore requiring vocal learning capacities as classically proposed, speech evolution cannot. Nonetheless, the vocal learning hypothesis will still require amelioration if it is to provide true evolutionary insight on the evolution of human song. For instance, to produce

different tones depends on vocal production learning, however, stringing different tones together to generate intervals, consonance and melodies relies on vocal contextual learning. Even for the evolution of human song, vocal production learning would have translated into a very limited scope of behavioural possibilities unless it was deployed in conjunction with other capacities. Ultimately, it is possible that the theoretical differentiation between vocal production learning versus vocal contextual learning is proven functionally incomplete and evolutionarily inaccurate.

5.2. Advanced vocal production learning in humans

Nowadays, it is by means of our capacity for multilingualism that vocal learning in modern humans mostly shines (even if not always fulfilled, i.e., by monolinguals). Great ape and modern language data indicate, however, that the minimal viable sound system for ancestral language(s) required unassuming vocal learning skills. Therefore, it seems that from the point of view of vocal learning, it is human contemporary multilingualism that requires evolutionary explanation, not language's origin per se. For instance, as long as multilingualism was an added benefit for the fitness and survival of early language-able populations, cumulative cultural evolution theory predicts that there would have been strong forces driving the advance of vocal learning thenceforward (89) – from an ape-like ancestor's modest forms to the blazing richness found across today's languages. Vocal learning would not have been a prerequisite, however, to the Ur-origin of language. In other words, advanced vocal learning capacities would have represented a consequence, not a cause, of the rise of the first language(s) in the human ancestral lineage.

6. Concluding remarks

Modern languages and great ape data show that the vocal learning hypothesis for language origin and evolution:

- Explains neither consonant, vowel, nor total repertoire size across the world's languages.
- Mischaracterizes great ape call repertoires in terms of size, composition and underlying neuro-control and behavioural capacities.
- Misrepresents the ancestral hominid sound repertoire and language minimal requirements.
- Is based on convoluted evolutionary scenarios defined by unknown behavioural precursors and unknown selective pressures.
- Neglects the central role of voiceless and consonant-like production in language.
- Attaches misplaced importance to vocal fold control for the expansion of vowel range.
- Fails to recognize the evolutionary importance of vocal *contextual* learning vs vocal *production* learning.

Other shortcomings and inconsistencies have been identified in the vocal learning hypothesis [reviewed in (68, 90)]. Although it may still prove useful for the study of communication and evolution in other taxa, its service as a framework for the study of language origin and evolution *in the human clade* has been exhausted and will probably require reform beyond recognition, if it is to be maintained at all. Instead, ape-human vocal-verbal homologies are set to help generate parsimonious and powerful testable predictions about the precursor conditions required for the evolution of speech and language.

References

1. S. Moran, D. McCloy, *PHOIBLE 2.0* (Max Planck Institute for the Science of Human History, 2019).
2. V. M. Janik, P. J. Slater, Vocal learning in mammals. *Advances in the Study of Behavior* **26**, 59–99 (1997).
3. M. H. Christiansen, S. Kirby, Language evolution: consensus and controversies. *Trends in Cognitive Sciences* **7**, 300–307 (2003).
4. J. J. Bolhuis, C. D. Wynne, Can evolution explain how minds work? *Nature* **458**, 832–833 (2009).
5. T. W. Fitch, E. D. Jams, "Birdsong and Other Animal Models for Human Speech, Song, and Vocal Learning" in *Language, Music, and the Brain*, (The MIT Press, 2013), pp. 499–540.
6. E. D. Jarvis, Evolution of vocal learning and spoken language. *Science* **366**, 50–54 (2019).
7. S. C. Vernes, V. M. Janik, W. T. Fitch, P. J. B. Slater, Vocal learning in animals and humans. *Phil. Trans. R. Soc. B* **376**, 20200234 (2021).
8. M. E. Wirthlin, *et al.*, Vocal learning–associated convergent evolution in mammalian proteins and regulatory elements. *Science* **383**, eabn3263 (2024).
9. S. Moran, D. Blasi, "Cross-linguistic comparison of complexity measures in phonological systems" in *Measuring Grammatical Complexity*, F. J. Newmeyer, L. B. Preston, Eds. (Oxford University Press, 2014), pp. 217–240.
10. H. Hammarström, R. Forkel, M. Haspelmath, S. Bank, *Glottolog 4.3* (Max Planck Institute for the Science of Human History, 2020).
11. A. Anikin, R. ath, T. Persson, Human Non-linguistic Vocal Repertoire: Call Types and Their Meaning. *Journal of Nonverbal Behavior* **37**, 1–28 (2017).
12. C. Perreault, S. Mathew, Dating the Origin of Language Using Phonemic Diversity. *PLoS ONE* **7**, e35289 (2012).
13. R. Salmi, K. Hammerschmidt, D.-S. M Diane, Western Gorilla Vocal Repertoire and Contextual Use of Vocalizations. *Ethology* **119**, n/a-n/a (2013).
14. M. M. Robbins, *et al.*, Behavioral Variation in Gorillas: Evidence of Potential Cultural Traits. *PLOS ONE* **11**, e0160483 (2016).
15. M. Perlman, N. Clark, Learned vocal and breathing behavior in an enculturated gorilla. *Animal cognition* **18**, 1165–1179 (2015).
16. R. Salmi, M. Szczupider, J. Carrigan, A novel attention-getting vocalization in zoo-housed western gorillas. *PLoS ONE* **17**, e0271871 (2022).
17. P. Marler, R. Tenaza, "Signaling behavior of apes with special reference to vocalizations" in *How Animals Communicate*, T. Sebeok, Ed. (Indiana University Press, 1977), pp. 965–1033.
18. M. Bermejo, A. Omedes, Preliminary Vocal Repertoire and Vocal Communication of Wild Bonobos (Pan paniscus) at Lilungu (Democratic Republic of Congo). *Folia Primatologica* **70**, 328–357 (1999).
19. F. Wegdell, *et al.*, An updated vocal repertoire of wild adult bonobos (*Pan paniscus*). [Preprint] (2025). Available at: http://biorxiv.org/lookup/doi/10.1101/2025.01.23.634282 [Accessed 31 January 2025].
20. C. Crockford, T. Gruber, K. Zuberbühler, Chimpanzee quiet hoo variants differ according to context. *R. Soc. open sci.* **5**, 172066 (2018).

21. A. K. Kalan, R. Mundry, C. Boesch, Wild chimpanzees modify food call structure with respect to tree size for a particular fruit species. *Animal Behaviour* **101**, 1–9 (2015).
22. K. E. Slocombe, K. Zuberbuhler, Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences* **104**, 17228–17233 (2007).
23. S. K. Watson, *et al.*, Vocal Learning in the Functionally Referential Food Grunts of Chimpanzees. *Current Biology* **25**, 495–499 (2015).
24. A. Arcadi, Phrase structure of wild chimpanzee pant hoots: Patterns of production and interpopulation variability. *American journal of primatology* **39**, 159–178 (1996).
25. C. Crockford, I. Herbinger, L. Vigilant, C. Boesch, Wild Chimpanzees Produce Group-Specific Calls: a Case for Vocal Learning? *Ethology* **110**, 221–243 (2004).
26. W. D. Hopkins, J. P. Taglialatela, D. A. Leavens, Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour* **73**, 281–286 (2007).
27. A. J. Marshall, R. W. Wrangham, A. Arcadi, Does learning affect the structure of vocalizations in chimpanzees? *Animal Behaviour* **58**, 825–830 (1999).
28. J. P. Taglialatela, L. Reamer, S. J. Schapiro, W. D. Hopkins, Social learning of a communicative signal in captive chimpanzees. *Biology letters* **8**, 498–501 (2012).
29. J. L. Russell, M. Joseph, W. D. Hopkins, J. P. Taglialatela, Vocal learning of a communicative signal in captive chimpanzees, Pan troglodytes. *Brain and Language* **127**, 520–525 (2013).
30. A. G. Ekström, Viki's First Words: A Comparative Phonetics Case Study. *Int J Primatol* (2023). https://doi.org/10.1007/s10764-023-00350-1.
31. A. Ekström, C. Gannon, J. Edlund, S. Moran, A. R. Lameira, Chimpanzee utterances refute purported missing links for novel vocalizations and syllabic speech. *Sci Rep* **14**, 17135 (2024).
32. M. E. Hardus, *et al.*, "A description of the orangutan's vocal and sound repertoire, with a focus on geographic variation" in *Orangutans*, S. Wich, M. T. Setia, S. S. Utami, C. Schaik, Eds. (Oxford University Press, 2009), pp. 49–60.
33. B. Spillmann, *et al.*, Acoustic properties of long calls given by flanged male orang-utans (Pongo pygmaeus wurmbii) reflect both individual identity and context. *Ethology* **116**, 385–395 (2010).
34. M. Hardus, *et al.*, Tool use in wild orang-utans modifies sound production: a functionally deceptive innovation? (2009). https://doi.org/10.1098/rspb.2009.1027.
35. S. A. Wich, *et al.*, Call cultures in orang-utans? *PloS one* **7**, e36180 (2012).
36. A. R. Lameira, *et al.*, Orangutan (Pongo spp.) whistling and implications for the emergence of an open-ended call repertoire: A replication and extension. *Journal of the Acoustical Society of America* **134**, 1–11 (2013).
37. A. R. Lameira, M. E. Hardus, A. Mielke, S. A. Wich, R. W. Shumaker, Vocal fold control beyond the species-specific repertoire in an orang-utan. *Scientific reports* **6**, 30315 (2016).
38. A. R. Lameira, R. W. Shumaker, Orangutans show active voicing through a membranophone. *Sci Rep* **9**, 12289 (2019).
39. S. Wich, *et al.*, A case of spontaneous acquisition of a human sound by an orangutan. *Primates* **50**, 56–64 (2009).
40. A. R. Lameira, *et al.*, Speech-like rhythm in a voiced and voiceless orangutan call. *PloS one* **10**, e116136 (2015).

41. A. Estrada, *et al.*, Impending extinction crisis of the world's primates: Why primates matter. e1600946 (2017). https://doi.org/10.1126/sciadv.1600946.
42. H. S. Kühl, *et al.*, Human impact erodes chimpanzee behavioral diversity. *Science* **363**, 1453–1455 (2019).
43. C. P. van Schaik, Fragility of Traditions: The Disturbance Hypothesis for the Loss of Local Traditions in Orangutans. *Int J Primatol* **23**, 527–538 (2002).
44. A. R. Lameira, *et al.*, Sociality predicts orangutan vocal phenotype. *Nat Ecol Evol* (2022). https://doi.org/10.1038/s41559-022-01689-z.
45. M. Garcia, M. Manser, Bound for Specific Sounds: Vocal Predisposition in Animal Communication. *Trends in Cognitive Sciences* S1364661320301376 (2020). https://doi.org/10.1016/j.tics.2020.05.013.
46. V. M. Janik, P. J. Slater, The different roles of social learning in vocal communication. *Animal Behaviour* **60**, 1–11 (2000).
47. E. Lattenkamp, V.-S. in Sciences, Vocal learning: a language-relevant trait in need of a broad cross-species approach. *Current Opinion in Behavioral Sciences* (2018).
48. T. W. Fitch, Empirical approaches to the study of language evolution. *Psychonomic Bulletin & Review* **24**, 1–31 (2017).
49. J. A. Soha, S. Peters, Vocal Learning in Songbirds and Humans: A Retrospective in Honor of Peter Marler. *Ethology* n/a-n/a (2015). https://doi.org/10.1111/eth.12415.
50. A. R. Lameira, *et al.*, Proto-consonants were information-dense via identical bioacoustic tags to proto-vowels. *Nature Human Behaviour* **1**, 0044 (2017).
51. P. K. Kuhl, Early language acquisition: cracking the speech code. *Nature reviews. Neuroscience* **5**, 831–843 (2004).
52. J. L. Locke, Babbling and early speech: continuity and individual differences. *First Lang* **9**, 191–205 (1989).
53. A. R. Lameira, The forgotten role of consonant-like calls in theories of speech evolution. *Behavioral and Brain Sciences* **37**, 559–560 (2014).
54. A. R. Lameira, I. Maddieson, K. Zuberbuhler, Primate feedstock for the evolution of consonants. *Trends in cognitive sciences* **18**, 60–62 (2014).
55. P. Ladefoged, I. Maddieson, *The Sounds of the World's Languages* (John Wiley {& Sons, 1996).
56. E. K. Sander, When are Speech Sounds Learned? *J Speech Hear Disord* **37**, 55–63 (1972).
57. Boë, Evidence of a Vocalic Proto-System in the Baboon (Papio papio) Suggests Pre-Hominin Speech Precursors. *PLOS ONE* **12**, e0169321 (2017).
58. L.-J. Boë, *et al.*, Which way to the dawn of speech?: Reanalyzing half a century of debates and data in light of speech science. *Sci. Adv.* **5**, eaaw3916 (2019).
59. T. W. Fitch, B. Boer, N. Mathur, A. A. Ghazanfar, Monkey vocal tracts are speech-ready. *Science Advances* **2**, e1600723–e1600723 (2016).
60. D. A. Puts, *et al.*, Sexual selection on male vocal fundamental frequency in humans and other anthropoids. *Proceedings. Biological sciences / The Royal Society* **283**, 20152830 (2016).
61. K. Johnson, P. Ladefoged, M. Lindau, Individual differences in vowel production. *The Journal of the Acoustical Society of America* **94**, 701–714 (1993).
62. G. Fant, *Acoustic theory of speech production* (Walter de Gruyter, 1970).
63. G. E. Peterson, H. L. Barney, Control Methods Used in a Study of the Vowels. *The Journal of the Acoustical Society of America* **24**, 175–184 (1952).

64. T. W. Fitch, The evolution of speech: a comparative review. *Trends in Cognitive Sciences* **4**, 258–267 (2000).
65. K. Pisanski, V. Cartei, M. Carolyn, J. Raine, D. Reby, Voice Modulation: A Window into the Origins of Human Vocal Control? *Trends in Cognitive Sciences* **20**, 304–318 (2016).
66. A. Ekström, *et al.*, Phonetic properties of chimpanzee, gorilla, and orangutan hoots tell a uniform story and point to new frontiers. (2024).
67. S. Moran, M. Maiolini, A. Lameira, Great Ape Vocal Repertoires Are All Similar in Size: Now What? in A. Ravignani, *et al.*, Eds. (2022).
68. A. R. Lameira, Bidding evidence for primate vocal learning and the cultural substrates for speech evolution. *Neuroscience & Biobehavioral Reviews* **83**, 429–439 (2017).
69. A. R. Lameira, R. Delgado, S. Wich, Review of geographic variation in terrestrial mammalian acoustic signals: Human speech variation in a comparative perspective. *Journal of Evolutionary Psychology* **8**, 309–332 (2010).
70. S. K. Watson, *et al.*, Reply to Fischer et al. *Current Biology* **25**, R1030–R1031 (2015).
71. C. P. van Schaik, *et al.*, Orangutan Cultures and the Evolution of Material Culture. *Science* **299**, 102–105 (2003).
72. J. C. Mitani, G.-L. Julie, Chorusing and Call Convergence in Chimpanzees: Tests of Three Hypotheses. *Behaviour* **135**, 1041–1064 (1998).
73. C. Crockford, R. M. Wittig, R. Mundry, K. Zuberbuhler, Wild chimpanzees inform ignorant group members of danger. *Current Biology* **22**, 142–146 (2012).
74. C. Crockford, R. M. Wittig, K. Zuberbuhler, Vocalizing in chimpanzees is influenced by social-cognitive processes. *Science Advances* **3**, e1701742 (2017).
75. A. R. Lameira, M. E. Hardus, Wild orangutans can simultaneously use two independent vocal sound sources similarly to songbirds and human beatboxers. *PNAS Nexus* **2**, pgad182 (2023).
76. M. Belyk, B. G. Schultz, J. Correia, D. S. Beal, S. A. Kotz, Whistling shares a common tongue with speech: bioacoustics from real-time MRI of the human vocal tract. *Proc. R. Soc. B* **286**, 20191116 (2019).
77. C. Hayes, *The ape in our house* (Harper, 1951).
78. W. H. Furness, Observations on the mentality of chimpanzees and orang-utans. *Proceedings of the American Philosophical Society* **55**, 281–290 (1916).
79. A. S. Pereira, E. Kavanagh, C. Hobaiter, K. E. Slocombe, A. R. Lameira, Chimpanzee lip-smacks confirm primate continuity for speech-rhythm evolution. *Biol. Lett.* **16**, 20200232 (2020).
80. A. R. Lameira, J. Call, Time-space–displaced responses in the orangutan vocal system. *Sci Adv* **4**, eaau3401 (2018).
81. C. P. van Schaik, L. Damerius, K. Isler, Wild Orangutan Males Plan and Communicate Their Travel Direction One Day in Advance. *Plos One* **8**, e74896 (2013).
82. A. R. Lameira, *et al.*, Population-specific use of the same tool-assisted alarm call between two wild orangutan populations (Pongo pygmaeus wurmbii) indicates functional arbitrariness. *Plos One* **8**, e69749 (2013).
83. B. Boer, S. A. Wich, M. E. Hardus, A. R. Lameira, Acoustic models of orangutan hand-assisted alarm calls. *The Journal of experimental biology* **218**, 907–914 (2015).
84. A. R. Lameira, M. E. Hardus, A. Ravignani, T. Raimondi, M. Gamba, "Recursive self-embedded vocal motifs in wild orangutans" (Animal Behavior and Cognition, 2023).
85. E. M. Panova, A. V. Agafonov, A beluga whale socialized with bottlenose dolphins imitates their whistles. *Animal Cognition* **20**, 1153–1160 (2017).

86. A. S. Stoeger, *et al.*, An Asian elephant imitates human speech. *Current Biology* **22**, 2144–2148 (2012).
87. S. M. Garcia, *et al.*, Evolution of Vocal Diversity through Morphological Adaptation without Vocal Learning or Complex Neural Control. *Current Biology* **27**, 2677-2683.e3 (2017).
88. A. R. Lameira, J. Call, Understanding Language Evolution: Beyond *Pan* -Centrism. *BioEssays* **42**, 1900102 (2020).
89. M. Derex, A. Mesoudi, Cumulative Cultural Evolution within Evolving Population Structures. *Trends in Cognitive Sciences* **24**, 654–667 (2020).
90. P. T. Martins, C. Boeckx, Vocal learning: Beyond the continuum. *PLoS Biol* **18**, e3000672 (2020).

Acknowledgements

Funding

UK Research and Innovation Future Leaders Fellowship grant MR/T04229X/1 (ARL)
Swiss National Science Foundation (Grant No. PCEFP1_186841) (SM)

Author contributions

Conceptualization: ARL and SM
Visualization: ARL and SM
Writing: ARL and SM

Competing interests

Authors declare that they have no competing interests.