Biol Philos (2017) 32:759–778
https://doi.org/10.1007/s10539-017-9607-x

Where did language come from? Connecting sign, song, and speech in hominin evolution

Anton Killin[1,2]

Received: 28 August 2017 / Accepted: 22 November 2017 / Published online: 30 November 2017

Abstract Recently theorists have developed competing accounts of the origins and nature of protolanguage and the subsequent evolution of language. Debate over these accounts is lively. Participants ask: Is music a direct precursor of language? Were the first languages gestural? Or is language continuous with primate vocalizations, such as the alarm calls of vervets? In this article I survey the leading hypotheses and lines of evidence, favouring a largely gestural conception of protolanguage. However, the "sticking point" of gestural accounts, to use Robbins Burling's phrase, is the need to explain how language shifted to a largely vocal medium. So with a critical eye I consider Michael Corballis's most recent expression of his ideas about this transition (2017's *The Truth About Language: What It Is And Where It Came From*). Corballis's view is an excellent foil to mine and I present it as such. Contrary to Corballis's account, and developing Burling's conjecture that musicality played some role, I argue that the foundations of an evolving musicality (i.e., evolving largely independently of *language*) provided the means and medium for the shift from gestural to vocal dominance in language. In other words, I suggest that an independently evolving musicality prepared ancient hominins, morphologically and cognitively, for intentional articulate vocal production, enabling the evolution of speech.

Keywords Evolution of language · Gesture · Great ape communication · Musicality · Protolanguage · Protomusic · Speech · Vocalization

✉ Anton Killin
anton.killin@anu.edu.au

[1] School of Philosophy and Centre of Excellence for the Dynamics of Language, Australian National University, Acton, ACT 2601, Australia

[2] Philosophy Department, Florida International University, Miami, FL 33199, USA

Introduction

Language evolution is a fascinating puzzle—or more accurately, series of puzzles—requiring a piecemeal and integrative approach. The puzzle piece that I consider in this article has become known as the (hypothesised) gesture-to-vocal transition. The road ahead is as follows. In the next section, "Protolanguage", I consider the leading hypotheses about the nature/origins of protolanguage.[1] I argue for a largely gestural conception of protolanguage, though that is not to say that vocalizations were absent from the picture.[2] Even so, theorists that posit that the origins of language are (largely) gestural, as I do, owe an explanation of the shift from a (largely) manual form of protolanguage to full speech ("Finding Voice?"). I consider gestural theorist Michael Corballis's most recent explanation of such a shift.[3] I critique aspects of Corballis's account ("Corballis tackles the gesture-speech transition"), and present my alternative, bringing musicality into the story ("Musicking across the gap"). Note that even if one does not accept the gesture-dominant conception of protolanguage, considerations of vocal musicality are still relevant in explaining how hominins shifted from having fairly non-flexible, chimpanzee-like vocal communication to a flexible vocal protolanguage (perhaps heavily assisted with gestures). Finally I identify some priorities for future research, including echo phonology and prospects for its import to evolutionary theorizing ("Towards arbitrariness: the echo phonology hypothesis"), and then sum up.

Protolanguage

In this section I consider three leading hypotheses about the origins and nature of hominin protolanguage. The first amounts to a *musical* conception of protolanguage, sometimes called a *musilanguage* (Darwin 1871; Brown 2000; Mithen 2005; Fitch 2010; Lawson 2014). The idea is that music and language share a single common ancestor—a vocal musical protolanguage—although the specifics differ from advocate to advocate. On this view, protolanguage in the ancestral hominin lineage is discontinuous with nonhominin great ape intentional communication (great apes do not sing), and is instead an example of convergence (e.g., with gibbon song, whale song, bird song). The protosyntax of musical protolanguage is generally considered by advocates of the hypothesis to be *holistic*—consisting of complete and unanalysable phrases with a "whole message",[4] which fractionate over time as syntax evolves.

[1] Terms like "protolanguage" and "language-ready" are a convenience, not meant to imply an element of purpose or teleology.

[2] Indeed, it is likely that language evolution took hold in the kinesic and oral-aural modalities together (see Kendon 2016) to some extent—and in coevolutionary tandem.

[3] Corballis (2017); for earlier versions see Corballis (2002, 2009).

[4] Fitch puts it like this: 'To the extent that there was a lexicon, it was a simple list of tunes or "riffs"—complex, multi-unit phrases linked to whole, context-bound events' (Fitch 2010, p. 476).

According to the second hypothesis, although protolanguage is vocal, it is not "musical" but *word-like* or "lexical". On this view, protolanguage resembles not animal *song*, but a referentially displaced version of typical primate vocalization, for example the much-discussed vervet monkey alarm calls (three distinct vocal calls for three different predators/evasion strategies). Our hominin ancestors, one version of the story goes, gained more and more complex mental representation systems until they could divide the world up into "word-size" pieces (Bickerton 1990). Once that occurred, protolanguage "popped out" as the pieces received lexical labels. This enabled basic protolinguistic communication: 'in protolanguage, the speaker thought of a word and then transmitted it directly to the organs of speech, then the next, and the next, without linking them in the brain prior to utterance' (Bickerton 2009, p. 232).

However, voice-dominant hypotheses about protolanguage are not well supported. Primate vocalization and human language are handled by different brain areas. Homologues of Wernicke's and Broca's areas (crucial to language processing in humans) exist in primates but don't handle vocalization (but rather gesture; see Rizzolatti and Arbib 1998; Kohler et al. 2002). Neocortical neural structures in general are not used in primate vocalization, which appears to be controlled by the limbic system. Primate vocalization is largely involuntary.[5] The vervet alarm calls, to take an example, are *symptomatic*. They are typical reactive and automatic responses to perceived stimuli. Typically vervets cannot omit, or produce-on-demand, such calls.

Interestingly, some apes seem to be "aware" of the largely involuntary nature of their vocalizations. Jane Goodall (1986) tells of a young chimp, excited upon discovering a banana, that ingeniously suppressed its pant hoots by muffling its mouth with its hand, so as not to inform the bigger chimpanzees of its discovery. Notice that the vocalization was *affective* and *automatic*, not intentional; the chimp's voluntary control was in the manual, not vocal, domain. Even gibbon song, as impressive as it is to our musical sensibilities, is not flexible, under voluntary control, but is reactive/automatic and affective (see Geissmann 2002). Top-down (voluntary/intentional) aspects of great ape communication are largely gestural, not vocal (Hobaiter and Byrne 2014; Pika and Mitani 2006). That said, recent research indicates that chimpanzees have the capacity for voluntary vocal control (Slocombe and Zuberbühler 2007; Watson et al. 2015; Schel et al. 2013; Crockford et al. 2012; Fitch and Zuberbühler 2013). So, presumably, some basic and restricted form of intentional vocal control, and intentional response to conspecific vocalization, was available to the *Pan-Homo* last common ancestor, which ancient hominins (and chimpanzees and bonobos) have been incrementally building on.

The third hypothesis, then, amounts to a *gestural* conception of protolanguage (more accurately: a gesture-*dominant* conception—as just noted, vocalizations were present too[6]). Advocates of this view include Corballis (2002, 2009, 2017), Sterelny

[5] Largely, but not exclusively, voluntary; recent evidence suggests great ape vocalization may be a little more flexible than previously thought. For discussion, see Kendon (2016); Irvine (2016).

[6] I thus resist the widespread (e.g. Kendon 2016) 'gesture-first' versus 'voice-first' nomenclature.

(2012, 2018), Hurford (2014), Tomasello (2008) and Hewes (1973).[7] This view posits straightforward continuity with great ape intentional communication. Intentional human gestural and vocal communication today employ similar neural systems (Newman et al. 2002; Kimura 1993). These systems are largely distinct from areas of the brain associated with emotion—the areas typically employed in primate vocal communication, and in automatic human utterances (e.g., in reflex swearing and grunting—these are regulated by different neural structures than ordinary language).[8] And to the extent that ontogeny is thought to recapitulate phylogeny, note that voluntary gestural communication is prior to voluntary vocal communication in human infant development (see e.g. Goldin-Meadow and Alibali 2013; Esteve-Gibert and Prieto 2014). From about 10 months, infants point in order to single out an object of interest, for example. And pointing, both by infants and carers, is a means by which the infant reduces referential ambiguity (Kalagher and Yu 2006; O'Neill et al. 2005). From 10 to 24 months, infants mimic the pantomime and playful gestures of carers. And young toddlers spontaneously produce iconic gestures from around 2 years of age—that is, around the same time as they begin to make two-word utterances (see Behne et al. 2014). Young deaf children exposed to sign language learn it in much the same way as they would learn speech, including going through a manual babbling phrase (Petitto and Marentette 1991). And deaf carers using sign language tend to produce slower, larger, exaggerated gestures, and more repetition, when interacting with deaf infants than with signing adults (Masataka 1992)—a gestural version of "motherese" (i.e., infant-directed speech).

In the gestural domain there is more scope for iconicity[9] (with due caution: see Irvine 2016 for developmental limitations). The hands and arms can mime physical shapes of objects, people, animals, and actions. And the intended meaning can often be easily comprehended (perhaps after a little practice/experience). Kim Shaw-Williams has pointed out to me just how difficult it is to describe the rather basic concept *spiral* in words; yet all it takes to communicate the idea is draw a spiral shape in the air with a pointed finger—plausibly, the same goes for communicating directions and pathways (e.g., which route we shall take on today's hunt—and indeed for communicating during the hunt). The meaning of an iconic pot-stirring gesture, in the right context (say, a child helping a parent cook), is easily expressed and grasped. As Begby (2017) points out, such iconic gestures are spontaneously

[7] Later work by Bickerton (e.g. 2009) explicitly allows for a role for gestures in establishing referential displacement. He envisions a scenario in which an individual has located a carcass somewhere in the distance and needs to recruit others to go and help exploit that carcass: 'waving and screaming and pointing "Thataway!", the elephant noise or the hippo noise or whatever it was could have only one meaning: dead megabeast, food for the taking only a short march away' (2009, p. 161).

[8] Also laughter, sobbing, and so on; these reactive, affective vocalizations may be continuous with primate vocalization.

[9] For example: 'We can't tell, just by the sound of the words, what kind of motion is indicated by *walk, run, swim, fly,* or *crawl*. But in sign languages, some iconicity has often been retained, and the meanings of the signs for these motion-types can be more successfully guessed at' (Hurford 2014, p. 107). Perniss and Vigliocco (2014) argue that iconicity played a key role in language evolution in establishing displacement (i.e., reference to that which is beyond the immediate here-and-now) and continues to play a key role in ontogeny in supporting the development of word-referentiality, crucial to vocabulary acquisition and meaningful communication.

utilised and developed into an idiosyncratic homesign system by deaf children born to non-sign-language-using parents. These children are without access to a public conventional language yet they, and their families, have no problem establishing novel gestural means of communication, and over time these homesign systems develop to exhibit morphological and syntactic structure (Goldin-Meadow 2003). Moreover, even today, modern sign languages comprise many iconic gestures: see Fig. 1 for the signs for *hammer* (resembling the action of using a hammer) and *deer* (resembling the shape of deer's antlers) in British Sign Language (BSL).[10]

Advocates of the gestural hypothesis stress gestural communication's plausibility as a means for getting protolanguage established in our smaller brained hominin ancestors.[11] And they stress the intentional gestural communication of chimpanzees and bonobos, our closest great ape cousins (e.g. Pollick and de Waal 2007). Hostetter et al. (2001) point out that chimps will use pointing gestures with human experimenters to signify a desire (e.g., they will point to a banana, wanting the human to pass it to them), if the human is looking (indicating that it is intentional; Tomasello 2008). And they often gesture to one another. Hobaiter and Byrne's important study, over the course of 266 days in Budongo National Park in Uganda, noted 4397 gestures of 66 kinds, such as *stop that, follow me, go away, mate with me* (see Hobaiter and Byrne 2011, 2014). Anecdotal reports of field-researchers describe (admittedly rare) instances of chimpanzee gestural "showing-how" communication. For example, a chimpanzee mother has been observed miming, in slow motion, how to hold a stone for nut-cracking, to her young daughter who hasn't quite got the hang of it yet (Boesch 1993; see also Boesch and Boesch-Achermann 2000; Russon and Andrews 2011). Although attempts to teach apes to speak have been abject failures, human-trained great apes such as Kanzi have been taught sign systems rather effectively (albeit with limits on the number of elements that could be combined in a sentence that even young human children can supersede) and with an impressive mastery of the distinction between proper names and general categories. Individual great apes can be trained to communicate via sign because their genetic infrastructure enables it—thus great ape species are merely *selection-limited* with respect to protolanguage, not *variation-limited* (see Számadó and Szathmáry 2006 for the distinction).[12] In other words, as concerns protolanguage, there is no need to posit a discontinuity: chimpanzees and bonobos have "proto-protolanguage capacities" and are thus protolanguage-ready; they just need

[10] According to Thompson et al. (2012), children learning BSL tend to acquire iconic sign production and comprehension before that of non-iconic signs.

[11] The following is a typical articulation of this: 'In a group of people just beginning to signal meanings to each other, many more meanings could be guessed from manual and facial gestures than from attempts to express them vocally. It would be easier to get a gestural language off the ground in the first place than a speech-based one' (Hurford 2014, p. 107). Indeed we have seen the birth of Nicaraguan Sign Language, a newly established, full language, built from the idiosyncratic homesign gestures of deaf children (Senghas 1995; Senghas et al. 2004; see Begby 2017 for philosophical implications).

[12] 'A transition is variation limited when the available genetic variation in the given lineage does not offer even a partial solution to the problem at hand, and it takes considerable time (in evolutionary terms) for the necessary variation to arise. By contrast, a transition is selection limited if the necessary genetic prerequisites of a possible transition are present, but the given transition is not selected for as this would require a specific ecological or social context' (Számadó and Szathmáry 2006, p. 555).

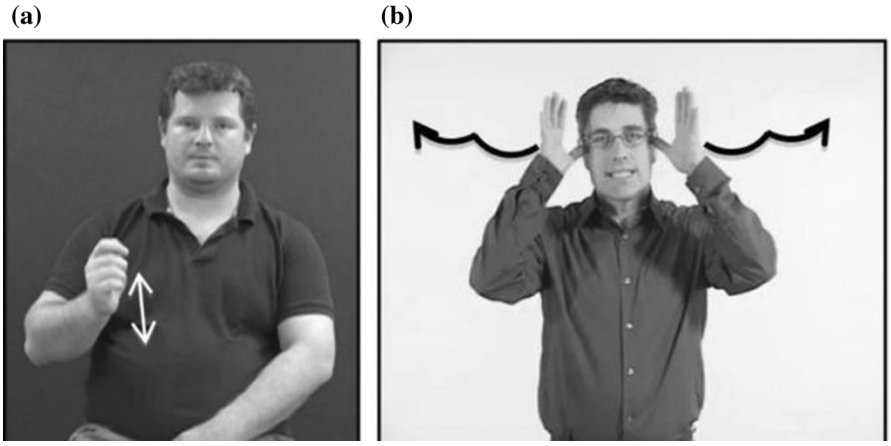

Fig. 1 Iconic signs in BSL: **a** *hammer*; **b** *deer*. Extracted from Perniss and Vigliocco (2014), reproducible under the terms of the Creative Commons Attribution License 3.0

a context in which it pays them to develop protolanguage. I suspect we can safely presume that the *Pan-Homo* last common ancestor was likewise protolanguage-ready.[13]

The brain science available also supports the gestural view. Recall that Broca's area is a major neural site of language processing in humans (e.g. Novick et al. 2010); a homologous site is present in other primates but does not respond to conspecific vocalizations (Kohler et al. 2002). In macaques, this is where the famous mirror neurons were discovered that correlate the observation and production of gestures (e.g. Gallese et al. 1996; Rizzolatti et al. 1988; Rizzolatti and Arbib 1998). In humans today there is much overlap in the brain's response/control of intentional gesture/sign language and speech (Rizzolatti and Craighero 2004; Capek et al. 2008; Newman et al. 2010), suggesting that in hominin evolution brain regions for primate gestural communication were recruited for modern speech (Hurford 2014).

So, henceforth, I shall assume a gesture-dominant view of protolanguage.[14] Yet the "sticking point" of gestural accounts is the need to explain how language shifted to a largely vocal medium (Burling 2005). In the next section, I spell out this sticking point.

[13] For more on great ape gestural communication in the context of language evolution see e.g. Moore (this issue), Sterelny (this issue).

[14] The evidence in its favour is not decisive, of course. However taken together it makes for a very compelling package, in my view.

Finding voice?

One explanatory challenge for gestural theorists is to specify how a predominantly gestural intentional communicative system transitioned to a predominantly vocal one (at least, among hearing individuals). Yet notice that we all (or at least the vast majority of us) continue to gesture while communicating vocally, sometimes without realizing. Many of us cannot give a lecture, conference presentation, or public speech, provide directions to a lost tourist, or explain how to execute a physical task to a novice, for example, without moving our arms about or gesturing in an expressive way. We point to things as an index; we accompany everyday speech with iconic pantomimic gestures (e.g., *push* gestures, *rotate* gestures).[15] We cup our ears to signify that we didn't hear our interlocutor; we wave, we give thumbs-up, thumbs-down, and we flip the bird. Nonetheless the fact remains that for real-time intentional linguistic communication, vocalizations are now primary.

It is perhaps unsurprising that if language's origins were gestural, such a transition would have taken place. There are a number of plausibly adaptive advantages for such a shift (*properly qualified*, of course; modern sign languages are harnessed by deaf people effectively just as spoken languages are by hearing people). For our hominin ancestors these include:

1. Lower time-costs; speaking is often quicker than gesturing.
2. Use in the dark, in dimly lit spaces (especially as fire control takes off, extending the usable hours of day for communication and other social pursuits), and in environments with limited conspecific visibility (like wooded or tall-sedge areas).
3. Use when one's hands are occupied, or body occluded in general (e.g. in wading or bathing), as well as for pedagogical purposes (e.g. in "showing how" demonstrative teaching and learning) and so on.
4. To better address individuals not in one's front line of sight (Elizabeth Irvine has pointed out that there are serious constraints on gestural communication when there are greater than 3 or 4 interlocutors—see also Sterelny 2016).
5. Freeing the eyes in general. Interlocutors can visually focus on what each other is doing (e.g., in toolmaking or teaching/learning) or look outside of the immediate conversational context.
6. For signalling *signalhood* (making articulate speech sounds is only good for speech; a wave might be to say hello or to shoo flies, an arm-up gesture could be confused with a stretch).
7. And once the transition is made, speech is metabolically less costly than gesture, requiring little more energy expenditure than that of breathing.

Yet the task remains to explain how the transition could have occurred—to ask what enabled it, its adaptive advantages notwithstanding. I take it that there are some salient constraints on giving a satisfactory explanation. The explanation has to

[15] Such pantomimic additions to speech enhance comprehension when congruent, and stifle it when incongruent. It appears that the dynamic influence of gesture on speech and vice versa is obligatory (Kelly et al. 2010).

be phylogenetically plausible. It has to be empirically constrained (i.e., tied where possible to the palaeoanthropological record) in order to explain why the transition occurred when it did. It has to not be reliant on low probability events or "magic bullet" scenarios. It must explain the move to a system comprising predominantly arbitrary symbols, from iconic and indexical communication. And finally, it must be compatible with the great complexity of vocal language production and comprehension (which in turn enables the great diversity of language: consider that there are over 1500 possible human speech sounds, though of course the phoneme inventories of all known languages each comprise only a portion of these; Evans 2009). Is there an explanation that is up to the job?

In the following section I describe and critique Corballis's recent version of events (Corballis 2017). Corballis is a gestural theorist; indeed, one that sees the importance in delivering a hypothesis about how this stretch of the "Rubicon"—by which he means (in this context) the gap between gesture and speech (see p. 155)—was crossed, evolutionarily speaking.[16] Corballis emphasises three aspects: he argues that speech is more like gesture than other theorists have supposed, he stresses the role of mirror neurons, and he foregrounds the integration of the hand and mouth in eating, connecting the upshot of this with littoral theory (i.e., a version of the aquatic ape hypothesis), to explain increases in volitional vocal production.

Corballis tackles the gesture-speech transition

Corballis points out that characterizing the transition from gesture to speech as one of a shift in communicative modality (i.e., from the visual modality for perceiving gesture, to the auditory modality for perceiving speech) is too simplistic. He argues that gesture and speech are more alike than other theorists have supposed. This seems reasonable. He says:

> …speech is itself a system of gestures, made up of movements of the lips, the velum, the larynx, and the blade, body, and root of the tongue. One might suppose, then, that the production of language shifted from one set of gestures to another (Corballis 2017, p. 148).

Moreover, it even seems as though modern humans typically comprehend speech *as gestures*, rather than as pure auditory patterns. The *phonemes* we perceive in speech do not map one–one onto sonic profiles. Corballis's examples include the *b* sounds in "battle", "bottle", "beer", "bug", "rabbit", "flibbertigibbet". These *b* sounds probably sound much the same to ordinary listeners in ordinary conditions. However they have distinct acoustic profiles. So why do we group these *b* sounds together as a unitary phoneme? They are *produced* similarly by our vocal apparatus. Indeed sometimes we hear the speech we "see", *in spite of* its sonic profile (McGurk

[16] Corballis co-opts Max Müller's famous use of "Rubicon"—the divide between humans and other species with respect to language: 'the one great barrier between brute and man is *Language*. Man speaks, and no brute has ever uttered a word. Language is our Rubicon, and no brute will dare cross it' (Müller 1861, p. 340).

and MacDonald 1976).[17] Of course, it's not that we are completely insensitive to speech sounds *as sounds*: 'both the way we produce speech and the auditory input itself contribute to the way we hear speech' (Corballis 2017, p. 151). Both listening to speech and visual observation of lip movements for speech excite the motor units that underlie speech production (Watkins et al. 2003).

Corballis also emphasises the role of mirror neurons:

> …some mirror neurons in the monkey brain respond to the sounds of actions, such as the tearing of paper or the cracking of nuts. But in monkeys, unlike humans, mirror neurons are deaf to vocalization—they don't respond to the sounds of other monkeys. [But they *do* to gesture.] Only later in primate evolution did the mirror system encompass vocal sounds, enabling us to perceive speech in terms of the way it is produced rather than in terms of how it actually sounds. (Corballis 2017, p. 150, text in square brackets is mine).

Corballis points out that we are the only great apes to do this. But how? Here Corballis reaches for the dual-stream theory of speech perception (Hickok and Poeppel 2007; Hickok 2012) that distinguishes roles played by the dorsal and ventral streams, see Fig. 2. The ventral stream is responsible for understanding speech, and is shared in other animals (consider the ability of trained dogs, apes, and so on, to appropriately respond to distinct human vocalizations). The dorsal stream—part of the mirror system—is responsible for the production of human articulate speech, and although shared in other animals, what is *absent*, at least as far as the other great apes are concerned, is its cooption for vocalization production and learning. (Homologous neural circuitry distinguishes for example the vocal learning

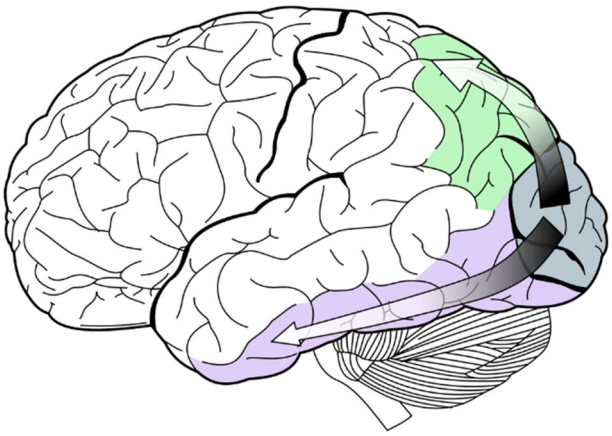

Fig. 2 Dorsal and ventral streams: the upwards arrow indicates the dorsal stream; the downwards arrow indicates the ventral stream. Image by Selket, available at https://commons.wikimedia.org/w/index.php?curid=1679336. Accessed 20 July 2017. Reproducible under the terms of the Creative Commons Attribution-ShareAlike License 3.0

[17] For a short BBC video clip demonstrating the McGurk effect, see https://www.youtube.com/watch?v=G-lN8vWm3m0. Accessed 20 July 2017.

songbirds from other birds; Pfenning et al. 2014.) In humans the dorsal stream maps auditory/sonic representations onto articulatory motor representations, thereby coding and maintaining 'instances of speech sounds, and [using] these sensory traces to guide the tuning of speech gestures so that the sounds are accurately reproduced' (Hickok and Poeppel 2007, p. 399).

Assuming this framework is on the right track—and it seems to be: despite some criticisms,[18] the dual-stream framework has been positively influential and is widely accepted in general (Eysenck and Keane 2015)—what remains to be explained is the cooption of the dorsal stream for vocalization production and learning, presumably after the split from the *Pan-Homo* LCA.

As we will see in the following section, in my view, increasingly intentional vocalization for emotion/affect, intentional listening, vocal grooming, infant-directed crooning, animal call mimicry, and so on (i.e., the mosaic comprising the evolution of "musicality"; Killin 2017) plausibly explains the cooption of the dorsal stream for intentional speech production and learning, in time. The idea: musicality provided both the medium (the voice), for articulate speech to transition from gesture, and the means (incrementally increasing upgrades in volitional control over vocal production apparatus). Corballis, however, takes a rather different approach.

Corballis notes that the hand and mouth are integrated, not only in communication, but in eating: 'More to the point, though, throughout primate evolution the hands and mouth are connected through the process of eating. People and monkeys bring food to the mouth in exquisitely coordinated fashion. Such coupling may well carry over to language' (Corballis 2017, p. 148). Corballis's idea is that the primary focus of (proto)language passed from gesture, to the face (e.g., 'The use of facial expressions as social signals', p. 155), to vocalization (speech being a "swallowed" facial expression, requiring intentional production of sound), following the passage of food from hand to face to inside throat. Many of our facial expressions are reactive/emotional/automatic, of course, but we humans are also capable of fine voluntary control over facial movements (and good actors excel in this). Great apes are capable of it too, though to much less of an extent (i.e., much less flexible and less fine-tuned control). So, the idea is, over the course of evolution, 'voluntary communication probably shifted increasingly from the hand to the face' (p. 155); 'It was but a small step from the external surface of the face to the movable parts inside it' (p. 156). But, of course, we cannot see the intentional movements of parts "inside it", at least not at all clearly enough for effective communication, so sound had to be produced *in order* to stand in as a proxy for a visually perceivable gesture.

Corballis spells out his hypothesis as follows. I quote:

> In primates, hand and mouth are closely linked both in the brain and in behaviour. In the motor cortex of the brain, responsible for initiating body movements, the so-called hand area is adjacent to the mouth area. Some neurons in the frontal lobe of the monkey are activated when the animal makes a grasping movement either with the hand or with the mouth… [There is] a close connection between movements of hand and mouth in people as well. If

[18] E.g., that the two streams don't in fact operate independently of each other, as some dual-stream theorists have supposed (see McIntosh and Schenk 2009).

> human subjects are told to open their mouths while grasping objects with their hands, the size of the mouth opening increases with the size of the grasped object. [Presumably this is tied to preparing the mouth for the food it is about to receive.] If people are asked to say "ba" while grasping an object, or even while watching someone else grasp an object, the syllable itself is affected by the size of the object grasped. The larger the object, the wider the opening of the mouth, with consequent effects on the speech sounds. Even one-year-old babies show these effects [see e.g. Gentilucci 2003]. These links between hand and mouth probably originated in eating rather than communicating… [i.e., in] preparing the mouth to receive an object after the hand has grasped it, but they were adapted for gestural and finally vocal language (Corballis 2017, pp. 157–158, text in square brackets is mine).

Notice that *even if* we buy much of this (though there are reasons to be sceptical: for instance, the hands and face are under somewhat independent cortical control, and we might expect there to be more overlap if Corballis' hand-to-face-to-mouth hypothesis were true), the stretch of the Rubicon between great ape gesture and hominin vocal (proto)language production and learning hasn't yet been successfully crossed. The above facts about the connection between hand and mouth do not suggest increasing voluntary vocal control, but stimulus control: the size of the object *reflexively* influences mouth shape. So Corballis's argument does not yet account for the extension of *voluntary control* progressively in roughly that direction, from hand to face to vocal tract, let alone that *language* transitioned progressively, starting from gesture, adding facial expression, adding voice. So Corballis appeals to littoral theory (a modern variant of the aquatic ape hypothesis; Verhaegen 2013) in order to posit a selection pressure for mouth-related voluntary control—that is, *diving*. Now, the probable bathing, wading, fishing, and shoreline foraging/shellfish gathering of our ancestors aside (see e.g. Shaw-Williams 2017), *maybe* some diving happened that required nonstandard intake and retention of air, but even so, the extent to which this would have had an upshot on voluntary *vocal production control* is extremely unclear. At best, introducing this selection pressure only helps to continue foregrounding greater breath control, not vocal production control. (And Corballis indeed thinks of the aquatic/littoral phase as 'supporting the voluntary control of *vocalization*', p. 163, my emphasis.) I agree that breath control is important to the story. But it alone does not explain the full gamut of vocal language production—including, recall, 'movements of the lips, the velum, the larynx, and the blade, body, and root of the tongue' (p. 148), let alone other evolutionary advances required for greater vocal capacities, such as the enlargement of the hypoglossal nerve (for tongue control), co-evolving vocal and auditory structures (neural and morphological—see Morley 2013), and thoracic vertebrae nerve canal expansion, enhancing intentional control of vocal musculature (MacLarnon and Hewitt 1999). As Kendon notes, 'specializations for *speaking*, in regard to both the production of speech and its reception, are complex and extensive' (Kendon 2016, my emphasis), not adequately accounted for by Corballis's account.

Since Corballis's account doesn't get us to where we want to be, I think that we should shed some of his more shaky theoretical commitments, and consider another plausible in-road. In my view, an evolving vocal musicality helps to explain the transition towards vocal speech from gesture: it prepared hominins cognitively and anatomically for it. I expand on this hypothesis in the following section.

Musicking across the gap

Ancestral hominins communicating with one another presumably vocalized in combination with their intentional, meaningful gestures,[19] for instance to command attention of an individual looking another way, to imbue the gesture with *affect* (perhaps 'playful, flirtatious, affiliative, competitive, or agonistic'—to borrow Kendon 2016's examples), or to add *vocal mimicry* (of a mimed animal, say), and so on. In my view, much of this amounts to the addition of aspects of an evolving musicality.

Elsewhere I have developed a theory of Plio-Pleistocene hominin musicality, conceptualised as a mosaic or "package" of traits: increasing top-down control over affective vocalization, finer breath control (supported by anatomic changes), intentional listening, vocal imitation, turn taking, entrainment, lithic sound play, motherese, call mimicry, vocal grooming, and so on (Killin 2017; see also Killin 2016). I suggested that protomusical group behaviours/activities—group-based expressions of musicality not too dissimilar from some of those of ethnographically known foragers—are to be found in the socio-cultural/cognitive developments occurring, incrementally, during the "Late Acheulean", say from 500 or 400 Kya (Killin 2017). This is based on an argument from hominin socio-cognitive coevolution (big brained, larger group, *social* hominins, with more developed emotional lives) centralized around hearths; the plausibility of these hominins possessing at least a protoaesthetic sensitivity (consider the 500,000 year old finely-crafted handaxes; see Kohn and Mithen 1999); the upgrades in technological production evidenced around this time (e.g., hafting; see Barham 2013); and using the date associated with more common and continual hearths *as social magnets* as a proxy, following Gamble et al. (2011) and Gowlett et al. (2012). Moreover, the ethnographic record details many ways of being musical that are simple enough and would not have left archaeological traces were our ancestors to have realized them or something like them.

Yet prior to this date for the (hypothesised) emergence of social protomusic, the picture of musicality evolution that I favour has consequences for language evolution and the switch to speech. Some of the musicality ingredients are verbal language ingredients too. As call mimicry, motherese, vocal grooming, sound-based play, intentional listening and the like took hold, the anatomical and cognitive preconditions for speech would evolve. Vocalizations, increasing in complexity and

[19] After all, chimps and bonobos vocalize *a lot* while communicating via gesture. For example: 'Kanzi the bonobo vocalizes prolifically while communicating through gesture… but this is probably largely emotional and it is the gestures that provide the information' (Corballis 2017, p. 163).

diversity, would be incorporated into protolanguage *along with* gesture, coming under top-down control, in co-evolutionary tandem. This would be an incremental process, occurring throughout the long stretch of the Pleistocene, resulting in the modifications of the hominin vocal tract that distinguish it from that of the other great apes, specializations for speech.

Our *Hominini* ancestors split from those of *Pan* around 6–7 Mya. It is highly likely that their non-arboreal forager lifestyle (from 4 Mya) would have selected for an increase in vocalization use for conspecific communication and in group defence from predators. Stone tool production and use appears in the archaeological record as far back as 3.3 Mya. It is not a far stretch to think that vocalizations for coordinating group hunting/scavenging, driving predators from kills, and carcass carving would have occurred; these would also have sent signals to nearby bands. Hunting/scavenging and social tracking (Shaw-Williams 2014) would have rehearsed finer attention to sounds, bringing listening under increasing intentional control. The same is true of tool production: attentive, intentional listening is key in both knapping and diagnosing raw material for use. As has been suggested by Robin Dunbar (e.g. Dunbar 1996), emotional/affective vocalizations are likely to have entered the picture, supplementing the manual grooming that maintains social bonds in other great apes (intensifying selection for intentional signal control, listening, and coordination between voices), and potentially playing a role in parent-infant bonding and affecting infant arousal (e.g. "lullaby"-like vocalizations to soothe; "arousal" vocalizations to excite) especially as our female ancestors encountered the obstetrical dilemma and birthed earlier, less developed (and helpless) babies, requiring greater demands in care. Vocalizations may have also played a role in courtship (indeed, in social worlds such as those of the Pleistocene, hominins were presumably constantly scrutinizing one another and vocalizations would have played an inescapable, even if implicit, part in mate selection).

Palaeoanthropological evidence is consistent with this picture of incremental coevolution of vocalization and audition. From around 2 Mya onwards, *Homo* began to develop neural and anatomical changes presumably in tight lock-step: enlargement of the hypoglossal nerve (enabling finer tongue control), coevolving vocal and auditory structures (neural and anatomical), and in time (at least by *heidelbergensis*), expansion of the thoracic vertebrae nerve canal (enabling greater intentional control of breath and vocal musculature). For review, see Morley (2013).

At some point, as hominin life stages evolved throughout the Pleistocene, infant babbling began—a deeply entrenched, universal behaviour which allows the infant to rehearse finer control over vocalization production through play-like behaviour (Merker 2012), leading to improved intentional control over vocal musculature in adults. Other primate infants don't babble; among the great apes it is a hominin adaptation.

The various features indicated above emphasise an increasing phonological complexity. For example, call mimicry is a crucial upgrade to the skill set of hunters and foragers, and would have been an effective addition to gestural/pantomimic communication of hunting plans (e.g. Bickerton 2009), as well as in status-securing, pedagogical, or "blowing off steam" hunting stories. Mid-Pleistocene hominins did not merely live in the here-and-now: they were highly proficient medium- to large-

game hunters engaging in coordinated, cooperative, planned activities (Bunn and Pickering 2010a, b).

Finally, the production methods of the Late Acheulean industry reveal salient cognitive advances taking place: greater imitation, shared intentionality, greater episodic memory, mental templates, attentive focus, impulse control, greater intentional motor control, and importantly for language evolution, upgrades in communication and social learning and greater intentional listening (Killin 2017). As our ancestors' cognitive abilities developed, they would have become better at distinguishing similar vocal sounds and comprehending/understanding the intentional vocalizations of conspecifics. And from 800,000 years ago, an incremental, though momentous brain size increase took place. More effective, *multi-modal* protolinguistic communication, enabling better foraging and control of fire, teaching and learning, affective and social communication, and cooperation and coordination in general, are almost certainly among the factors driving this final surge in hominin encephalization. At least by *heidelbergensis*—the ancestor of Neanderthals and modern *sapiens*—the palaeoanthrological evidence is suggestive of near-modern vocal control and the morphological means to produce a near-modern vocal repertoire of sounds (Morley 2013; Dunbar 2014), which would continue to evolve and come under greater intentional control throughout the long passage to anatomical and cognitive modernity. With such anatomical and cognitive changes occurring from *ergaster/erectus* to *heidelbergensis* to *sapiens*—changes in the domain of affective vocalization, and eventually the emergence of group protomusic—the transition from predominantly gestural to verbal protolanguage is a predictable consequence, given the considerations listed earier.

Towards arbitrariness: the echo phonology hypothesis

It might be objected at this point that I have not yet provided an adequate explanation of the transition from gesture to speech, since it is hard to see how it provides a mechanism for a switch from a largely iconic and deictic/indexical form of communication to one in which mappings from symbol to meaning are largely *arbitrary*. In the context of the present dialectic, however, notice that Corballis's view does not either; in fact, Corballis's approach is to down-play the extent to which language is truly "arbitrary". (Indeed the extent to which spoken words resemble their referents has been debated at least since Plato's *Cratylus*.) However, Corballis's move here is hardly convincing, in my view. Even if language is less arbitrary than some theorists have supposed,[20] it is nonetheless widespread enough to be an explanandum requiring attention.

Admittedly, the objection that I am entertaining here may be overstated: it has been claimed that once language has gone largely vocal, it automatically goes largely arbitrary as a byproduct of this shift (e.g. Sterelny 2018); that is, there isn't two problems—the shift to vocal dominance and the shift to arbitrary sign dominance—just the first, the shift to vocal dominance. And in any case, iconic and

[20] Saussure, for one, famously thought of arbitrariness as a defining feature of language.

indexical gestures can become conventionalized and thus "arbitrarified" through familiar processes, drifting to the arbitrary as a result (Tomasello 2008). I am sympathetic to this rendering of the matter, however in this section I will consider more seriously those who require more to be said on the uptake of arbitrariness and the gesture–vocal transition in particular. Kendon (2016) puts it thusly: 'Even if we can suggest factors that might have contributed to the elaboration of complexity in vocal (and gestural) expression… the issue of how these gestures acquired symbolic significance still eludes us'. Settling this matter persuasively is not possible in the remaining space, and I suspect that there are a range of mechanics that tend to push symbol systems towards arbitrariness (Gasser 2004).[21] I nod towards one plausible answer: the *echo phonology* evolutionary hypothesis (Woll 2014).[22]

Echo phonology (Woll 2001) is the name given to the class of mouth gestures that obligatorily accompany signed gesture in contemporary sign languages. The mouth gesture is a motoric and visual "echo" of the hand/arm gesture, hence the term "echo phonology" (the mouth gesture is typically not voiced, and is not related to, or derived from, vocal movements for speech production). Consider:

> In the BSL sign *true*… the upper hand moves downwards to contact the lower hand, and this action is accompanied by mouth closure, synchronized with the hand contact… the mouth gesture forms part of the citation form of the manual sign [and does not carry additional meaning; e.g., it does not function to distinguish between gestural homonyms, or add an adverb to a signed action, etc.]… Signs with echo phonology appear incomplete or ill-formed in their citation form if the mouth gesture is not present (Woll 2014, p. 4, text in square brackets is mine; also see Fig. 3).

Neural studies indicate that network activation underlying the processing of signs with echo phonology lies somewhere in-between that of manual-sign-only processing and processing of signs accompanied by mouthings that function to disambiguate homonymous signs (which more closely resembles that of the processing of lips in proficient speechreaders) (Capek et al. 2008; for further discussion see Woll 2014).[23] What is impressive about Capek and colleagues' study is the internal consistency found across the various cases: in each case, 'the more active region was that which was more involved in processing hand movements than mouth movements' (Capek et al. 2008, pp. 1231–1232). Woll connects these findings with a plausible evolutionary hypothesis.

[21] I thank an anonymous referee for pushing this point.

[22] I am grateful to Lauren Reed for bringing this hypothesis to my attention.

[23] Capek et al. find that 'speech-derived mouthings (DM [disambiguating mouth signs]) generated relatively greater activation in a somewhat circumscribed region of the left middle and posterior portions of the superior temporal cortex, whereas for mouth gestures (EP [echo phonology signs]), which are not speech-derived, there was relatively greater posterior activation in both hemispheres' (Capek et al. 2008, p. 1231). This suggests that in EP the manual gesture drives the accompanying mouth gesture, providing a plausible cortical correlate. They continue: 'Although mouth actions can be of many different sorts, DM and EP show systematic differences in terms of their functional cortical correlates; DM resembles speechreading more closely, whereas EP resembles manual-only signs' (Capek et al. 2008, p. 1231).

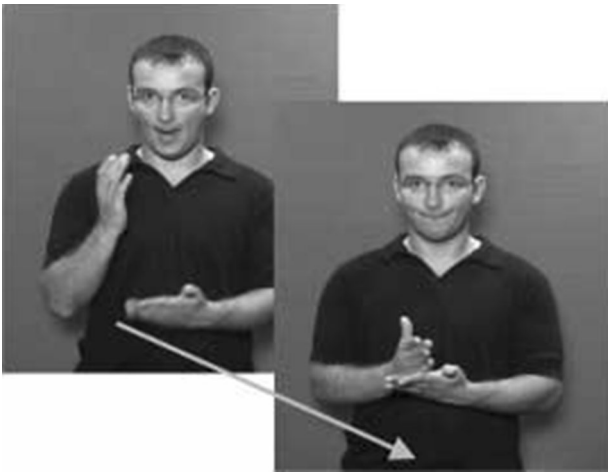

Fig. 3 Echo phonology: BSL *true*. Extracted from Capek et al. (2008), reproduced here with permission of MIT Press

Although the mouth gestures of echo phonology are usually not voiced by deaf signers, hearing signers sometimes add an audible vocal component. In these hearing bilinguals (e.g. those fluent in English and BSL), in contexts in which codes are blended, the gestural component of a sign may be dropped, and only the mouth component provided, with or without audible voicing (Woll 2014). According to Woll's evolutionary hypothesis, echo phonology points to a possible means for leaping from 'a situation where voicing accompanies these mouth gestures so that they begin to have independent existence as lexical items… Echo phonology illustrates a mechanism by which abstract concepts, which can be represented by iconic manual gestures, can be attached to abstract mouth gestures' (Woll 2014, pp. 5–6). The idea: an expression that our ancestors may have iconically gestured or pantomimed may have been (involuntarily or otherwise) accompanied by the mouth simultaneously performing the action of the hands/arms (the link between the hand and mouth in the mirror system presumably having some role in this). Here's a hypothetical pantomimic/iconic example (indulge me): outstretched forward-pointing arms snapping into one another, like a crocodile's snout, accompanied by a (voluntary or involuntary) snapping (voiced or silent) of the teeth.[24] And once familiar, the hand/arm gesture could be omitted (perhaps the arms are busy pointing in some direction or miming out a route or coordination plan) with the mouth gesture/voicing "taking over" for crocodile. In my view: an evolving musicality provides the impetus for such a transition in general and several plausible upshots (e.g., expanding the possibility space of producible phonemes, providing the physical means for intentional voicings in the absence of the associated manual

[24] This hypothetical example shouldn't seem too wild. There is evidence that ancient hominins were procuring aquatic prey including crocodiles, presumably a feat requiring coordinated action, 1.95 million years ago (Braun et al. 2010).

gesture, enabling coevolution with the hominin auditory channel to better align hearing ranges and vocal ranges).

Of course, the hypothesis is presented here somewhat tentatively. Further investigation into echo phonology and its import to the evolution of language debate is a priority for future research. One key question to investigate is why it happens. Some aspects of sign language may be present due to the fact that signers so often have to communicate with hearers, even if (some of) those hearers are also signers, so it would be worth knowing how variable and explicit the presence of echo phonology is across signers who mostly communicate with other deaf signers and those who do not.[25] Another, for example, is to gain a sense of how ancient echo phonology is, including whether and to what extent it accompanies great ape intentional gesture, and to gain a sense of how universal it is across human peoples, including whether and to what extent it appears in homesign systems. Needless to say, Woll is optimistic. I'll give her the final words of this section:

> One issue for those concerned with suggesting a link between gesture and word has always been how the arbitrary symbol-referent relationship of words in spoken language could have come from visually-motivated gestures. Echo phonology provides evidence for a possible mechanism. Firstly, the phenomenon appears to be fairly common across different sign languages… Secondly, the mouth actions found in echo phonology are themselves non-visually motivated… Thirdly, the actual inventory of elements in echo phonology looks very much like a system of maximal contrasts in a spoken language phonology…Fourthly, functional imaging research on the representation of signs and words in the brain suggests that echo phonology occupies an interesting intermediate position. (Woll 2014, p. 8).

Summary

I hope to have convinced you that the origins of language, as we know it today, are largely in the gestural domain, continuous with the intentionally communicative behaviours of the great apes. I have evaluated one recent account, put forward by an influential gestural theorist, of how gestural protolanguage might have transitioned to speech. I found that account wanting, and I argued that an independently evolving musicality played a key role in preparing ancient hominins for vocal language. Further research is required to develop this idea and generate testable hypotheses. I hope to have convinced you of a "proof of concept", nonetheless. (At least, it fares better against the set of constraints identified than Corballis's account does). I suggested that further research on echo phonology may shed light on a possible mechanism for the transition towards arbitrariness in speech.

Acknowledgements Thanks are due to Simon Greenhill and Kim Sterelny for compiling this special issue. I am grateful to Michael Corballis, Kim Sterelny, and an anonymous referee for helpful critical comments on previous versions of the manuscript. I thank audiences at the Empirical Philosophy

[25] Thanks to an anonymous referee for pushing this point.

Workshop 2017 at Victoria University of Wellington, the Centre of Excellence for the Dynamics of Language Seminar Series at the Australian National University, and the Australasian Association of Philosophy Conference 2017 at the University of Adelaide for helpful questions and feedback on this and related material. In particular I thank Matt Spike, Kim Shaw-Williams, and Lauren Reed.

References

Barham L (2013) From hand to handle: the first industrial revolution. Oxford University Press, Oxford
Begby E (2017) Language from the ground up: a study of homesign communication. Erkenntnis 82:693–714
Behne T, Carpenter M, Tomasello M (2014) Young children create iconic gestures to inform others. Dev Psychol 50(8):2049–2060
Bickerton D (1990) Language and species. University of Chicago Press, Chicago
Bickerton D (2009) Adam's tongue: how humans made language, how language made humans. Hill and Wang, New York
Boesch C (1993) Aspects of transmission of tool-use in wild chimpanzees. In: Gibson KR, Ingold T (eds) Tools, language, and cognition in human evolution. Cambridge University Press, Cambridge, pp 171–183
Boesch C, Boesch-Achermann H (2000) The chimpanzees of the Taï forest: behavioural ecology and evolution. Oxford University Press, Oxford
Braun DR, Harris JWK, Levin NE et al (2010) Early hominin diet included diverse terrestrial and aquatic animals 1.95 Ma in East Turkana, Kenya. Proc Natl Acad Sci 107(22):10002–10007
Brown S (2000) The "musilanguage" model of music evolution". In: Wallin NL, Merker B, Brown S (eds) The origins of music. MIT Press, Cambridge, pp 271–300
Bunn HT, Pickering TR (2010a) Methodological recommendations for ungulate mortality analyses in paleoanthropology. Quatern Res 74(3):388–394
Bunn HT, Pickering TR (2010b) Bovid mortality profiles in paleoecological context falsify hypotheses of endurance running–hunting and passive scavenging by early Pleistocene hominins. Quatern Res 74(3):395–404
Burling R (2005) The talking ape. Oxford University Press, New York
Capek CM et al (2008) Hand and mouth: cortical correlates of lexical processing in British Sign Language and speechreading English. J Cogn Neurosci 20(7):1220–1234
Corballis M (2002) From hand to mouth: the origins of language. Princeton University Press, Princeton
Corballis M (2009) The evolution of language. Ann N Y Acad Sci 1156:19–43
Corballis M (2017) The truth about language: what it is and where it came from. Auckland University Press, Auckland
Crockford C, Wittig R, Mundy R, Zuberbühler K (2012) Wild chimpanzees inform ignorant group members of danger. Curr Biol 22(2):142–146
Darwin C (1871) The descent of man, and selection in relation to sex. Appleton & Co., New York
Dunbar R (1996) Grooming, gossip and the evolution of language. Harvard University Press, Cambridge
Dunbar R (2014) Human evolution. Penguin, London
Esteve-Gibert N, Prieto P (2014) Infants temporally coordinate gesture–speech combinations before they produce their first words. Speech Commun 57:301–316
Evans N (2009) Dying words: endangered languages and what they have to tell us. Wiley-Blackwell, Oxford
Eysenck M, Keane M (2015) Cognitive psychology: a student's handbook, 7th edn. Psychology Press, New York
Fitch W Tecumseh (2010) The evolution of language. Cambridge University Press, Cambridge
Fitch WT, Zuberbühler K (2013) Primate precursors to human language: beyond discontinuity. In: Altenmüller E, Schmidt S, Zimmerman E (eds) Evolution of emotional communication. Oxford University Press, Oxford, pp 26–48
Gallese V, Fadiga L, Fogassi L, Rizzolatti G (1996) Action recognition in the premotor cortex. Brain 119(2):593–609
Gamble C, Gowlett J, Dunbar R (2011) The social brain and the shape of the Palaeolithic. Camb Archaeol J 21(1):115–135
Gasser M (2004) The origins of arbitrariness in language. Proc Annu Meeting Cogn Sci Soc 26:434–439

Geissmann T (2002) Duet-splitting and the evolution of gibbon songs. Biol Rev 77:57–76
Gentilucci M (2003) Grasp observation influences speech production. Eur J Neurosci 17(1):179–184
Goldin-Meadow S (2003) The resilience of language: what gesture creation in deaf children can tell us about how all children learn language. Psychology Press, New York
Goldin-Meadow S, Alibali MW (2013) Gesture's role in speaking, learning, and creating language. Annu Rev Psychol 64:257–283
Goodall J (1986) The chimpanzees of gombe: patterns of behavior. Harvard University Press, Cambridge
Gowlett J, Gamble C, Dunbar R (2012) Human evolution and the archaeology of the social brain. Curr Anthropol 53(6):693–722
Hewes GW (1973) Primate communication and the gestural origin of language. Curr Anthropol 14:5–24
Hickok G (2012) The cortical organization of speech processing: feedback control and predictive coding the context of a dual-stream model. J Commun Disord 45:393–402
Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8(5):393–402
Hobaiter C, Byrne RW (2011) Serial gesturing by wild chimpanzees: its nature and function for communication. Anim Cogn 14:827–838
Hobaiter C, Byrne RW (2014) The meaning of chimpanzee gestures. Curr Biol 24:1–5
Hostetter AB, Cantero M, Hopkins WD (2001) Differential use of vocal and gestural communication by chimpanzees (*Pan troglodytes*) in response to the attentional status of a human (*Homo sapiens*). J Comp Psychol 115(4):337–343
Hurford J (2014) Origins of language. Oxford University Press, Oxford
Irvine E (2016) Method and evidence: gesture and iconicity in the evolution of language. Mind Lang 31(2):221–247
Kalagher H, Yu C (2006) The effects of deictic pointing in word learning. In: Proceedings of the 5th international conference of development and learning. Indiana University Department of Psychological Brain Sciences, Bloomington. ISBN 0-9786456-0-X
Kelly SD, Özyürek A, Maris E (2010) Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. Psychol Sci 21(2):260–267
Kendon A (2016) Reflections on the 'gesture-first' hypothesis of language origins. Psychon Bull Rev. https://doi.org/10.3758/s13423-016-1117-3
Killin A (2016) Rethinking music's status as adaptation versus technology: a niche construction perspective. Ethnomusicol Forum 25(2):210–233
Killin A (2017) Plio-Pleistocene foundations of hominin musicality: coevolution of cognition, sociality, and music. Biol Theory 12(4):222–235
Kimura D (1993) Neuromotor mechanisms in human communication. Oxford University Press, Oxford
Kohler E et al (2002) Hearing sounds, understanding actions: action representation in mirror neurons. Science 297:846–848
Kohn M, Mithen S (1999) Handaxes: products of sexual selection. Antiquity 73:518–526
Lawson FRS (2014) Is music an adaptation or a technology? Ethnomusicological perspectives from the analysis of Chinese *shuochang*. Ethnomusicol Forum 23(1):3–26
MacLarnon A, Hewitt G (1999) The evolution of human speech: the role of enhanced breathing control. Am J Phys Anthropol 109:341–363
Masataka N (1992) Motherese in a signed language. Infant Behav Dev 15(4):453–460
McGurk H, MacDonald, J (1976) Hearing lips and seeing voices. Nature 264(5588):746–748
McIntosh RD, Schenk T (2009) Two visual streams for perception and action: current trends. Neuropsychologia 47(6):1391–1396
Merker B (2012) The vocal learning constellation: imitation, ritual culture, encephalization. In: Bannan N (ed) Music, language and human evolution. Oxford University Press, Oxford, pp 215–260
Mithen S (2005) The singing Neanderthals. Weidenfeld & Nicolson, Great Britain
Moore R (this issue) Social cognition, Stag Hunts, and the evolution of language. Biol Philos. https://doi.org/10.1007/s10539-017-9598-7
Morley I (2013) The prehistory of music: human evolution, archaeology, and the origins of musicality. Oxford University Press, Oxford
Müller M (1861) Lectures on the science of language (first series). Longman, Green, Longman and Roberts, London
Newman A et al (2002) A critical period for right hemisphere recruitment in American Sign Language processing. Nat Neurosci 5:76–80

Newman A et al (2010) Prosodic and narrative processing in American Sign Language: an fMRI study. Neuroimage 52:669–676
Novick JM, Trueswell JC, Thompson-Schill SL (2010) Broca's area and language processing: evidence for the cognitive control connection. Lang Linguist Compass 4(10):906–924
O'Neill M, Bard K, Linnell M, Fluck M (2005) Maternal gestures with 20-month-old infants in two contexts. Dev Sci 8(4):352–359
Perniss P, Vigliocco G (2014) The bridge of iconicity: from a world of experience to the experience of language. Philos Trans R Soc B 369:20130300. https://doi.org/10.1098/rstb.2013.0300
Petitto LA, Marentette PF (1991) Babbling in the manual mode: evidence for the ontogeny of language. Science 251:1493–1496
Pfenning AR et al (2014) Convergent transcriptional specializations in the brains of humans and song-learning birds. Science 346:1333–1346
Pika S, Mitani J (2006) Referential gestural communication in wild chimpanzees (*Pan troglodytes*). Curr Biol 16(6):R191–R192
Pollick A, de Waal F (2007) Ape gestures and language evolution. Proc Natl Acad Sci 104:8184–8189
Rizzolatti G, Arbib MA (1998) Language within our grasp. Trends Neurosci 21(5):188–194
Rizzolatti G, Craighero L (2004) The mirror-neuron system. Annu Rev Neurosci 27:169–192
Rizzolatti G et al (1988) Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. Exp Brain Res 71:491–507
Russon AE, Andrews K (2011) Pantomime in great apes: evidence and implications. Commun Integr Biol 4(3):315–317
Schel AM, Townsend SW, Machanda Z, Zuberbühler K, Slocombe KE (2013) Chimpanzee alarm call production meets key criteria for intentionality. PLoS ONE 8(10):e76674
Senghas A (1995) The development of Nicaraguan sign language via the language acquisition process. In: MacLaughlin D, McEwan S (eds) BUCLD 19: proceedings of the 19th annual Boston University conference on language development. Cascadilla Press, Boston
Senghas A, Kita S, Özyürek A (2004) Children creating core properties of language: evidence from an emerging sign language in Nicaragua. Science 305:1779–1782
Shaw-Williams K (2014) The social trackways theory of the evolution of cognition. Biol Theory 9(1):16–26
Shaw-Williams K (2017) The social trackways theory of the evolution of language. Biol Theory 12(4):195–210
Slocombe KE, Zuberbühler K (2007) Chimpanzees modify recruitment screams as a function of audience composition. Proc Natl Acad Sci USA 104:17228–17233
Sterelny K (2012) Language, gesture, skill: the coevolutionary foundations of language. Philos Trans R Soc B 367:2141–2151
Sterelny K (2016) Deacon's challenge: from calls to words. Topoi 35(1):271–282
Sterelny K (2018) Language: from how-possibly to how-probably? In: Joyce R (ed) Routledge handbook of evolution and philosophy. Routledge, Oxon/NY, pp 120–135
Sterelny K (this issue) From code to speaker meaning. Biol Philos. https://doi.org/10.1007/s10539-017-9597-8
Számadó S, Szathmáry E (2006) Selective scenarios for the emergence of natural language. Trends Ecol Evol 21(10):555–561
Thompson RL, Vinson DP, Woll B, Vigliocco G (2012) The road to language learning is iconic: evidence from British Sign Language. Psychol Sci 23(12):1443–1448
Tomasello M (2008) Origins of human communication. MIT Press, Cambridge
Verhaegen M (2013) The aquatic ape evolves: common misconceptions and unproven assumptions about the so-called aquatic ape hypothesis. Hum Evol 28:237–266
Watkins KE, Strafella AP, Paus T (2003) Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia 41(8):989–994
Watson SK, Townsend SW, Schel AM, Wilke C, Wallace EK, Chang L, West V, Slocombe KE (2015) Vocal learning in the functionally referential food grunts of chimpanzees. Curr Biol 25:495–499
Woll B (2001) The sign that dares to speak its name: echo phonology in British Sign Language (BSL). In: Boyes Braem P, Sutton-Spence R (eds) The hands are the head of the mouth: the mouth as articulator in sign language. Signum, Hamburg, pp 87–98
Woll B (2014) Moving from hand to mouth: echo phonology and the origins of language. Front Psychol 5:662. https://doi.org/10.3389/fpsyg.2014.00662