

Human Origins and Language Origins 2: From Actions via Pantomime to Language and Narrative

MICHAEL A. ARBIB*

University Professor Emeritus, University of Southern California, Los Angeles, California, USA; ORCID [0000-0002-6902-3580](https://orcid.org/0000-0002-6902-3580); <https://carta.anthropogeny.org/users/michael-arbib>; arbib@usc.edu

FRANCESCO D'ERRICO

Université de Bordeaux, PACEA UMR 5199, CNRS, Pessac, FRANCE; and, Centre for Early Sapiens Behaviour (SapienCE), Department of Archaeology, History, Cultural Studies and Religion, University of Bergen, Bergen, NORWAY. ORCID [0000-0002-2422-3079](https://orcid.org/0000-0002-2422-3079); <http://www.pacea.u-bordeaux.fr/d-ERRICO-Francesco,487.html>; francesco.derrico@u-bordeaux.fr

PETER GÄRDENFORS

Cognitive Science, Department of Philosophy and Cognitive Science, Lund University, Lund, SWEDEN; and Paleo-Research Institute, University of Johannesburg, Johannesburg, SOUTH AFRICA; ORCID [0000-0001-7423-828X](https://orcid.org/0000-0001-7423-828X); <https://portal.research.lu.se/en/persons/peter-gardenfors>; peter.gardenfors@lucs.lu.se

BERND HEINE

Institute of African Studies and Egyptology, Universität zu Köln, GERMANY; ORCID [0000-0002-9304-7814](https://orcid.org/0000-0002-9304-7814); www.prof-bernd-heine.de; Bernd.Heine@uni-koeln.de

TANIA KUTEVA

Department of English Language and Linguistics, Heinrich-Heine-Universität Düsseldorf, GERMANY; ORCID [0000-0002-4400-7277](https://orcid.org/0000-0002-4400-7277); <https://www.anglistik3.hhu.de/team/detailseite-kuteva>; kuteva@hhu.de

RUKMINI BHAYA NAIR

Department of Humanities and Social Sciences, Indian Institute of Technology, Delhi, INDIA; and, School of the Arts and the Institute for Humanities and Social Sciences, Queen Mary University of London, UNITED KINGDOM; ORCID [0000-0002-2185-1651](https://orcid.org/0000-0002-2185-1651); <https://hss.iitd.ac.in/faculty/rukmini-bhaya-nair>; <https://www.qmul.ac.uk/ihss/people/global-professorial-fellows/>; rukmini.nair@gmail.com

CHRIS SINHA

School of Psychology, University of East Anglia, UNITED KINGDOM; ORCID [0000-0002-7306-2400](https://orcid.org/0000-0002-7306-2400); Christopher Sinha - University of East Anglia; c.sinha@uea.ac.uk

*corresponding author: Michael A. Arbib; arbib@usc.edu

submitted: 8 April 2025; revised 15 May 2026; accepted: 25 May 2026

Handling Editor in Chief: Erella Hovers

ABSTRACT

The paper explores stages of (bio)cultural evolution particularly associated with language origins, adopting a co-evolutionary approach (EvoDevoSocio) to human origins that relates changes in the human organism to changes in the communicative, material, and symbolic niches created by and sustaining human societies. Rather than proposing a specific place or time for the origin of language, we advocate a “polycentric” approach, criticizing saltationist theories. Social communication of new skills, we claim, was supported by pantomime for teaching and social coordination, which in turn supported the development of manual and vocal gestures that crystallized into protolanguages and, in turn, into languages. Narrative and conversation not only played a crucial role in establishing social cohesion, transmitting knowledge, and sharing experiences and values, but also were major drivers of language change, with grammaticalization as a powerful engine for such change. We explore the intertwined roles of narrative and empathy in building human groups and societies, with notes on the emergence of drawing and writing. Based on the archaeological record of material and symbolic culture in the companion paper, “Human Origins and Language Origins 1” (HOLO 1), we offer an evolutionary perspective on human cognitive architecture and propose time ranges for the polycentric emergence of protolanguages and evolutionarily modern languages.

1. INTRODUCTION: THE POLYCENTRIC EVOLUTION OF HUMAN BIOCULTURAL NICHES

This paper is HOLO 2, a companion to HOLO 1, “Human Origins and Language Origins 1: a polycentric scenario for biocultural niche expansion” (d’Errico et al. 2027 [this issue]). Based in archaeology and paleoanthropology, HOLO 1 used evolutionary and developmental biology and sign theory to identify major transitions from *Australopithecus* to the diverse patterns of evolution of *Homo* populations. It then explored the emergence of human biocultural niches supporting developmental processes grounded in learning socially constructed patterns of praxis and communication, introducing two key theoretical concepts—the *semiosphere* (after Lotman 1990) of language and its diverse precursors and complements; and the *technosphere* of material artifacts and constructive praxis, including human-wrought changes to the physical environment, such as agriculture and buildings. HOLO 2 exploits comparative primatology, action analysis, comparative linguistics, and narratology to chart language origins. Communication is necessary for the survival of social creatures and therefore represents a significant selection pressure, but language is a special form of communication whose particularities must be emphasized. It depends on a shared understanding of varied situations, an ability to represent an *open-ended* set of objects and actions in the environment, and the ability to use symbols in ways that allow for mutual adjustment and cooperation among members of a community. We thus emphasize that changes in the semiosphere co-evolved with changes in the technosphere. §1.1 summarizes how the papers complement each other and will also help those who wish to read HOLO 2 without having read HOLO 1.

The evolution of the biocultural niche of language (Sinha 2009) depended upon many other co-evolving dimensions. For example, Dunbar (1993) hypothesizes that social grooming evolved into the cohesive use of language for communicating gossip. We do not discount this, but emphasize the role of language in teaching and in social coordination, with a concern for its role in conversation and narrative. The scope of the *niches* we study can vary. One broadly defined niche is the language niche; another is the developmental niche anchored in the extended developmental period of childhood and the role of caregiving. However, in tracing human origins, we may focus on a more specific material and symbolic niche to explore its potential for biocultural evolution. Where much evolutionary discussion of niche construction focuses on competition and cooperation between different species, our concern with *biocultural niche evolution* (Laland et al. 2000) will be with niches socially constructed through material and symbolic change (*technosphere* and *semiosphere*) as a complement to phenogenotypic selection in populations of humans and their predecessors. We view language and culture as integral to human ecology because human organisms continuously transform their environmental niche through linguistic and cultural practices as well as material actions (Sinha 2009, 2015a).

1.1. HOLO 1 AND HOLO 2 IN RELATIONSHIP

Figure 1 provides an overview of the human and language origins charted by HOLO 1 and 2. HOLO 1 reviews archaeological evidence for episodes of niche expansion and their relationship, supporting inferences concerning the practical and communicative processes behind the production and use of artifacts. These studies also explore the extent to which culturally developed niches exert adaptive pressures that influence further natural selection and cultural evolution. The biological evolution of *Homo* that increased both *altriciality* (being born in an immature, helpless state and requiring extended parental care) and *neoteny* opened (and helped open) the increased dependence of group members on behaviors that emerge through cultural evolution. HOLO 1 not only enhances understanding of past human behaviors and cognitive abilities but also illustrates the dynamic interplay between cultural and biological evolution in shaping human history.

HOLO 1 reviewed the archaeological record from the technosphere for evidence of the emergence of an early human semiosphere entrenched in material signs. Some semiotic innovations were made by hominins well before the emergence of *Homo sapiens*, while others seem to have been made independently by *H. sapiens* and Neanderthals, with some interchange between these and other *Homo* fossil species. We do not treat language as a separable “add-on” to other capabilities but instead chart a path where developments in physical and communicative skills are interwoven. In HOLO 2, we posit a crucial role for pantomime not only in coordinating human actions but also, for example, in teaching the making of stone tools even before the use of pantomime was conventionalized into protosigns and enriched by protospeech. Without discounting the importance of speech (see §1.2), we argue that *pantomime provided powerful interaction between technosphere and semiosphere*.

We make a crucial distinction between protolanguages and languages. For us, a *protolanguage* is a potential precursor to language with only a limited range of conventionalized “protowords” (the basic units of meaning in a protolanguage) whether signed or spoken, and at most limited means to combine them (Arbib and Bickerton 2010)¹, whereas *languages* have richer lexicons and grammars with many *constructions* (in the linguist’s sense) for combining words and phrases, possibly hierarchically. Such constructions not only assemble the forms of words, but combine the meanings of constituents to create new meanings (Croft 2001; Fillmore and Kay 1993; Goldberg 1995).

Both HOLO papers are informed by an EvoDevoSocio framework (Arbib et al. 2018; Dor and Jablonka 2014), in which biocultural evolution yields organisms that develop according to the physical and social environment but which can in turn change that environment by their behavior and the social systems that construct it, extending the EvoDevo framework (Jablonka and Lamb 2005; Müller 2007). Crucially, humans are constrained by their current biocultural niche and yet can actively reconfigure that niche by a process of cultural evolution that has been increasingly decoupled from gene-changing processes of biological evolution.

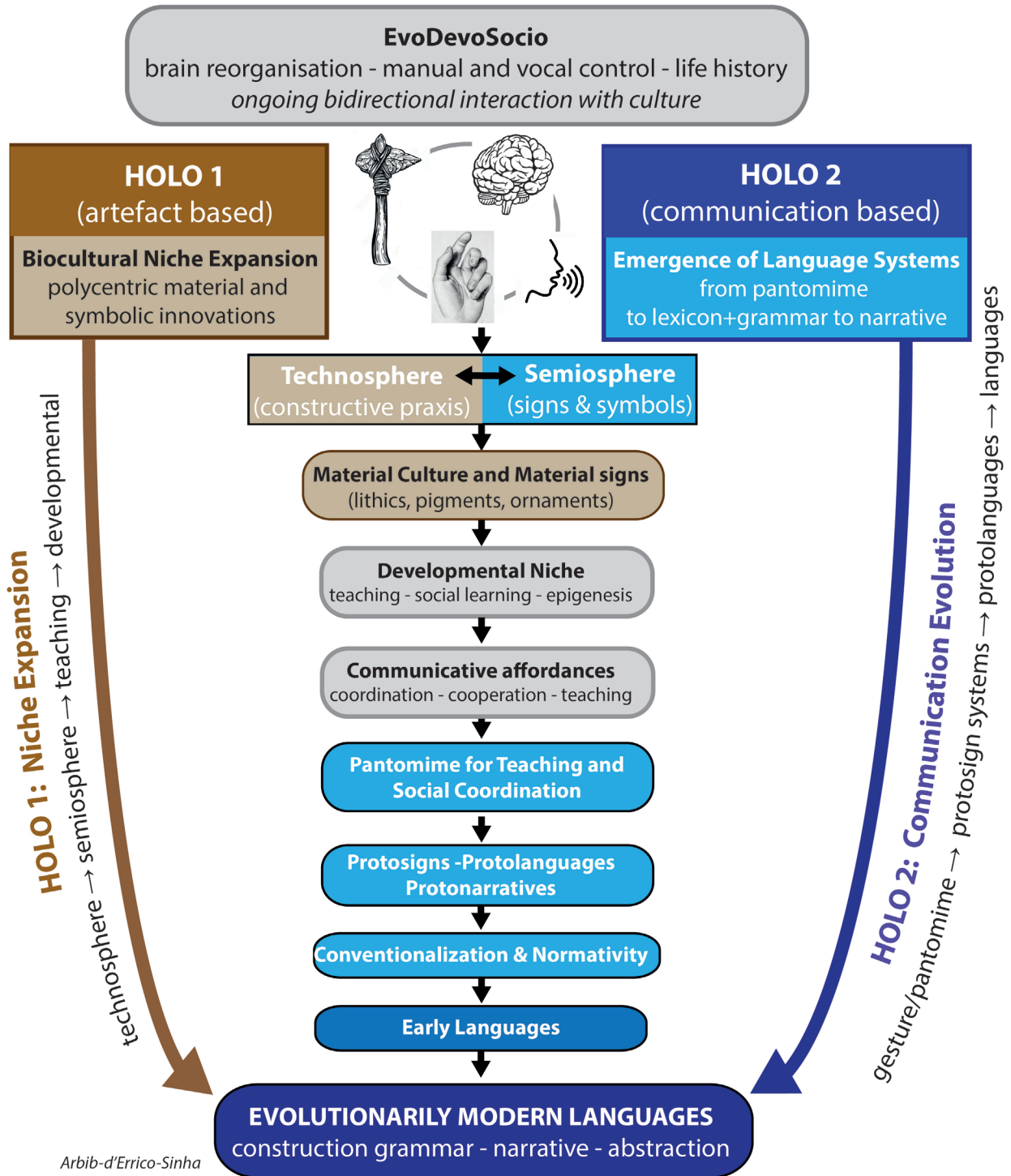


Figure 1. Integrated biocultural model linking HOLO 1 and HOLO 2. HOLO 1 documents the polycentric expansion of human biocultural niches through the co-evolution of material practices (technosphere) and symbolic systems (semiosphere), mediated by ontogenetic development in evolving social contexts (EvoDevoSocio). Increasingly complex systems of communication, described in HOLO 2, involved the successive emergence of pantomime, protosigns, protolanguages, early languages, and ultimately evolutionarily modern languages (see Table 1). Throughout this trajectory, biological evolution (e.g., brain reorganization, manual and vocal control, life history changes) interacted dynamically with cultural processes, while cultural evolution progressively gained relative autonomy permitting accelerated cultural diversification.

Processes important in cultural transmission are more widespread in other species than earlier recognized, suggesting possible continuities between animal and human culture (Whiten et al. 2011). Nonetheless, given the differences between the communication systems of non-human primates and languages, we emphasize the diversification of human cultures, together with the impact of cultural evolution in the predispositions of human minds for cultural transmission. Morgan and Feldman (2024) argue that human culture is uniquely open-ended rather than uniquely cumulative. Rendell et al. (2011) advert to our theme of imitation (§2) by analyzing how copying affects the persistence of cultural knowledge, while d’Errico and Stringer (2011) weighed the merits of evolution, revolution, and saltation scenarios for the emergence of modern cultures.

HOLO 2 outlines the evolutionary paths that led to language. It is more speculative than HOLO 1, but many aspects are grounded in diverse studies in comparative primatology, human development, and historical linguistics, among others, while other aspects have begun to attract empirical testing. A core hypothesis is that biocultural evolution of humans (in the broad sense) yielded brains, bodies, and social interactions that were *language-ready* (Arbib 2002) long before cultural evolution yielded human communities using expressively open-ended languages. By an X-ready species, we mean one that had the ability to eventually develop, primarily through cultural evolution, the capability to use X. For example, early *Homo sapiens* were not only language-ready but were also culture-ready (Whitehead 2010), city-ready, drawing-ready, and “reading-ready”—Dehaene et al. (2010) found that an area serving face recognition becomes adapted for the recognition of visual word forms through experience of reading, not through genetic change—supporting the notion of *cultural recycling of brain regions* to serve purposes that supplied none of the adaptive pressures in their biological evolution that produced them (see also M.L. Anderson 2010; Colagè and d’Errico 2020)². Similarly, the use of tally marks or symbolic notations for simple record-keeping could have been exapted, driven by changing social and environmental contexts, into more sophisticated systems of writing and numerical representation.

In his book *The Symbolic Species*, Deacon (1997), following Saussure (1916), posits that a symbol’s meaning depends *paradigmatically* on how it contrasts with alternative possible symbols and *syntagmatically* on its patterns of cooccurrence and ordering with other symbols across time. This seems not to occur at the level of basic pantomime (§2). There, a pantomime succeeds precisely because it evokes some aspect of a practical behavior whose relevance may be apparent to the observer because of the context in which it occurs rather than through its relation to other pantomimes. However, as protolanguages emerge, conventions become adopted that begin to modify protosigns based on socially emergent contrasts. Thus, where Deacon asserts that symbols arise only as relational entities, we see pantomimes as taking a symbolic role even before interlocking relations emerge.

Some theorists postulate an all-or-none emergence of language based on a single mutation that created a “language gene” as recently as 50 Ka (e.g., Berwick and Chomsky 2015). Others posit that “language” must have evolved as far back as *Homo erectus* (e.g., Barham and Everett 2021), using the term “language” to include even systems with few words and only rudimentary means of combining them, whereas we emphasize the long haul of biocultural evolution in converting such protolanguages into the complexity of evolutionarily modern human languages. The huge variety of lexicons and grammars of today’s languages (Haspelmath et al. 2005) reinforces the view that different innovations occurred at different places at different times. Some remained localized, while others spread far and wide, whether through contact diffusion, migration, warfare, or colonization. This reinforces our polycentric view of the origin of languages rather than an all-or none emergence of Language-with-a-capital-L.

1.2. THE DEBATE OVER MANUAL VERSUS VOCAL ORIGINS: THE CRUCIAL ROLE OF PANTOMIME

We do not deny the importance of speech, but do argue for the crucial role of the body (and especially the hands) in human and language origins. HOLO 2 explores polycentric paths to the diversity of evolutionarily modern languages, beginning with basic pantomime (§2). For us, basic pantomime involves the discovery that performing versions of actions in the absence of a target object may convey a context-dependent message about the action and/or object—just as miming holding the handle and lifting a cup toward the mouth may signify “please give me some tea” or “would you like some tea?” This is a social change for both mimer and observer, since both realize that communication need not be limited to familiar gestures. Communication becomes open-ended. In §3.2 we show how conventionalization provides the transition from pantomime to a protolanguage shared by a group.

Before considering the evolution of spoken languages, we argue that manual gesture played a crucial role in the evolution of languages (Armstrong et al. 1994; Hewes 1976; Kimura 1993; Stokoe 2001). We emphasize the role of pantomime as a crucial bridge between the technosphere of practical manual action and the semiosphere of protolanguages. Darwin (1871) recognized that some evolution of hominid cognition must have occurred “before even the most imperfect form of speech could have come into use”. Consider, for example, the facial expression of emotions (Darwin 1872). He supported the role of social communication in language and located its origin in techniques of courtship. We consider social exchange as part of a matrix in which social and referential functions of gesture and vocalization intertwined, combining perceptual-motor skills and cultural learning. Tomasello et al. (1993) suggest that this can be tied to developmental changes in social cognition. Donald (1991, 1998, 1999) stresses the revolution in communication when pictorial and other representations came to provide external memory (see §4.4 of HOLO 1:

d’Errico et al. 2027 [this issue]), while Gamble (1994) suggests that the demands of increasing social structure to deal with long-term variations in climate and food resources (see §3.2 of HOLO 1) were involved in the *co-evolution* of increasingly subtle perceptual-motor skills involving working memory, causal cognition, theory of mind, planning, teaching ... and language.

Pantomime remains relevant today. Pantomime, especially in context, can provide information about some current event, seek coordination with another person, or even support simple protonarratives when no shared language is available. It occurs in conventionalized (not basic) form in drama and theater (the narrative arts, Brown 2024), and in the cospeech gestures that enrich spoken language (Skipper et al. 2009). The creation of systems of symbolic communication, and therewith of symbolic biocultural niches, has occurred only during human evolution and, as suggested in Figure 2 below (§7), evolutionarily modern human languages have emerged rather recently when viewed on an archeological timescale.

Not all languages are spoken—the *signed languages* of the Deaf are fully-formed languages (Klima and Bellugi 1979; Polich 2005). As argued in HOLO 1, the increasing complexity of material culture—particularly in lithic production sequences requiring planning, error correction, and teaching—provides a key inferential basis for reconstructing the communicative demands that likely drove the emergence of pantomime as a bridge between action and representation. Once a protoword is conventionalized for use in the context of certain practical activities across a group, each fresh usage by one person is likely to be understood by its intended recipient (this is the crucial notion of *parity*). In spoken or signed languages, the material substrate for signified meanings is a conventional vocal or manual action, which (unlike the material signs explored in HOLO 1) leaves no trace in the archaeological record. Written texts, with a graphic substrate, can offer enduring records, but only go back a few thousand years.

Given the preponderance of spoken languages today, many researchers question the notion that a range of protowords emerged from manual gesture before our ancestors gained flexible vocal control and learning (Aboitiz 2013; Fogassi et al. 2013; MacNeilage and Davis 2005; Mithen 2024; but see van Mazijk 2024). However, Goldin-Meadow (2024) argues that not only is gesture an intrinsic part of modern-day human communication but may have been so from early times³, while Fay and Walker (2024) argue that current studies of modern humans tasked with developing new symbols support the gesture-first theory of language origin by showing that gesture supports the production of iconic signs more readily than does attempted sound symbolism, and these are crucial to communicative success. But, if so, how did vocal communication become the predominant mode of language? Gärdenfors and Ekström (2026) specify several intrinsically iconic aspects of vocal behavior, related to internal bodily or cognitive states. More generally, making use of pre-existing relationships between sound and non-sound (including physical move-

ment and internal physiological states) may have paved the way for a shift from the relatively inflexible vocal systems to flexible, open-ended spoken language.

“Speech” has at least two meanings in English:

- Speech-1 is the ability to deploy a vocal apparatus with many degrees of freedom, and to be able to flexibly learn novel vocal patterns, whether or not language is involved, while
- Speech-2 is the ability to deploy Speech-1 to produce utterances of a language⁴.

The debate, here, is whether hominins had *full* vocal control and learning (Speech-1) before they had any protosign, or whether increasing (but not necessarily extensive) use of protosign provided adaptive pressure to extend the neural mechanisms of conventionalization of communicative gestures from hand to mouth (Corballis 2002; Fay and Lim 2012; Leavens et al. 2014). §3.1 suggests that an initial predominance of protosign may have added to the adaptive pressure for the biocultural evolution of neural and laryngeal mechanisms to support protospeech and, eventually, Speech-1. But the debate does not change our crucial point—that manual skills, their imitation, and their role in pantomime provided a crucial driver for the evolution of the language-ready brain, ensuring the tight (though polycentrically varied) linkage between the cultural evolution of technosphere and semiosphere (Sinha 2015b). Thus, we see pantomime as crucial to the origin of language on the path from LCA-ga (our last common ancestors with the great apes)⁵ while noting some ways in which voice may have played a complementary role.

Given this, what we call POLE (Pantomimic Origin of Language Evolution) should be read as charting the essential, but not necessarily exclusive, role of pantomime in language origins. We reject views that see language and its evolution as confined to speech.

When might Speech-1 have emerged in our timeline? Nishimura et al. (2022) assessed a wide range of extant primates to conclude that it was evolutionary loss of vocal membranes that allows human speech to better tune acoustic control compared to other primate vocalizations. Such tissue is not preserved in the fossil record. Comparisons of the morphology of the phonatory and auditory apparatuses of Neanderthals and *H. sapiens* show similarities (Barney et al. 2012; Boë et al. 2002; Conde-Valverde et al. 2021) but cannot directly compare vocal membranes.⁶ Note that this comparison neither proves nor disproves that Neanderthals had the neural mechanisms that could learn to control their peripheral apparatus to produce meaningful conventionalized protowords in vocal form or those for recognizing their meaning from auditory input. Although genetic studies have suggested the existence of some cognitive differences between Neanderthals and modern humans (Pinson et al. 2022) there are also paleontological and archeological observations in support of the hypothesis that Neanderthals were articulate beings with a lexicon of (proto)words for communicating about spatial locations, hunting and gathering, fauna and flora, social relations, technologies, and their symbolic world (Dediu and

TABLE 1. A FOUR-STAGE MODEL OF LANGUAGE EVOLUTION
(adapted from Heine and Kuteva 2007).*

Stage	Linguistic characteristics	Main innovation	Possible correlations
1 (PL): Protolanguage ("Simple Protolanguage")	An open-ended inventory of protowords as form-meaning elements (interactive forms, holophrases)	Symbol creation	Early <i>Homo erectus</i>
2 (EPL): Expanded Protolanguage	Combining symbols into coordinate constructions ('flat syntax')	Concatenation, early constructions	Late <i>Homo erectus</i>
3 (EL): Early Language ("Simple Language")	Modification: Conventionalized patterns of modifier-modified constructions, resulting in phrasal recursion	Early grammaticalization	Different <i>Homo</i> populations (160 ka – 60 ka)
4 (EML): Evolutionarily Modern Language ("Complex Language")	Extending modification to clause combining: Modifier-modified constructions now used for clause subordination, resulting in recursion at the sentence level. Full-fledged languages as we know them today	Later grammaticalization	<i>Homo sapiens</i> and possibly Neanderthals (from 60 ka, continuing up to the present)

*Table 2 below schematizes the POLE account of the role of pantomime in the biocultural evolution of the language-ready brains that supported the first protolanguages.

Levinson 2013, 2018; Dediu et al. 2021; see also HOLO 1, §§4.2–4.4). Thus it is possible that 40,000 years ago, when Neanderthals faced assimilation by coeval African immigrants, both they and incoming *Homo sapiens* were on the cusp between Early Language and Evolutionarily Modern Languages, in the scheme of Table 1. Our treatment of the evolutionary path from protolanguages to evolutionarily modern languages in §§3 and 4, following the stages of Table 1, will be agnostic as to whether the emphasis is on vocal or pantomimic origins. We emphasize that there are no strict boundaries between the stages, and that the progression was polycentric.

Nothing in the above suggests that different groups of *Homo* developed the same cultural elements or protolanguages as others unless they were in social contact. Some innovations in general form were developed independently with their own particularities in different groups, whereas others may have spread through cultural contact as well as migration and/or assimilation of populations yielding results that may have blended in varied proportions the varied contributions.

1.3. A BRIEF LOOK AT EMOTION

Our account links language evolution to interaction of people engaged in practical activities. We will downplay

(but not ignore) the role of emotion in human interactions in much of the paper, but once we have elaborated our account of the pantomimic origins of language, welcomed voice into the equation, and charted the emergence of evolutionarily modern languages, we devote §5 to "Empathy, sociality, and narrative," and so a note on emotion seems appropriate here.

The great diversity of human emotions has evolved "atop" what Denton et al. (2009) called the "primordial emotions," based on drives such as fear, hunger, thirst, sex, and caregiving (Koelsch et al. 2015). Note that the first three reflect the state of the individual, while the latter two are inherently social, thus requiring some communication. However, for social animals, hunger and thirst can also require a group to coordinate its behavior. Thus, as emotions evolved, each could have the *internal* role of being "drivers" for action—we act very differently if we are angry or are courting a loved one—and the *external* role whereby facial and bodily expression can convey emotional state to others so that they can respond more-or-less appropriately (Arbib and Fellous 2004).

As previously noted, Darwin (1872) observed—see also Ekman (1998)—that there are parallels between the facial expression of emotions in humans and other animals (especially mammals). However, recent research (e.g., Bar-

rett 2017) does not lend support to Darwin's and Ekman's stipulation that each "basic" emotion is a universal, in-built reflex and has a distinct facial expression as a consistent "fingerprint/signature." The evolution of neocortical systems and their reshaping of "ancient" emotion related systems (LeDoux 2012; Rolls 2011) has come to support even the "aesthetic emotions," such as those engaged in the appreciation of art and music (Scherer 2013). Emotions vary across cultures. As a result, human emotions

"are not [merely] remnants of our pre-sapient past but rather they are important characteristics of an active, searching, thinking human being. Novelty, discrepancy and interruption generate visceral responses while our cognitive system interprets the world as threatening, elating, frightening or joyful" (Mandler 1984: 220).

Emotions play a strong role in conversation. Thus, despite emphasizing narrative rather than conversation in HOLO 2 (see §6.1), we recognize that an evolutionary analysis of conversation would also have great value (Dessalles 2007; Dingemanse et al. 2013; Roberts and Levinson 2017). Narratives have been shown to actively modulate emotions at the neural level (Altavilla et al. 2025). Emotional sequencing within narrative, for example, fear → relief, sadness → hope, is key to the structuring of a narrative (Knight et al. 2025). Universally, stories may have an emotional "crisis" at their center that various narrative formats seek to resolve, with particular emotions requiring different narrative strategies. Contemporary narratives include detective fiction that invites curiosity and offer surprise endings, religious stories that induce awe or invite contemplation, and tragedies that evoke grief. Furthermore, as individual differences develop in cultures, we begin to think of ourselves, for instance, as "calm" people or as "quick to anger" or "easy to please." Such reflexive personal self-images are the result of a lifetime of narrative "roller-coasting," starting off with the childhood training we receive in learning to accurately read the barometer of community emotions (Nair 2003b, 2011b).

Emotional traits such as compassion systematically shape narrative structure forming "redemption arcs" through affective tonalities (Zieba 2025). Narrative, in this respect, appears to offer a "temporal grammar" of emotion in EMLs that helps us understand how proto-narratives might have worked in early evolution to enable adaptive response to environments that the community deemed fear-inducing, surprising, loving, and so on. However, to state this is to offer challenges that go well beyond the insights we offer in §5.

1.4. INTENTIONS, REPRESENTATIONS, AND THE GREAT MOVE

Humans share many capacities with other primates and some—such as those engaged in finding food, evading predators, and reproduction—can be found in various forms far back in the evolutionary tree (Cisek 2022); humans have facial expressions of emotions that are related to those exhibited by diverse mammals (Darwin 1872). Much

of what we do employs our biologically evolved versions of these survival traits that we share with other species that make no use of language, though they may employ a small repertoire of communicative signals. Our challenge here has been to explore the biological capacities that readied us for the cultural evolution that made it possible for us to categorize and name these emotions and survival capabilities, tell stories about them, and to support waves of increasing social complexity, from domestication of animals and agriculture to urbanization. Below, we introduce the key notion of the Great Move. But first we need some words of caution about three terms, "intention," "aboutness," and "intentionality."

We use the word *intention* for whatever it is that an animal or human wants, needs, or plans to do, its goal, whether it is an animal seeking prey to eat or a human intending to write a book. When we speak of the *aboutness* of language *at its phylogenetic base*, we are arguing that, as for the infant and young child expressing its needs, at its root (proto)language expresses states of the physical world of the immediate environment or of immediately intended states of that world, as when one forager communicates to another the presence of fruit on a tree that appear ripe and encourages action that will transfer the fruit from the tree to a basket to be carried back to the tribe. This grounding in the here-and-now-and-next is the sense of "aboutness" we emphasize in seeking the origins of (proto)language.

Nonetheless, as languages become richer, one can move from looking despondent if the fruit is not yet ripe to expressing, for example, the intention to return in two weeks when the fruit is likely to be ripe. Symbolic expressions as used today *stand for* or *represent* something outside the sign itself; they are *about* objects, events, and/or states of affairs that may include requests or commands to coordinate ongoing action, but may also be displaced in space and time, or may involve imagined or hypothetical scenarios. Human symbolic repertoires changed as social interactions, social systems and technology changed. As emphasized in HOLO 1 (d'Errico et al. 2027 [this issue]), symbolic systems are inherently *normative*, in the sense that they rely on socially shared conventions for their transmission and interpretation. The emergence of protolanguages can thus be seen as a critical threshold at which these normative constraints became increasingly explicit, stabilized, and collectively regulated by human groups within expanding and interacting biocultural niches. Thus, through cultural evolution as well as individual development, the grounding aboutness of language in the physical and personal here and now expands to its current richness.

Unfortunately, the philosophical term *intentionality* is defined in a way that can sow confusion. The Oxford English Dictionary defines intentionality as "The distinguishing property of mental phenomena of being necessarily directed upon an object, whether real or imaginary." For A. Byrne (2006):

... some things are *about*, or are *directed on*, or *represent*, other things. For example, the sentence "Cats are ani-

mals” is about cats (and about animals), a map of Boston is about Boston, ... [but] the city of Boston [is] not about anything. Many mental states and events also have “aboutness”: the belief that cats are animals is about cats, as is the fear of cats ... Arguably some mental states and events are not about anything ... This – rather vaguely characterized – phenomenon of “aboutness” is called *intentionality*. ... This medieval terminology was reintroduced by the Austrian philosopher Franz Brentano in his 1874 book *Psychology from an Empirical Standpoint* (Brentano, 1995) ...

The point here is that language can be used to express statements about “situations” that do not exist, as in the example “Adult elephants are usually found in teacups.” In some sense, then the voyage in HOLO 2 is from “real” aboutness that we chart for language origins in §2 and §3 to the power of narrative in myth and storytelling as presented in §5.

But now we come to something momentous that occurred in evolution. To frame this, we summarize the “the great move” as charted in the foundational document for the term *cognitive architecture* in modern cognitive science, *Unified Theories of Cognition* by Allen Newell (1990, 1992). For Newell (1992: 426–427),

Knowledge must be represented in order to be used. The concept of representation is captured by the representation law. In an external world, entity (X) is transformed (T) into entity (Y). A representation of X-T-Y occurs in a medium within some system when an encoding of X to an entity in the medium (x) and an encoding of T into an internal transformation in the medium (t) produces an internal entity (y), which can be decoded to the external world to correspond to Y.

Our concern is that this fails to distinguish two levels of representation. Consider a frog snapping at a fly. There must be neural circuitry that responds to the prey-object by triggering and passing parameters to the neural circuitry that controls the motor apparatus for approaching, orienting, and snapping to catch the fly. But we would not want to say that the frog “knows” about catching flies—or we might compromise by saying its knowledge is *implicit*. The crucial fact, which will be exemplified in the transition from demonstration for teaching to pantomime for teaching (§2.3.1), is that the brain being able to represent *symbols* for objects, actions, and more, is distinct from having the neural circuitry for recognizing objects and executing actions. Language provides the ability to develop diverse sentences employing symbolic structures that represent the knowledge *explicitly*. Although elaborate neural mechanisms support our breathing and our digestion, we rarely talk about them unless they malfunction. If we visit a pond on a sunny day, seeing and smelling the flowers by the pond, and reveling in the play of light on the water and of fish in the water, we have an emotionally satisfying experience of a totality that may or may not guide our action (smile and move on, sit awhile by the pond, or dive right in) but cannot be exhausted by a symbolic description. We

reject the notion (Fodor 1975) of a language-like structure for our awareness and thought.

However, Newell’s (1992: 427) crucial insight is the “Great Move”:

Obtaining a representation for a given external situation seems to require discovering an internal medium with the appropriate natural transformations – this is the essence of analog representation. But as external situations become more diverse, complex, and abstract, discovering adequate analogs becomes increasingly difficult, and at last impossible. A radically different solution exists (the great move), however, where the internal medium becomes freely manipulable with combinatorially many states and all the representational work is done by being able to compose internal transformations to satisfy representational laws. ... This leads to the familiar, but incredible, results from computer science about *universal* computational systems that can attain the ultimate in flexibility. They can produce, by being instructed, all the functions that can be produced by any class of machines, however diverse.

Newell wrote this at a time when AI, thanks in great part to his contributions with Herbert Simon (Newell and Simon 1972; Newell et al. 1959), had learned to program digital computers to do a fairly good job at various problem-solving tasks that had previously been seen as requiring human intelligence. Given recent breakthroughs in adaptive artificial neural networks, we see that “instruction” in the sense of an external agent providing an explicit program is no longer required.

Arbib and Liaw (1995) turned from the technology of digital computers to the biology of neural networks in supporting the behaviors of diverse animals. In particular, they contrasted the special-purpose vision of a frog with processes of distributed planning that supports human interpretation of visual scenes (Draper et al. 1989):

... a satisfactory account of Newell’s “Great Move” should not seek a complete break from using specialized materials to support different [sensory and motor] schemas to using a medium in which it is possible to compose copies of whatever schemas are needed to form novel representations. Rather, we should [assess] how instinctive behavior provides a basis for, and is intertwined with, rational behavior. When we study frogs, we see the animal’s behavior mediated by the dynamic interaction of multiple special-purpose schemas implemented in dedicated neural circuitry. But when we seek to understand human vision, we combine a model of low-level vision implemented across a set of dedicated brain regions with a variety of “general-purpose” media in which ... schema instances ... can be assembled, parameterized, and bound to regions of the image as they compete and cooperate in a process of distributed planning which creates an interpretation of a visual scene.

Their claim is that the mechanisms of distributed interaction seen in visual interpretation extend to distributed goal-directed planning. This includes the use of language (as shown, e.g., by Barrès and Lee 2014).

[In this view] the “Great Move” takes us from dedicated circuitry for every schema to the creation of novel schemas distributed across relatively general-purpose machinery. ... The hedge “relatively” ... is important. ... [W]hat were once thought to be large undifferentiated [human] brain regions can in fact be discriminated into far smaller regions with distinctive patterns of input and output connections (Goldman-Rakic, 1991). Presumably, each of these regions can access a distinct set of other regions, and as these sets of subsets can be developed hierarchically, the result can be a “general-purpose” representation-but one which is distributed across more or less specialized partial representations each in a specified brain region, rather than ... the totally uniform computational medium of the serial general-purpose computers used in Newell’s many contributions to classical AI.

This Great Move emphasizes changes in the generality of representation within a computer or brain. Another Great Move involves the development of externalized memory. We thus review Merlin Donald’s (1991, 1998, 1999) theory of the evolution of culture and cognition, cited briefly above. He posits three stages that separate the modern human mind from that of LCA-ga:

1. The introduction of a supramodal motor-modeling capacity which he calls *mimesis*, and the exploitation of this capability to create representations with the critical property of voluntary retrievability.
2. The addition of a capacity for lexical invention, and a high-speed phonological apparatus, the latter being a specialized mimetic subsystem.
3. The introduction of *external memory* storage and retrieval, and a new working memory architecture.

Ad (1): POLE is broadly consistent with this general framework, but places special emphasis on “mimesis” as occurring originally as complex action recognition and imitation in relation to the visual control of hand movements going beyond the emulation capacity of great apes (R.W. Byrne and Russon 1998), and thus presumably of LCA-ga (§2.3.1), to provide the basis for pantomime as a key innovation on the evolutionary path that led to the emergence of the language-ready brain. The voluntary retrievability of Donald’s stage 1 requires detached mental representations, thus tying back to the Great Move.

Ad (2): Communicating by symbolic representations (recall HOLO 1) goes beyond the signaling systems of animal communication (and our earlier discussion insists that “having a thought” is not the same as being able to put it into words). Taken together, Donald’s stages 1 and 2 may take us to protolanguage (for body and voice) but not yet to languages. In the rest of §3 and continuing in §4, we add a further component—the passage to languages via the co-evolution of lexicon and grammatical constructions.

Ad (3): However, Donald’s Stage 3 introduces a vital component that complements the POLE framework—the creation of external memory. This includes permanent body modification, deliberate sequential marking, carving, painting scenes, and writing. HOLO 1 (especially §4) scoured the archeological record for evidence that early hu-

mans and Neanderthals created artifacts of symbolic value, constituting a foreshadowing of Donald’s stage 3. However, his Stage 3 emphasizes that artifacts (drawings, paintings, enduring marks of other kinds) are then created as part of a symbolic system that complements and/or relates to other aspects of symbolic power, thus providing external systems for recording post-Great Move symbolic representations. We return to this topic in §3.2.

1.5. PROSPECTUS

In laying the basis for the evolution of protolanguages, §2 integrates an account of the role of pantomime in teaching new skills, the *Technological Pedagogy Hypothesis* (Gärdenfors and Högberg 2017; Stout 2018), with the earlier stages of the so-called *Mirror System Hypothesis* (MSH: Arbib 2012; Rizzolatti and Arbib 1998), rebranded as the *Beyond the Mirror Hypothesis* (BtMH), in rooting protosign in pantomime (§2.2.1). §2.3 explores the role of pantomime in the evolution of teaching and social coordination, introducing the key notion of the coevolutionary expanding spiral, in this case relating communication and praxis. §2.4 introduces the notions of protonarrative and protoconversation.

§3 then describes how the brain became language-ready, supporting simple protolanguages and then supporting the coevolution of words and constructions. §3.2 extends our discussion linking language-readiness to material praxis. Our synthesis, outlined in Table 2, is called POLE for “Pantomimic Origin of Language Evolution.”

§4.1 explores how grammaticalization played a key role in the emergence of early languages and then evolutionarily modern languages. §4.2 complements this by viewing flexible construal as another driver for grammatical complexity. Throughout, we see no need to postulate a hypothetical “language gene,” but rather show how languages evolved from their precursors over at least the last 300,000 years based increasingly on the evolutionary trajectories of different human populations (from *Homo erectus* to *Homo sapiens*, Neanderthals, and, possibly, others such as Denisovans), in which cultural evolution attained increasing autonomy from biology. HOLO 1 (d’Errico et al. 2027 [this issue]) documented growing evidence demonstrating gene exchange between Neanderthals, Denisovans, and *Homo* immigrants to Europe and Asia (L. Li et al. 2024). This suggests that each of these species may have contributed to the exchange of cultural practices and knowledge, facilitated by social interactions and interbreeding. This in turn may have played a crucial role in the development of protolanguage skills across these different human populations, but does not attest to how far cultural evolution may have taken Neanderthals or Denisovans along paths towards modern-like language use.

§5 introduces the notion that “narratives are empathy machines” (Nair 2011a) and uses properties of narratives in recent millennia to chart a path from protonarrative to narrative; we suggest how the demands of narrative contributed to further innovations in grammar. It then looks at emotion in relation to individual understanding, before exploring how the increasing complexity of language sup-

TABLE 2. STAGES POSITED FOR POLE, OUR HYPOTHESIS ON “PANTOMIMIC ORIGIN OF LANGUAGE EVOLUTION.”*

LCA-m	POLE-1	LCA-m had manual dexterity, an ability to recognize the actions of conspecifics, and the ability to learn new manual skills, but little in the way of imitation. Communication involved a limited set of species-specific calls and facial gestures.
LCA-ga	POLE-2	LCA-ga preserved the ability for dexterous manual actions and for recognition of actions by others, plus a <i>simple imitation (emulation) system</i> for manual actions based on trying to replicate achievement of subgoals. A small innate repertoire of manual gestures could be augmented by gestures developed by dyads.
Hominin Biocultural Evolution Post LCA-ga	POLE-3	Evolution post-LCA-ga led to hominins employing both <i>complex action recognition</i> and <i>complex imitation</i> based on recognizing subgoals and <i>various movement details</i> of the actions involved.
	POLE-4	POLE-4a - Hominins built on complex action recognition and imitation to develop <i>teaching via demonstration</i> . POLE-4b - Teachers came to also control the acquisition of certain skills via <i>pantomime</i>
	POLE-5	<i>Ad hoc pantomime for communication</i> : Performing actions in the absence of objects to convey to others the need for an object or some associated behavior was coupled with the ability of others to recognize when a performance is an intended communicative pantomime rather than an attempted practical action. This is the root of open-ended communication.
	POLE-6	<i>Basic pantomime</i> – ad hoc pantomime and ritualizations that may be specific to dyads or small groups but are not conventionalized – could support protonarratives and protoconversations.
	POLE-7	The emergence of early <i>protolanguages</i> , of being able to freely create novel communicative acts while meeting the parity condition. At this stage, <i>protowords</i> may be available in either modality, signed or spoken POLE-7a – <i>Protosign (Bodily Symbolic Gesture)</i> emerges as a manual-based communication system, which includes (but is not restricted to) pantomimes that have become conventionalized within communities to allow meanings to be conveyed more economically and less ambiguously. POLE-7b – Nonhuman primates have very limited vocal learning. <i>Protospeech (Vocal Symbolic Gesture)</i> may rest on the “invasion” of the vocal apparatus by collaterals from the communication system based on manual control and recognition.
Primarily Cultural Evolution	Post-POLE	Cultural rather than biological evolution dominated as language emerged from (vocal and signed) protolanguage, with a widening of lexicons and grammars supporting <i>compositional semantics</i> to express and comprehend an unlimited range of novel meanings. Expanded protolanguages → early languages → evolutionarily modern languages

*Recall Endnote 5 for the meaning of LCA-m and LCA-ga. The details will emerge across §§2 and 3. As stressed in §1.2, voice may play an earlier role, though here we stress its relation to pantomime as developing at stage POLE-4b. We suggest that the earliest protolanguages need not have been compositional. A crucial part of POLE is to suggest how protolanguages became increasingly compositional as they became more complex.

ports the “social construction of reality” in both semiosphere and technosphere as well as unleashing the human imagination.

§6 focuses on cognitive architecture. It first places Ferretti’s triad of mental time travel, mental navigation, and mind reading, within an evolutionary perspective. It then offers an expanded view of the evolution of human cognitive architecture, building on general cognitive abilities of animals and expanding via POLE to an EvoDevoSocio view of the cognitive abilities that support modern humans and their rich use of language in a complex technosphere and semiosphere.

Finally, §7 offers a somewhat controversial timeline for our account of the trajectory from the polycentric emergence of material and symbolic practices to the cultural evolution of increasingly complex communicative systems, including the emergence of evolutionarily modern languages.

The HOLO project compels a re-evaluation of theories of the origins of languages, placing these firmly within the evolution of human biocultural niches characterized by interacting developments in the technosphere and semiosphere, while emphasizing a *polycentric* emergence of protolanguages and thence languages, consistent with the diversity of evolutionarily modern languages and human cultures.

2. FROM INTERACTION TO INNOVATIVE COMMUNICATION VIA PANTOMIME

Comparative primatology informs hypotheses on how biocultural evolution came to support *language-ready* brains for humans long before cultural evolution yielded the ever-expanding biocultural niches that supported language-using communities. Modern monkeys use innate (though somewhat adaptable) vocalizations to communicate (Seyfarth et al. 1980a); great apes in the wild also communicate via a small range of apparently innate manual gestures while having the ability to add a very few gestures to that repertoire (Hobaiter and Byrne 2014). Pika and Fröhlich (2019) support the social negotiation hypothesis that “gestures emerge from an exchange of social behaviors between interactants” while Graham et al. (2024) propose that great ape gestures recruit features of their existing behavioral repertoire for communicative purposes. This suggests that LCA-ga had restricted communicative repertoires, but some flexibility for developing a limited repertoire of novel manual gestures.

For Żywicznyński et al. (2018),

Pantomime represents a communication mode characterized by the absence of language and relying on movements of the whole body—mainly manual gestures, body movements, facial expressions or voluntarily enacted mannerisms. Accordingly, the communicative potential of pantomime depends on the intentional use of integrated ... movement of ... body parts to convey meaning.

In what follows, our concern is with how simple forms of pantomime—related to coordination of actions between individuals as well as developing from demonstration in the teaching of skills—could contribute to the early evolution of symbolic communication in humans.⁷ §2.4 argues that *pantomime* could bridge from communication about actions to communication in terms of protonarratives and protoconversations. Our emphasis on polycentric human and language origins makes clear that “fragments” of later evolutionary stages may have accompanied earlier stages in different ways and to different extents in different communities as varied (proto)languages developed.

2.1. DISPLACEMENT, TRUST, AND CUMULATIVE CULTURE

Apes communicate with a limited range of gestures that are almost exclusively *cued* (Gärdenfors 1995), that is, they refer to something in the present or immediate future, whereas human communication is typically *detached* or *displaced* whether referring to what is elsewhere, or to the past or an imagined future (so-called “mental time travel”: Corballis 2019; Suddendorf and Corballis 2007), or expressing fantasies (§6.1). Detached communication exploits a symbolic or iconic system, while animal communication can be seen as a system of signals.

Waciewicz and Żywicznyński (2018) argue that the emergence of complex languages depends on a “platform of trust” enabling “cheap but honest” communication that goes beyond animal signal systems. An important move from §2.3 to §3.2 will be the transition from ritualized forms of communication between members of a dyad to conventions that allow symbols to become used and understood across a community—the dyad precedes the group but the group then shapes the individual. Nair (2003b) discusses cooperation as a means of building community from basic conversational dyads to large polyphonic groups but also notes that deception is an integral part of human communication (Nair 2026). As modern humans, although we often assume that the information people give us is correct, we also continually assess whether their interests differ from ours. This suggests that, rather than looking at evolution of communication as building on an established platform of trust, we should factor evaluation of (dis)alignment of interests into the evolution right from the start. More generally, Tomasello and Vaish (2013) consider the origins of human cooperation and morality. Patterns of coercion are outside our scope but are crucial for an understanding of the emergence of complex cultures as in urbanization and slavery.

Apes have cultures in the sense (Whiten et al. 1999) that different groups exhibit certain traits that are not determined either by the available resources, by genotypic differences between groups, or by the geographic distance between them. Gunasekaram et al. (2024) have recently found that complex behaviors in chimpanzees are transmitted across generations and communities via female migrations. Although these findings illuminate chimpanzee cultural evolution, a crucial difference with human culture

is that changes are at so low a rate in the absence of human intervention that we have the impression that they do not have the extended capacity for *cumulative* cultures exhibited by humans.⁸

Transmission fidelity and population size and complexity all play a role in marking the human difference as does the fact that the language-ready brain can support “ratcheting” whereby old skills can be modified and combined to support new skills to an extent far beyond that seen in present-day great apes (Tennie et al. 2009; van Schaik et al. 2019). For Sterelny (2012), ratcheting occurred because each generation of humans came to enrich the learning environment of the next generation. However, a more complex analysis is required. As discussed below, demonstration may have sufficed for Oldowan technology, but explicit teaching seems necessary for transmission of Acheulean technology. And yet there were hundreds of millennia of stasis. We suggest that the crucial transition to ratcheting may have come as expanded protolanguages became early languages, so that plans could be discussed using single words to name novel skills and their key subcomponents. This supported both thinking about and discussing ways in which one or more skills could be adapted and combined to achieve goals that were hitherto unattainable. Ratcheting is thus crucial to human cumulative culture both in the technosphere and in the semiosphere, with language “going hierarchical” being a form of ratcheting that can build more complex plans that thus feed back into the technosphere.

2.2. THE EMERGENCE OF PANTOMIME

Social interaction and the motivation to communicate about the world (and more) are central to the human condition. We now chart how recognizing the actions of others as they go about their behaviors may have given rise to pantomime as an open-ended form of communication that supplements the limited set of vocal signals and manual gestures used by other primates. The success of pantomime requires that a group comes not only to generate pantomimic gestures but also to understand that such gestures performed by others are intended to communicate “something.” A reference point here is that apes may acquire a few novel gestures through *ontogenetic ritualization* (Tomasello et al. 1997) in which repeated interaction between members of a dyad may become reduced to a mutually understood gesture to request that interaction—but this novel gesture is not created intentionally, and may remain for use just by the dyad. However, symbols created through conventionalization across a community (§3.1) play no role in either great ape culture or communication in the wild. Human-enculturated apes can be taught to use a limited range of ASL (American Sign Language) handshapes or lexigrams to communicate with humans (Lyn 2024) but there is no reliable evidence of widespread, entrenched symbolic practices or the use of syntax in any non-human species. This leads to the conclusion that *symbolic niches* (in whatever diversity) emerged after LCA-ga.

2.2.1. Rebranding: The Beyond the Mirror Hypothesis

Mirror neurons, as first discovered in macaque monkeys, are active both when the monkey executes a certain type of manual action and when it observes others performing similar actions (di Pellegrino et al. 1992; Gallese et al. 1996). Rizzolatti and Arbib (1998) sketched a possible evolutionary path from manual skills to language, the Mirror System Hypothesis (MSH). They noted that the area of premotor cortex in the macaque containing these neurons is homologous to Broca’s region, a human brain region long understood to be engaged in speech production. But why is a speech area related to manual gesture? The key is that Deaf people⁹ exhibit an aphasia comparable to that of speaking people’s impairment in spoken language from comparable lesions (Poizner et al. 1987). In other words [sic], to the extent that Broca’s region is a language area, it is not just a speech area. These findings set the foundation for MSH, but the overall shape of MSH emerged only later (Arbib 2005b, 2012) and is still undergoing development (Arbib 2020).

MSH exploits *comparative neuroprimatology* to compare brain, behavior, and communication in monkeys, apes, and humans as the basis for assessing the capabilities of LCA-m (our Last Common Ancestor with the monkeys) and LCA-ga to yield hypotheses on the biological and cultural evolution $LCA-m \rightarrow LCA-ga \rightarrow H. sapiens$. While modern chimpanzees and bonobos have the closest genomes among all current non-human species to modern humans, we now prefer to consider the last common ancestors of humans and great apes generally rather than with chimpanzees and bonobos alone. MSH starts with the observation that, although no other extant primate has extensive vocal control and learning, all have manual dexterity and can acquire new manual skills (Arbib et al. 2008). Moreover, apes can develop a limited number of novel gestures. MSH develops the case that the path from LCA-ga to modern humans built upon manual flexibility to yield extended vocal flexibility. It is notable, for both MSH and the Technological Pedagogy Hypothesis, that processing language partly shares a neural genetic basis with processing tools and body parts (Wen et al. 2024).

Crucially, mirror neurons fire only for actions *within the animal’s repertoire*, whether executing the action itself or recognizing a similar action performed by others. All animals can recognize the actions of others and act upon them, whether it be chasing prey, avoiding a predator, or engaging in some form of social interaction—action recognition in general is crucial, and is not limited to recognition of actions one can oneself deploy. MSH emphasizes the ability to imitate—to recognize that an action performed by others could be worth adding to one’s own repertoire—rather than recognizing one’s own actions. We thus see the grounding of MSH not in the mirror neurons themselves, but in the way brain systems “beyond the mirror” can support the learning of new actions. Thus, despite two decades of using the MSH label, we now (following Arbib and Cuccio 2026) relabel the theory as the *Beyond the Mir-*

ror Hypothesis (BtMH) to emphasize our concern with the dramatic changes in parts of the brain beyond the mirror system that have enabled humans to evolve and use languages. BtMH thus seeks not only to explain how the structure of language evolved but how it did so in such a way that newcomers could learn the manual and vocal gestures of a language while others can innovate, with such innovations being ignored or being learned and shared by others. All this exhibits how BtMH lives within the EvoDevoSocio framework.

2.2.2. Early Forms of Action Recognition and Imitation

A crucial distinction, based on comparative studies of capacities for imitation in diverse species of present-day monkeys and apes, was to posit a corresponding progression in ancestral forms:

- **POLE-1.** LCA-m had manual dexterity, an ability to recognize the actions of conspecifics, and the ability to learn new manual skills, but little in the way of imitation. Communication involved a limited set of species-specific calls and facial gestures.
- **POLE-2.** LCA-ga preserved the ability for dexterous manual actions and for recognition of actions by others, plus a *simple imitation (emulation) system* for manual actions based on trying to replicate achievement of subgoals. A small innate repertoire of manual gestures could be augmented by gestures developed by dyads.

Note that if we define an action as combining a goal and the movements that achieve it, this is *not* an action-copying system, but rather a means of coming to recognize subgoals to be achieved in completing a behavior, and then developing actions (possibly by trial and error) that can achieve those subgoals. This is a case of *emulation* but for our evolutionary account we refer to it as *simple imitation* in contrast to complex imitation as defined below. For LCA-ga, we posit that a small innate repertoire of manual gestures could be augmented by novel gestures developed within a group to mediate communication.

The model of simple imitation known as *behavior parsing* (R.W. Byrne 2003) posits that an ape seeks to emulate a behavior exhibited by another because the goal it achieves appears desirable and that certain conditions (we call them subgoals) recur as the behavior is repeated with variations dependent on current circumstances, but the means for bridging from one subgoal to another is often unfamiliar and must be learned through trial-and-error (clearly, new tasks can be mastered more readily when these bridging actions are familiar).

- **POLE-3.** Evolution post LCA-ga en-route to modern humans involved the linked emergence of both:
 - *Complex action recognition*, the ability to recognize *various movement details* of actions within the behavior of others (with limitations based on the familiarity of the component actions and the complexity of the assemblage) and the subgoals they achieved; and,
 - The ability to exploit this to support *complex imitation* of observed skills, based on trying to replicate not only achievement of subgoals but also the action

trajectories used by another to achieve them.

Complex imitation preserves the ability to recognize goals and subgoals, but the acquisition of bridging actions came to be accelerated by attention to the movement details of the other's actions. *Humans* may develop their own way of reaching a subgoal (by trial-and-error, possibly serving simple imitation) or can learn by complex imitation, paying attention to the details of the demonstrator's actions, and modifying their actions accordingly. In either case, much practice may be needed to develop skilled sensorimotor coordination.

The evidence of *apprenticeship* among various species of *Homo* but not in other non-human animals (Colagé and d'Errico 2025; Donald 2012; Sterelny 2020) may be taken as evidence for complex imitation emerging post LCA-ga. The systematic exploitation of lithic cores to produce multiple blanks of similar lengths, as observed at the 2.3 Myr old Lokalelei site in Kenya, indicates a level of technological standardization within the Oldowan (see HOLO 1, §3.4.1). Such standardization suggests that transmission of this knowledge required complex imitation—recognizing goals and subgoals, attending to movement details of another's actions, and modifying one's behavior through practice to achieve skilled sensorimotor coordination, but may not have involved teaching (Gärdenfors and Högberg 2017). These capabilities imply that the hominins responsible (possibly *Australopithecus*) possessed the cognitive and social mechanisms necessary for at least limited forms of complex imitation.

Fractionation: Since a novel behavior may contain actions that are not variants of familiar actions, complex action recognition supports the *crucial* added ability for the observer to *fractionate* a behavior into *apparent* subactions, and then approximate the performance on the basis of observed movements, with increasing practice yielding increasing skill in performing the new action in varied situations (or failure leading to inaction, or to trying a new fractionation).

2.3. PANTOMIME FOR TEACHING AND SOCIAL COORDINATION

Complex imitation sets the stage for pedagogy and the role played there by pantomime (§2.3.1). A human may acquire a new behavior upon their own initiative not only because the result seems desirable, but also because a caregiver or teacher supervises the process of imitation, whether or not the goal seems desirable to the initiate. Another occasion for acquiring a new behavior is the “desire to belong,” in which a skill is acquired because others are doing it. We add that whether based on imitation or just the trying out and tuning of sensorimotor skills, *play* has a major role in the development, individual and social, of the young of many species. For example, young *Cebus* monkeys play at nut cracking (apparently enjoying the movements and the banging sounds) long before they are able to effectively use a hammer stone to crack a nut placed on an anvil (Fragaszy et al. 2017).

We next examine the role of pantomime for teaching. By considering it as providing a step beyond the demonstration of particular actions, we clarify the way in which pantomime can mediate between expansions of the technosphere and the semiosphere as increased complexity of the production of material artifacts drives and is driven by increased complexity of pantomime and its successors in protosign. We then turn to a more general discussion of the role that pantomime may have played to support social coordination. Each could contribute to the later emergence of protolanguages, long before humans had even early languages.

2.3.1. Demonstration and Pantomime for Teaching

The *Technological Pedagogy Hypothesis* (Stout 2018) asserts, in particular, that hominin reliance on stone toolmaking generated selection pressure for the emergence of teaching and language. By extension, there were other skills whose acquisition could be aided by teaching rather than mere observation. Although many animal species, especially primates, engage in diverse forms of social learning (Boesch et al. 2019; De Waal 2016), intentional teaching within non-human species is very limited, if it exists at all (Hoppitt et al. 2008; Thornton and Raihani 2008). Hoppitt et al. (2008) also explain why teaching—in various forms—is present in all human cultures (Gärdenfors and Högborg 2017; Kline 2015). Several studies emphasize intentional teaching as important in cultural transmission when learning complex, cognitively opaque skills, and Colagé and d’Errico (2025) offer a scenario for the spatial, temporal, and social evolution of cultural transmission in our lineage. This emphasis on teaching extends the developmental niche highlighted in HOLO 1 (d’Errico et al. 2027 [this issue]), where socially guided learning and extended ontogeny are seen as central mechanisms in the co-evolution of cognition, material practices, and communication.

- **POLE-4.** Evolution post LCA-ga that led to modern humans complemented complex action recognition and imitation with the development of *teaching relationships* wherein teachers came to control the acquisition of certain skills via first demonstration and later pantomime for teaching.

Demonstration involves *intentionally* showing somebody else how to perform a task by explicitly aiding the complex action recognition that supports complex imitation—both by emphasizing the actions involved in manipulating objects, perhaps by slowing them down or exaggerating them, and by drawing attention to the affordances for each action. Gärdenfors and Högborg (2017) and d’Errico and Banks (2015) argue that demonstration was already used in teaching some forms of the Oldowan stone knapping technology but that learning this technology requires little if any symbolic form of communication. Morgan et al. (2015) report how using language can aid teaching Oldowan knapping to modern humans, but (unlike students being taught these skills today) learners in Oldowan times would have lived in a community where they had observed many instances of stone knapping and the use of the re-

sultant flakes. Nonetheless, Morgan et al. (2015) see their data as advancing the technological pedagogy hypothesis. They suggest that low-fidelity social transmission, such as imitation/emulation, may have contributed to the ~700,000 years stasis of the Oldowan technocomplex, and that teaching or proto-language may have been pre-requisites for the appearance of the later Acheulean technology (see §3.3.1 of HOLO 1). However, the relevance of language to more complex skills does not imply that language replaces the relevance of demonstration and pantomime. Imagine trying to teach the skill of fabricating an Acheulean stone axe by simply providing a verbal description. In particular, there is the issue of how the student acquired the relevant action-terms in the first place (see Castro et al. 2025 for recent insights on this topic).

Intentional evaluative feedback can benefit all forms of coordination between agents, but it is especially significant in the development of teaching, whether for a mother attending to her child or someone explicitly teaching a new skill to an apprentice. The communication that is required for early forms of teaching may involve only evaluative expressions, say a grunt or a warning sound. However, teachers may wish to direct the learner’s attention to something relevant in a learning situation. This is typically done by declarative pointing toward a focal object.

For *demonstration*, the brain of the teacher need only engage the affordances and effectivities needed for the performance of the skill. Teaching by demonstration involves a change from POLE-3. The teacher is performing the skill, but now he is not only aware that he is being observed by the student but wants the student to learn from his performance. This requires students to understand that they are to learn by attempting to imitate the demonstration. The teacher may thus modify the performance by pausing when each subgoal is reached and exaggerating certain of the intervening movements to help the student’s observation of the behavior in the service of acquiring a new skill (Gärdenfors 2017; Zukow-Goldring 2006). For *pantomime* the teacher’s brain must hold the new motor skill of performing the pantomime, associate it with the action it represents, and develop new affordances for teaching, observing the occasions in the student’s performance when performance of the pantomime is appropriate. Here the pantomime may initially be an ad hoc variation on portions of the demonstration but—as we shall discuss in §2.4—a teacher may come to ritualize the pantomimes needed to teach each skill.

Variations in the trajectory of the pantomimed action may indicate some of the parameters of the suggested action. Such modulation of trajectory has a limited role in modern signed languages (though note the emotional inflection that can modulate signing), whereas the trajectory variation can play a crucial role in the cospeech gestures that may accompany speech (as in displacing the hands while saying “the fish was yay big”). Signed languages have co-sign gestures (Sandler 2013).

Crucially, students must recognize that the pantomime is intended to remind them of the possible actions they

should next perform, and thus this form of communication is open-ended.

Both demonstration and pantomime presume that the interactors can achieve joint attention (Tomasello 1999; Zlatev et al. 2005). The evolution of joint attention may be a central factor for why humans are, more or less, alone in having intentional teaching. Another factor related to mind-reading is that when the learner tries to imitate the demonstrated action, the teacher can also react with evaluative feedback, and, if necessary, with renewed demonstration.

“Conscious awareness” of the substages involved in a skill is required for a teacher to track the progress of the apprentice and provide corrections. Moreover, the cognitive machinery for informing and teaching must include the motivation to share the skill and the student’s motivation to learn. It could be claimed that this puts a premium on “honest information,” but the counter-claim would be that the teacher has been observed to successfully perform a skill and the students have been motivated to acquire the skill and can learn with no *general* notion of “honest information” involved (recall our §2.1 discussion of a “platform of trust”). The pragmatics of communication is thus an important driver of the evolution of teaching and remains crucial to the emergence of early and evolutionarily modern languages.

We emphasize the key difference between demonstration and teaching via pantomime by splitting POLE-4 into two:

- **POLE-4a:** Hominins built on complex action recognition and imitation with the development of *teaching* via *demonstration*.
- **POLE-4b:** Teachers came to also control the acquisition of certain skills via *pantomime*.

As explored in HOLO 1 (d’Errico et al. 2027 [this issue]), many of the emerging skills in hominins involve the making and use of objects, some of which serve symbolic roles. Thus, while there is no lockstep in the development of communicative symbols and symbolic objects, cultural evolution will engage both in an expanding spiral. The notion of *coevolutionary expanding spiral* is a key part of our EvoDevoSocio framework—(bio)cultural evolution in A can provide new opportunities for the evolution of B, and these in turn drive and benefit from further changes in A, and so the spiral continues as the growth of A and B continues. However, there is no implication of a one-to-one relation between changes in A and changes in B, nor a denial that such coevolution may be polycentric, with changes diffusing between groups and being further modified in the process.

2.3.2. Pantomime for Communication More Generally

Pantomime is not just for teaching. Pantomiming a request, command, warning, etc., can serve social coordination more generally. We define *ad hoc pantomime* as embracing the case where any pattern of movement is mapped “spontaneously” (i.e., without the ritualization or conventionalization discussed below) onto movements of arms and

hands (especially) and other effectors, without necessary dependence on the objects that constrain “real-life” performance of these movements during manipulation, locomotion, hunting or child-care, etc.¹⁰ We add the stipulation of *partial parity*—that others will be able to understand the intended message within its current context, though perhaps with limited success. This includes both *direct mimesis* (employing the same effectors as another human would in performing the pantomimed behavior) and *indirect mimesis* where this condition is relaxed—e.g., to use arms or the two hands to indicate the flapping wings of a bird or butterfly; to use two fingers to represent legs walking; to move the hand through a trajectory to represent the path taken by a person walking or a bird flying, and thus involves action recognition that is not restricted to actions in one’s own repertoire.

- **POLE-5:** *Ad hoc pantomime for communication*: Performing actions in the absence of objects to convey to others the need for an object or some associated behavior was coupled with the ability of others to recognize when a performance is an intended communicative pantomime rather than an attempted practical action. This is the root of open-ended communication.

POLE-5, which subsumes POLE-4b, marks a crucial point in the evolution of symbol-readiness in the sense of being able to freely create novel communicative acts while meeting, more or less successfully, the *parity condition* of shared meaning. Another form of pantomime involves moving the hands as if tracing contours of an object to represent the object—a skill that may be relevant to the later emergence of drawing (§5.2). Communication has become open-ended. Note how different this is from the specialized, but tunable, neural machinery required for the three alarm calls associated with leopard, snake, and eagle of the vervet monkey (Seyfarth et al. 1980b).

2.4. Ritualization, Protonarratives, and Protoconversations

A teacher who repeatedly teaches the same skill will come to develop ritualized pantomimes, rather than making ad hoc gestures on each occasion.

- **Ritualization**—an individual simplifies the form of a performance with another who (eventually) recognizes the communicative goal. Note that the other may have had to participate as a witness to the transitions involved to recognize the cumulative result.

We use the term *basic pantomime* to cover ad hoc pantomimes and their extension by ritualization, but without the process of conventionalization to be described in §3.1. In contrast to the sign and language systems we will discuss later, it should be noted that such ritualization may be primarily *dyadic* with the communicative goal shared by only two individuals. However, a teacher may use a sign gained in one dyad to teach it to another, perhaps at first using demonstration or ad hoc pantomime to convey its meaning. Unlike the ontogenetic ritualization specific to a great ape dyad, a teacher may employ the same basic pantomime for different students, and other teachers may choose to im-

TABLE 3. FIVE FACTORS DISTINGUISHING DEMONSTRATION, PANTOMIME FOR TEACHING, AND PANTOMIME FOR COMMUNICATION (Gärdenfors 2021)*.

	Demonstration for teaching	Pantomime for teaching	Pantomime for communication
Displacement	None	Minimal	Potential
Pragmatic form	Imperative (instruction)	Imperative (instruction)	Declarative
Represented actor	Demonstrator showing learner's action	Mimer representing learner's action	Mimer representing various roles
Perspective	Egocentric	Egocentric	Often allocentric
Pretense	None	Limited	Extensive

*However, since teaching involves a form of communication, aspects of the teaching column can be added to the communication column.

itate some of these communicative strategies. However, the parity principle does not apply—a ritualized pantomime performed by a teacher may be unlikely to be performed by a learner.

The teacher deploys a sequence of basic pantomimes over the course of a session, based on procedural memory for the skill and of associated pantomimes. With this we are on the verge of *protonarrative*: eliciting the sequence based not on the demands of current behavior but when recalling past events or in sketching a plan of action in coordinating with others for some upcoming activity.

- **POLE-6:** *Basic pantomime*—ad hoc pantomime and ritualizations that may be specific to dyads or small groups but are not conventionalized—could support protonarratives and protoconversations.

This can occur at the stage of ritualization where members of a dyad or a group can understand some of the other's pantomimes whereas others cannot. POLE-6 asserts that biocultural evolution has endowed the group with an understanding that basic pantomime can be used to communicate and thus that certain performances by others may be seen as pantomimes, so that the observer is to seek to understand the communicative goal. Crucially, someone who uses a (possibly ritualized) pantomime and is not understood has the option of creating another (possibly ad hoc) pantomime to help convey the intended meaning. The notion of “understanding” here would be far simpler than any philosophy of modern language would countenance. In §3.1 we discuss the process of conventionalization whereby pantomimes may yield the protowords of an early protolanguage.

Table 3 presents a possible evolutionary ordering (Gärdenfors 2021) in which *demonstration for teaching* precedes *pantomime for teaching*, which in turn precedes *pantomime for communication* where, by the stage of POLE-6, pantomimes may be ad hoc or ritualized. However, pantomime can support communication in many ways, with no necessary priority for the emergence of pantomime as a tool for teaching. Consider coordinating behavior in childcare or in a hunt, indicating where another might move

and what action they might take. Such uses of pantomime for social coordination could well have preceded or been contemporaneous with its use in teaching. All this sets the stage for the emergence of diverse “vocabularies” of pantomime from which protolanguages could later evolve.

In Table 3, the five factors concern different aspects of the actions performed by the demonstrator/pantomimer. *Displacement* concerns the extent to which entities relevant to the pantomimed actions are absent in the context. *Represented actor* concerns whether the communication only involves actor and receiver and objects present in the context, or whether the actor intends to represent also other individuals or objects that are not present in the context. In *Egocentric pantomime*, the reference is determined in relation to the miming person's own body, while in *Allocentric pantomime* the body of the mimer enacts another person or object located at another place.

The observation that pantomime can support protonarratives is an important addition to early contributions to MSH and the technological pedagogy hypothesis, and was inspired by the work of Ferretti and Adornetti (Ferretti 2021, 2024; Ferretti and Adornetti 2021, 2024) for whom the development of narrative and conversation is central to the origin of language. One aspect of this work is the claim that narrative actually precedes language. However, pantomime has limited expressive power compared to present-day languages. We thus suggested the term *protonarrative* for the simple narratives that could be expressed in pantomime or early protolanguages (Arbib 2024). Given this qualification, we can accept the slogan “Stories come first” [prior to language] of Ferretti (2025), while insisting that the complexity of stories and language developed through an expanding co-evolutionary spiral.

To illustrate this, here is a “just-so” story. A long-ago hunter returns with the carcass of an animal he has killed and wants to tell the tribe about the successful hunt. To do so in pantomime, he might perhaps alternate between egocentric pantomime, miming his movements stalking the prey and then stabbing it with his spear, and allocentric pantomime of the movements of the prey seeking to evade

him, then turning on him, and finally succumbing to his spear thrusts. This narrative works because the tribe can see that he has just completed a successful hunt and that the pantomime must be a re-enactment of the hunt and the actions of its two protagonists. Body posture and facial expression can indicate whether hunter or prey is represented at each turn. We can extend our “just so” story to include “protoconversation.” Consider a version of the hunter’s protonarrative instructing a group on how an imminent hunt is to be conducted. Some individuals might suggest alternatives, and these would be discussed back and forth.

Ferretti (2021) holds that narrative is supported by a triadic *cognitive architecture* grounded in mental time travel, mental spatial travel, and mindreading. The hunter’s first protonarrative is based on the recall of past events while his instructions for the upcoming hunt suggest future behavior, and so each involves some aspect of the form of displacement called “mental time travel.” We will place the three elements of Ferretti’s cognitive architecture in an evolutionary framework in §6.1, adding causal cognition to the triad.

3. FROM PANTOMIME TO PROTOLANGUAGES AND ON TO LANGUAGES

Pantomime opened up the semantic space for communication, expanding and exploiting the primates’ open-ended manual dexterity, although in many cases the dyadic relations involved did not exemplify the parity principle. We now move beyond the first six stages of POLE to complete POLE by incorporating the BtMH account of the final steps of biocultural evolution of language-ready brains and protolanguages they supported. But first we discuss the Great Move in primate brain evolution that provided the neural space for this transition to occur in early humans. The rest of the section emphasizes technosphere ↔ semiosphere interaction and then takes us to the stage of expanded protolanguages. §4 will then offer a Post-POLE account of the emergence of early and then evolutionarily modern languages, while §5 will examine some of the narrative powers that modern languages support.

3.1. FROM PANTOMIME VIA CONVENTIONALIZATION TO PROTOSIGN AND PROTOSPEECH

One of the challenges of discussing language origins is to dispel assumptions that modern humans take for granted. Even as a relatively young child, whether or not we understand that there may be other languages, we understand the notion of a language as combining words and that we can keep learning new words. However, at the stage we are positing now, humans may spontaneously pantomime (ad hoc pantomime) when they need to communicate with others. Unconsciously, if such situations arise repeatedly, such pantomimes become ritualized. We now explore how conventionalization may have provided the final step in creating a language-ready brain as humans became aware of the possibility of creating novel protosigns and protospeech to express new meanings and have them understood by oth-

ers. §3.3 will then carry our account forward on the path toward early languages.

Moving beyond basic pantomime, based on ritualization of ad hoc pantomimes, further biocultural evolution is required to pass from ritualization to conventionalization:¹¹

- **Conventionalization** occurs when *novel* gestures with similar communicative goals become shaped and shared by a community to yield a *protosign* whose performance and interpretation is then shared by that community (establishing parity for the community). Whereas basic pantomime need not involve imitation, conventionalization must involve imitation.

Although a teacher may ritualize the pantomimes during teaching to find versions that work best with diverse students, a protosign is typically a *social* construct involving a group of mutually communicating individuals. However, there is no need to suppose that basic pantomime comes to an end as conventionalization takes place.

Caldwell and Smith (2012) use experiments with modern humans to explore how graphical signs emerged over chains of participants when the most experienced members of a group were continually replaced by naïve participants. Signs were initially iconic but rapidly became symbolic in that they were mutually incomprehensible across experienced members of different chains, and new entrants needed to learn conventionalized meanings. Moreover, signs became progressively simplified over successive usage. This provides critical support for the view that human communicative symbols could have evolved culturally from iconic representations, but conventionalization in a human society (as distinct from a group in a lab) may inhibit further changes by new members.

Żywiczyński et al. (2021) speak of pantomimic fossils in modern human communication, exploring circumstances in which people bootstrap communication by means of pantomime, stressing that when these circumstances persist, conventionalization occurs so that newly emergent pantomimic communication becomes increasingly language-like. However, our concern is with the process of conventionalization before there is any language to be like.

Ambiguity in pantomime may have provided further selective pressure for coming up with “protosigns” as conventionalized forms with restricted ranges of meaning in certain contexts. In one anachronistic example from ASL (American Sign Language), a pantomime of an airplane flying (hand shaped like a plane and “flown through the air”) became conventionalized in two different ways—with the handshape moved back and forth on a short trajectory, it means AIRPLANE, whereas with the conventionalized form of a single traverse of the long trajectory it means FLY.¹² Note that this conventionalization could have been the other way round. While no airplanes flew when early hominins first developed protosigns, the principle exemplified here already applied. As a pantomime becomes familiar to a group, it becomes conventionalized—which may involve simplification, but may also involve arbitrary changes to distinguish different aspects of the original cloud of meanings—and thence may become a symbol rec-

ognized only by members of the group. With this, we present POLE-7 via two means of expression:

- **POLE-7a:** *Protosign (Bodily Symbolic Gesture)* emerges as a manual-based communication system, which includes (but is not restricted to) pantomimes that have become conventionalized within communities to allow meanings to be conveyed more economically and less ambiguously.

Recall that nonhuman primates have flexible manual control but limited vocal control—a major motivation for linking manual control to the earliest stages of changes that in due course supported a language-ready brain. Specifically, then, BtMH posited that the next stage involved the evolution of flexible vocal control.

- **POLE-7b:** Nonhuman primates have very limited vocal learning. *Protospeech (Vocal Symbolic Gesture)* may rest on the “invasion” of the vocal apparatus by collaterals from the communication system based on manual control and recognition.

Together, these constitute one sub-analysis of:

- **POLE-7:** The emergence of early *protolanguages*, of being able to freely create novel communicative acts while meeting the parity condition. At this stage, *protowords* may be available in either modality, signed or spoken.

The notion of a protolanguage here is not only that it is an open-ended system of communicative symbols (protowords) but that it is an explicit part of cognition—that humans with a protolanguage understand that this is a system separate from, e.g., facial gestures and basic pantomime that they can learn to use, even if few venture to create new protowords.

As we saw in §1.2, the relative role of manual and bodily gestures versus vocal gestures in the origin of language remains controversial, but we recall the earlier observation that, despite the engagement with different sensory and motor systems, Broca’s area plays similar roles in the compositionality of spoken and signed languages (Corina et al. 1992; Marshall et al. 2004). Despite privileging POLE-7a, BtMH does *not* claim that humans had signed languages before they developed vocal control, but rather that protosign opened a communicative niche for a diversity of conventionalized gestures to express meaning, thus providing a selective advantage for further evolution of vocal control and learning to expand the range of available gestures while adding sound symbolism to the communicative repertoire. In other words, it posits another *expanding spiral* linking the biocultural evolution of protosign and speech (Arbib 2005b).

What is crucial for us is that there was an early manual component (as suggested in POLE-1 through POLE-6) that provided a grounding phylogenetic link between action and communication, just as today there is an ontogenetic link that in modern infants sees action gestures as dominant in the first 18 months of life in supporting the acquisition of words, whether in signed or spoken language (Carpici and Volterra 2008; Iverson and Thelen 1999; Volterra et al. 2018).

As protolanguages emerged, they would have been of use in diverse aspects of social coordination (hunting, teaching, childcare, food preparation, and more). Our hypothesis is thus that early humans with different responsibilities within social groups may have developed *micro-protolanguages* whose conventionalized protowords were specific to communicating about their own activities (while still relying greatly on context) long before the emergence of larger protolanguages shared by the whole community, in part through integrating portions of these diverse micro-protolanguages.

Given the diverse skills that were perhaps in place by the time of the Acheulean technology, and given that today’s student of a skill might become tomorrow’s teacher, genes that contributed to improved teacher-student dyads might have become favored across the population, while advances in pantomime and protolanguage repertoires would have contributed to the separate advancement of a variety of skills. One can imagine that traits for mental adaptability and emotional stability (e.g., patience) might have been part of the mix. We do not titrate selective advantage between biological and cultural change, but suggest as a basis for further discussion that the steps that led to the language-ready brain involved a strong biocultural involvement whereas thereafter, human societies were able to develop on the primary (but not exclusive) basis of cultural change, with the Socio coming to predominate over the Evo in the EvoDevoSocio cycle across the generations.

We next examine how POLE helps us understand the linkage of language-readiness to material praxis. Then, in the Post-POLE analysis that follows, we use examples from both manual and vocal domains to show how protowords led to words and constructions (§3.4) and then how these features of language ramified on the path from expanded protolanguages to evolutionarily modern languages (§4).

3.2. LINKING LANGUAGE-READINESS TO MATERIAL PRAXIS

The grammar of a human language comprises a set of *constructions* in the linguist’s sense of “tools for putting words together” (§3.3) but the archeological record reviewed in HOLO 1 (d’Errico et al. 2027 [this issue]) invited us to consider *constructions* in the general sense that includes transforming and assembling material objects. A crucial element of HOLO is that, because our theory of language evolution is strongly linked via complex imitation and pantomime to manual praxis, a language-ready brain *in our sense* can support constructions in both senses (Arbib et al. 2023)—an inference that is not readily supported by theories that address the evolution of spoken languages in isolation. Here, then, we further argue that “the language-ready brain” was indeed a *culture-ready brain* (Whitehead 2010).

Recalling Merlin Donald’s (1991) stage 3 (§3.1), the introduction of *external memory* storage and retrieval, and a new working memory architecture, we briefly consider the emergence of drawings and, much later, writing. to ask: “What must be added to a brain that can pantomime the shape of an animal (as well as hunting and butchering it) to

get a brain that can draw the animal?” By running a hand around the real animal, butchering it, or moving a finger so that it looks as if it is tracing around it, we can trace its contours in three dimensions and thus pantomime the shape of the animal. The breakthrough to drawing was the ability to go from this evanescent tracing of contours in three dimensions to the more-or-less enduring marking of contours on a two-dimensional surface. This may well have emerged without archeological trace in scratching contours with a stick in sand or mud, but the cognitive transformation from making a mimetic gesture in mid-air to an *enduring* mark on a surface *cannot occur without the availability of a supporting technology*, such as using a burnt stick to mark a cave wall (see §4.4 of HOLO 1).

The making of abstract patterns of possible communicative value provides the precursor of the representational use of ochre as part of a technology for creating enduring symbolic representations (Davidson 2020; Davidson and Nowell 2021). The behavior, including the relevant tool use, is clearly within the range of creatures whose brains have a flexible capacity for pantomime, yet involves cultural evolution to yield communities that can use drawings as a memory system to expand the reach of their interactions. The two related issues are (i) the motor skill and its technological support, and (ii) the representational intention. We note, but do not develop, the related concern with the development of 3-dimensional representations in the form of figurines (see HOLO 1, §4.5).

As emphasized by Stibbard-Hawkes (2025), ethnographic evidence from contemporary hunter-gatherers shows that cognitively complex populations may leave only sparse or ambiguous traces of symbolic or non-utilitarian activities, due to the widespread use of perishable materials. Stone tools dominate the archaeological record, now dating back before *Homo* (HOLO 1, §3.3). The absence of records of other behaviors in the deep past should therefore be treated with caution, as it may reflect preservation biases rather than genuine cognitive or cultural limitations. It may not have been until the Acheulean that the type of pedagogy suggested in §2.2.1 might have arisen. However, recalling our earlier discussion of *micro-protolanguages*, it would be mistaken to limit the use of pantomime and protolanguage to tool making. Presumably, when teaching first arose, it must also have transmitted crucial life experiences, including observations of flora and fauna, the tools and stances required for hunting, the sociality of communal dancing and the like.

Consider the ancient uses of ochre by hominins (§4.1 of HOLO 1) that only much later would serve as a complement to the pantomimic exploitation of contours. Observation of the bright plumage of male birds in their mating rituals may have encouraged the use of ochre to decorate members of a group during rites of passage. By 140 Ka there were human groups in Africa that used ochre for coloring various ornaments supporting symbolically significant ritualistic practices and social communication. Body painting does not in itself allow a long-term storage of exosomatic information but does exemplify how the evolution

of means of communication has been gradual, with new words and constructions of language emerging to complement the increasing sophistication of social interaction, practical skills, and symbolic objects.

Although Donald (1991) posits that external memory emerged with the Upper Paleolithic in Europe, instances of external memory systems have been sporadically found in Africa and Europe dated as far back as 44 Ka (d’Errico and Colagè 2018; d’Errico et al. 2018). Nair (2024) sees finger flutings on cave walls by young children circa 13,000 BCE as *archewriting*, anticipating the important cultural innovation of writing. Archewriting may also have served as a precursor for the development of aesthetic criteria such as linearity, color contrast (compare the use of ochre in body markings), flowing designs, crisscross patterns, especially since there appear to be incipient “teaching” as well as emotional elements involved with children being communally “encouraged” to do their finger-flutings on the upper surfaces of caves. But writing was then still far in the future so that certain people were tasked with becoming the “external memory” for spoken narratives of special power to the community. The narratives of Homer developed within an oral tradition before being written down, and relied on people of exceptional memory to tell and retell the stories.

The oldest reliable examples of writing systems date back only to around 3400 BCE in Mesopotamia, with other notable occurrences at 3300 BCE in Egypt and 1800 BCE in Greece, while the inscriptions on bones and shells (called “oracle bones”) dated from around 1400–1200 BCE laid the basis for Chinese writing (de Voogt 2024). Mayan writing systems emerged much later. Cuneiform arose as a system for keeping accounts and only later became expanded to record narratives. In many cases, pictograms that conveyed the content of existing words became adapted to capture something of the sound structure of a spoken language, but with diverse approaches to this in current writing systems. The Mediterranean, Chinese, and Mesoamerican systems emerged independently and with very different characteristics—a dramatic example of polycentric cultural innovation that yields multiple approaches to a given problem.

The manufacture and use of knapped stone and bone tools, which require the coordination of vision and haptic feedback, likely had important implications for early communication. Such activities within a community may have involved both (proto)language communication and non-verbal signaling with a gradual shift toward the former as technology became more complex. Similarly, the use of ochre required prehistoric humans to intentionally leave their caves to first gather and then grind down ochre and other natural materials, make them into paste, possibly using resin and water, before actually painting. All this would have required dedicated labor and the allocation of temporal and brain/mental resources spent *not* on hunting/gathering but on a different sort of activity—namely, empathetic sociality and inference-making. We regard this as a form of “symbolic survival” that is crucial to the development of the human niche. The downtime of hunter/foragers created the opportunity for cultural evolution to yield

activities that people take pleasure in, whether or not they make a positive contribution to survival, but (for good or ill) promote tribal identity (Nair 2012).

3.3. FROM PROTOLANGUAGE TO WORDS AND CONSTRUCTIONS

POLE-1 through POLE-7 involve biological evolution (Evo) of the developmental mechanisms that shape brain and body (Devo), though cultural evolution (Socio) is needed for a group to learn to exploit these abilities, in turn constructing a socio-cognitive niche for further evolution (Iriki and Taoka 2012). POLE-7 asserted that these stages yielded a “*language-ready*” brain, but that early humans had only *protolanguages* (a limited vocabulary of protowords, signed or spoken, with little or no grammar). The emergence of *expanded protolanguages* and then *early languages* (with extended lexicons and grammars) thereafter involved primarily cultural, rather than biological, evolution among humans with language-ready brains. Nonetheless, we will argue (§6.2.2) that the notable changes in anatomy from *Homo erectus* to successor species like *Homo sapiens* may have converted a “*ready-for-complex-protolanguages*” brain to a “*ready-for-evolutionarily-modern-languages*” brain, in part through changes in brain size.

Additionally, although a few humans are incapable of learning or using a full modern language (whether spoken or signed) and some are more fluent in their language use than others (whether due to genetic make-up or a beneficial environment) and despite huge variations in the genome of human populations across the globe and genetic variations within those populations, essentially all humans are able to acquire and use any language that can support their behavior in their physical, social, and symbolic environment. We reiterate that the emergence of diverse languages is part of the polygenesis of symbolic culture presented in HOLO 1 (d’Errico et al. 2027 [this issue]), involving distribution of innovations across space and time.

- **Post-POLE.** With the emergence of recent human lineages, cultural evolution predominated as languages emerged from protolanguages, with not only a widening of the lexicon but also the emergence of *compositional semantics* supported by grammars for combination of words to express and comprehend an unlimited range of novel meanings.

We call this Post-POLE to emphasize that it holds whether or not the protowords involve speech or bodily gesture. Today’s humans bear the marks of cultural evolution that progressed through Post-POLE in polycentric fashion all the way along the continuum of Table 1 (no sharp boundaries) to the diversity of today’s human languages (and many others that have become extinct). The use of protowords in teaching a basic skill or recounting a simple protonarrative need have no grammar as such, but simply the ability to emit protowords that correspond to events or actions in chronological sequence. Nonetheless, the definition of simple protolanguages (protosign and protospeech) offered by POLE-7 does not exclude that a protolanguage may have limited compositionality. But

how do we get from the “*protowords*” of protolanguages to the words and constructions of languages? Speculation about the early stages that led toward Early Languages will engage us in the present section. The progression from there to and within Evolutionarily Modern Languages will be the focus of §4.

Bickerton’s (1995) *compositional view* of protolanguages is that protowords were akin to nouns and verbs, possibly with a few strung together but with no syntax to organize them. Following Chomsky’s (1995) emphasis on “*Merge*” as an operation putting two words or phrases together, language evolved for Bickerton (2009) by “*just adding Merge*” to build syntactic structures that integrated these words, though he later replaced Merge by Attach, a combinator that builds in semantic conditions (Arbib 2025; Bickerton 2014). By contrast, BtMH adopts the *holophrastic view* (Arbib 2005a; Wray 2000), positing that the earliest protolanguages employed meaningful “*protowords*” without any internal semantic structure or constructions to combine them. These protowords were *holophrases* (aka *unitary utterances*), many of them based on pantomimes, that capture events such as an action being performed on an object or another person, and are thus akin in meaning to phrases of modern languages. It is a mistake to see holophrasis and Bickerton’s view as in simple opposition, but here we focus on the holophrastic view of how words and constructions co-evolve.

We are positing here a process before words in the modern sense exist. Recall from §2.2.2 (POLE-3) that complex action recognition supports the ability for the observer to *fractionate* a behavior into *apparent* subactions, and then approximate the performance on the basis of observed movements, with practice yielding increased skill to perform a new action in varied situations. To see how this might have applied to the transition from holophrases to words and constructions, consider pantomimes for *open-a-door* and *close-a-door*. Neither contains an explicit component for *door*. But what if an observer, having the general idea of protosigns and doors, seeks to fractionate the performance to seek the (not yet existent) protosign for “*door*”? A possible solution (Arbib 2012: Chapter 10) is that, because it is common to the pantomimes for *open-a-door* and *close-a-door*, the pantomime of turning the handle becomes fractionated out as the protosign for *door* – *even though it did not originate as a separate pantomime for door*. The two directions of moving the hand yield protosigns for *open* and *close*. It is complex action recognition and imitation (recognizing variants of an action as ways of combining subactions in flexible ways that achieve subgoals under varying circumstances) and thus fractionation applied to communication, that makes this possible. Moreover—and here we sow seeds of the compositional/holophrastic reconciliation suggested above—the existence of even primitive constructions might scaffold the invention of new (proto)words, for example, a rapid opening and closing of the mouth might become adopted as the protosign for *mouth* that could provide a new X to be available to fill the slot in *open-X* and *close-X*. This example marks an important elision—from pantomime of

manual gesture to pantomime of an oro-facial gesture. All this is in contrast to the “language fossils” to be described in §4.1 that survive into modern languages despite not being linked to constructions.

In imperative form, a pantomime does not need to specify the agent who is supposed to perform the action that is pantomimed. However, when pantomime takes a declarative function, there is often a need to specify *who* is the agent of the communicated action and sometimes also *what* is the object acted on or *where* it is performed (Brown et al. 2019). Considering pantomimes as one class of gestures, a second class includes pointing (Tomasello 2010). The characteristic references for pointing gestures are locations or directions (and see the earlier discussion of joint attention—the notion is that pointing goes beyond “look in that general direction” to “pay attention to whatever in that direction is relevant to our interaction”). A third type of gesture refers to objects or object properties such as shape, size, length, height, and depth that can be signaled by a complementary use of pantomime. Gärdenfors (2017) sees these three categories as *proto-verbs*, *proto-demonstratives*, and *proto-nouns/adjectives*, respectively, and posits that once these protoword classes are in use, it is natural that gestures are combined and thereby more complex messages can be communicated. However, BtMH argues that syntactic categories are a later generalization. In *open-X*, the “slot-filler” *X* must meet a *semantic* criterion of being something-that-opens. We are a long way from generalized constructions of modern languages whose slot-fillers are defined by syntactic categories like “noun” or “verb,” but even at this early stage, we see that protowords are beginning to become words since they are now enriched by their linkage to constructions. As “protowords” were *fractionated* or elaborated, they yielded words for constituents of their original meaning. In a complementary process, *constructions* developed to arrange the words to reconstitute those original meanings and (the advantage of this transition) many other meanings beside. Moreover, not all present-day constructions have slot fillers that are syntactically defined. For example, “The woman in red” gets its meaning from a very specific “*X* in *Y*” construction with *X* required to denote a person and *Y* denoting the color of their garment.

On this account, brain mechanisms that evolved to support the imitation of manual skill and later the development of teaching were “lifted” from circuitry involved in praxis to circuitry involved in symbolism to support essential stages of transition to expanded protolanguages and early languages. However, pantomime is not the only source of meaning, and fractionation is not the only path to words and constructions. For example, the process could have been one where the meaning of the proto-unit was reduced to one of its constituent parts as a result of frequent use, with the effect that a holistic unit meaning “open the door” came to acquire either “open” or “door” as its new meaning. And, as we saw in the ASL example of PLANE versus FLY, possibly arbitrary changes in an initial proto-

word could yield different words to exhibit different aspects of the initial meaning.

BtMH suggests that as protolanguages became more complex, they employed many ad hoc constructions, but with little in the way of “generalized slot fillers.” Subsequently, generalizing across similar constructions would yield more general constructions, while generalizing across the slot-fillers would yield increasingly general categories, passing from the semantic to “semantico-syntactic” and, after much further cultural evolution, yielded the diverse syntactic categories of modern languages. We further suggest that such processes are enough to support the grammaticalization that formed evolutionarily modern languages (charted in §4.1) *once there were enough words and simple grammar to support it*. The claim is then not that grammaticalization arose as a completely novel mechanism, but rather that sufficiently complex Early Languages had to emerge to support the type of grammaticalization seen in Evolutionarily Modern Languages that can be inferred from the historical record.

A process like the fractionation of holophrases would also have yielded, piecemeal, the phonology of a (proto) language, i.e., the regularization of the articulatory components into a limited set of meaningless elements such as phonemes or syllables in speech, and position, hand shape, and trajectory in sign. Sandler and Aronoff (2007) offer relevant observations on the cultural evolution of ABSL, the Al-Sayyid Bedouin Sign Language. Chapter 12 of Arbib (2012) discusses the emergence of Nicaraguan Sign Language, NSL (Kegl et al. 1999; Polich 2005; Senghas 2003) as well as ABSL, noting the contribution of ambient language use, culture, and community on the rapid emergence of these languages. More recently, Mineiro and Moita (2024) report the roots of Sao Tome and Principe Sign Language in pantomime. In each case, the signed language develops in contact with people who already use an evolutionarily modern language, whereas our assessment of how protolanguages first emerged from pantomime (and vocal gestures) must assume no such contact.

3.4. A Brief Look at Metaphor

In evolutionarily modern languages (EMLs), metaphor typically transfers meaning from one conceptual domain to another (Black 1962; Hesse 1988; Indurkha 1992; Lakoff and Johnson 1980; Ortony 1979; Searle 1979). An example attributed to Aristotle (4th century BCE) is: “The ship ploughed through the sea.” This sentence has the analogical pattern: “the ship is to the sea as a plough is to the field.” Here the missing term is “field.” Metaphor is often considered a hallmark of modern languages, especially in its use in literary and poetic expression. However, metaphor seems “inevitable” even in protolanguages. There, in a novel context an early human may not have had a proto-sign or word in their vocabulary that expressed what they wanted to say. Instead, they might “analogically” employ an expression from another context that expressed much of what they wanted to convey—with the expectation that

the recipient will not be misled into “importing” features from that other context that do not fit the present context. As a parallel, Elbers (1988) finds “overextensions” of existing vocabulary among children—a child might say “sticks of the chair” instead of the “arms of the chair” or “ears of the plane” instead of “wings of the plane” (both examples from Nair 2022). Of course, the crucial difference is that the child’s language community already has the related word, whereas the protosigner’s community does not.

Ellison and Reinöhl (2024) argue that metaphor may have driven the transition from protolanguage to full syntax. Without metaphor, language remained concrete; with it, abstraction, and possibly grammar, emerged (Arbib and Cuccio 2026). Xu et al. (2017) computationally analyzed 5000 or so metaphorical mappings over 1000+ years of English and found that metaphors tend to map from embodied/concrete to abstract/internal domains. Mangiaterra et al. (2026) similarly use AI models to track metaphor difficulty and structure across centuries, finding that metaphor evolves differently across genres and historical periods, which in turn indicates that modern metaphor evolution is shaped by literary and communicative cultural ecosystems that did not exist at the protolanguage stage.

4. THE EVOLUTION OF GRAMMAR

We have now sketched the early stages of the co-emergence of words and constructions that yield protolanguages of increasing complexity. In §4.1, we consider how this process continues in the emergence of early languages and on to evolutionarily modern languages. Observations on linguistic change suggest that the stages were not discrete; rather, there must have been gradual transitions from one stage into the next in different communities and at different times. Data from historical and comparative linguistics has shown how grammaticalization has complemented expansion of the lexicon in shaping evolutionarily modern languages.

An important challenge is to understand how (proto) languages made the move from requests, commands, and protoconversation in the here and now (or next) to narratives that were divorced from “veridical” accounts of the past. Myths and just-so stories, etcetera, are often divorced from truth conditions, though some theories of narrative posit coherence and consistency conditions. §§ 4.2 and 5 will begin to explore the way in which language evolution depends not only on modifying grammar, but also on finding ways to express thoughts that extend the semiosphere to actions, perceptions, emotions, wishes, and so on that are related both to the experience of real worlds and the imagination of fictional worlds. Such efforts challenge our understanding of languages and their evolution in a way that is unsupported by, for example, Fodor’s (1975) language of thought which pays no attention to the open-ended richness of our experience. No words can exhaust our experience of sunsets.

4.1. FOUR OVERLAPPING STAGES IN THE EVOLUTION FROM PROTOLANGUAGES TO LANGUAGES

We now consider further the progression of Table 1:

1. (PL): Protolanguage (“Simple Protolanguage”)
2. (EPL): Expanded Protolanguage
3. (EL): Early Language (“Simple Language”)
4. (EML): Evolutionarily Modern Language (“Complex Language”)

One dimension for construction of protolanguages in the vocal domain (i.e., from the emergence of protospeech onwards) rests on observations made in present-day languages, which exhibit elements strongly suggestive of “linguistic fossils” (Jackendoff 1999, 2002). Here we give spoken examples, whereas we have seen that Żywicznyński et al. (2021) offered examples of *pantomimic* fossils in modern human communication. Such fossils, called interactives, are essentially grammarless (holophrastic). They are syntactically and semantically independent forms, not being part of a sentence, and are normally accompanied by gestures involving the hands, the head, and/or the entire body. They include ideophones used for depicting events and states (cf. English *bang*, *boing*, *thud*), interjections for expressing emotive states (cf. English *wow*, *ouch*, *yuck*), attention signals (cf. English *hey*, *oi*, *yo*), response signals (cf. English *mhm*, *uh huh*), and response elicitors (*eh?*, *huh?*). The study of such linguistic fossils suggests that protolanguages included indexical interactive forms for elementary social communications (Heine 2023) separate from the emergence of constructions. This is a crucial addition to the repertoire considered in §2 and §3 in the path from protolanguage towards early languages. These examples remind us that the back and forth between two or more people requires further communicative tools, especially in support of conversation, but their very presence as postulated in protolanguages offers further seeds for the growth of language.¹³ A simple *eh?* or *uh?* does signal confusion but does not convey what caused confusion in the prior utterance. This lays the basis for adding more explicit questions to the conversational stock of early languages and the later development of existential questions that lie at the heart of science, philosophy, and religion.

While the linguistic forms (protowords) of protolanguages must have been used to form utterances on their own, the development of Stage 2 of “expanded protolanguage” (EPL) involved the limited ability to combine various forms through primitive constructions. Conventionalized sequences of interactives yield a kind of “rudimentary coordinating syntax” where a change in order can change the overall meaning, as in “John-Hit-Bill” versus “Bill-Hit-John,” where no grammatical markers have yet emerged. Jackendoff (1999) called this pre-syntactic, claiming it relies on pragmatic principles (heuristics) like “agent first.” It is debatable where one wants to say that syntax starts (we are still in the world of porous boundaries), but the

point is that the notion of “agent” is a semantic category, not a syntactic category. Moreover, the general notion of “agent” may be relatively late. Our discussion of early constructions showed that different classes of actions may have different constructions with slots for different types of agents—the “semantic verging on syntactic” generalization is itself an emergent. As PLs became EPLs, we suggest, few if any grammatical markers had yet emerged. Rather than “syntactic rules,” ordering of protowords must have been based on the pragmatic principle of iconic coding, where the sequence of coding immediately reflected the order in which the event described took place in real life. Thus, the ordering in “John-Hit-Bill” reflects the fact that there was a speech participant (“John”) engaged in an action (“Hit”) with the effect that another participant (“Bill”) was subsequently affected.

In addition, as already noted in §2.3, just as we may string sentences together to create a simple narrative preserving temporal order, so might protonarratives have been formed by stringing such protophrases together, preserving the way in which a pantomime could offer a protonarrative. It would require later grammatical features to keep track of a narrative when temporal order was not preserved, as supported by the discussion of pantomimed narratives in §5.1.

In §3.3, we examined competing hypotheses about how words and constructions might have first emerged, with the process accelerating perhaps over the last 160 Ka to 60 Ka during the development of early languages (ELs). Whatever the differences, it seems clear that eventually, these emerging languages would have had words to describe objects (as the basis for a syntactic category of nouns) and words to describe actions (as the basis for a syntactic category of verbs), and so we may assume that categories akin to *noun* and *verb* began to emerge in the *expanded proto-languages* of Stage 2. But note the word “began,” since §3.3 argues that at the early stages, no general syntactic notion of “noun” or “verb” would exist—but rather semantic categories corresponding, rather, to objects of a particular type, and similarly for actions. However, the syntactic boundary between the words treated like verbs and those categorized as nouns and the extended sets of words for “objects and other entities” and “actions and other processes” admitted to these syntactic categories varies greatly between present-day languages (Kemmerer 2005).

We now turn from the *longue durée* to consider how the diverse syntactic categories of current human languages may have emerged, taking place during Stages 3 and 4. On the basis of comparative linguistic evidence we can assume that around 15 to 10 Ka, and probably earlier, languages had semantic and syntactic structures (but not, of course, the lexicons) that sufficed for any cultural and technological innovations in the lead up, at the latest by 10 Ka, to languages that in their general forms may be essentially indistinguishable from languages as we know them today (Heine and Kuteva 2007, 2011).¹⁴

Early languages (Stage 3) may have been the first where language evolution made extensive use of *grammati-*

calization—that is, the emergence of grammatical forms and constructions from a prior way of putting words together. However, the process continues in EMLs. Noting the diversity of constructions and attendant categories of slot-fillers across the world’s languages today (Haspelmath et al. 2005), we can speculate that precursors of today’s broad and diverse language families may have been in place toward the end of Stage 3, but evolutionarily modern languages must have developed further diversity, grammaticalizing into analytic languages in some parts of the world and in other parts, into synthetic or polysynthetic languages. Analytic languages are suggestive of a limited amount of grammaticalization. Chinese, for instance, is an analytic language (note, however, that in its early stages of development, Chinese was a largely polysynthetic language and it was over the centuries that it lost its morphosyntactic structures, now being an analytic language). Polysynthetic languages, in contrast, are the result of extensive grammaticalization, such as Navajo, whose verb morphology can combine subject, object, aspect, mode, valence, and direction. Besides, in Navajo, it is possible to incorporate nouns into a single verbal complex, for example, expressing “I am dragging a boy” in a single word (Beniamine et al. 2017).

Moreover, language change can be stimulated by language contact. English results from changes following a Germanic language (Anglo-Saxon) receiving a strong overlay from a Romance language (Norman French), expanding the lexicon and bringing somewhat different grammatical systems into contact in a manner that provides rich material for further change that may simplify or complexify the grammar. Similarly, the study of pidgins and creoles has offered rich insight into the impact of different patterns of language contact in supporting language change (Adone and Gabel 2025). Such interactions have been in operation from the earliest times of protolanguage complexification and continue today.

Grammaticalization had the effect that, for example, in combinations of two independent nouns, like “dress” and “orange,” one assumed the function of a modifier of the other, turning into an adjectival modifier of the other (“orange dress”), eventually giving rise to new word categories, like adjectives in the present example. This strategy was extended further at Stage 4, when combinations of any kind of linguistic units gave rise to new modifying categories. For example, in combinations of two independent sentences, like “I went to bed” and “I was tired,” one developed into a modifier of the other, resulting in clause subordination, as in “I went to bed because I was tired,” where the former is the main clause and the second a subordinate clause. The overall effect of such processes of grammaticalization was that language acquired a hierarchical syntactic organization, the transition being one in degree rather than of kind.

The mechanism central to grammaticalization at both Stage 3 and Stage 4 is *context extension*, having the effect that the uses of various forms were extended to new contexts where specific semantic features were highlighted while other features were suppressed (Heine 2002; Heine

et al. 2013; Heine and Kuteva 2002)—and recall our note on metaphor (§3.4)—suggesting that grammaticalization exploits early processes in new ways once EPLs/ELs offer enough structure for context extension to act on phrases and longer utterances. Research across corpora of diverse languages led to charting the emergence and development of grammatical forms and constructions across space and time and on this basis to infer ancestral forms for present-day languages (Heine and Kuteva 2007).

One important cognitive strategy underlying context extension is that words or constructions having concrete, physically perceptible meanings are gradually extended to also express more abstract meanings, including grammatical functions (Heine 1997). Diverse languages across the world have grammaticalized their verb for “have” into a grammatical marker. In doing so, they used this verb in contexts where its actual meaning of physical possession is suppressed and new meanings are highlighted. This has also happened in the history of English when the verb “have,” originally referring to physical possession as in “Jeff has a new coat,” was extended to non-physical complements, as in “Jeff has no mercy,” and finally to verbal complements, like “Jeff has come,” where “have” no longer expresses verbal meaning but rather a grammatical function, giving rise to the English perfect aspect.

Grammatical and other linguistic changes do not normally lead in discrete steps (or “leaps”) from one stage X to another Y but exhibit an overlapping structure, and these changes are described by means of the *overlap model* of Heine (2002) whereby the change from X to Y involves an intermediate stage where X and Y co-occur, being used side-by-side, before in the final stage Y becomes the main option: $X \rightarrow X/Y \rightarrow Y$. The overlap model may also account for a multimodal approach from manual to vocal language evolution. It helps to explain the transition from bodily gesture to vocal communication—a gesture (X), like a nod, is perfectly alright to express consent or agreement. But one may combine the nod with a vocal expression (X/Y), like *Yes!*, before finally relying mainly or only on the vocal expression (Y).

We see that combining or employing existing elements in new ways for novel purposes is an elementary but pervasive human strategy on which the linguistic changes captured by Table 1 are based. Accordingly, such changes need not have happened only once in hominin development. Presumably, somewhat similar paths from protolanguage to languages as we know them today took place independently in different parts of the world. While the motivating factors and overall properties of grammaticalization were presumably of the same kind everywhere, the pathways and outcomes differed from one community to another in a number of ways, in particular in which linguistic material was affected by it and in how extensively it was applied. Such differences can be held responsible for much of the linguistic diversity observed in modern language families, consistent with the view of polycentrism espoused in HOLO 1 (d’Errico et al. 2027 [this issue]) and in §3 above.

Our proposal in Table 1 reconciles recent findings about interactive grammar (Heine 2023) and grammaticalization (Heine and Kuteva 2007) with other theoretical proposals that support the timeline for the emergence of symbols within Simple and Expanded Protolanguages with *Homo erectus* and more complex grammars and Early and Evolutionarily Modern Languages emerging with later *Homo* populations. It is also compatible with recent research identifying multilingualism rather than monolingualism as playing a central role in language evolution. This is compatible with a multi-source model of human language genesis based on the insight that languages most likely evolved not in a single, monolingual, isolated community but rather in a multilingual setting, whereby evolutionary innovations made by different small-size groups at different times and places could have been shared through bilingual individuals (thanks to a combination of small societal scale, trade, and out-marriage) before gradually being brought together in a more powerful combination (Evans 2017).

4.2. FLEXIBLE CONSTRUAL AND GRAMMATICAL COMPLEXITY

In §3.4, we examined how metaphor as a literary strategy in evolutionarily modern languages probably had its roots in the attempt, starting with simple protolanguages, to adapt the meanings of (proto)words in one domain to cover somewhat related meanings in other domains while, at a more general level, §4.1 charted how the emergence of evolutionarily modern languages provided an expanding lexicon and grammar that helped support language to engage with, and help drive, an ever-expanding semiosphere and technosphere. Here we expand on this by considering the impact of conversation and impactful narratives in expanding grammar. Both require the speaker to adopt and speak from different perspectives; to know when to express known and unknown information, when to leave it to the audience to infer it, and to craft a coherent string of utterances reflecting the development of a shared understanding of what is told. Here, the perspectives of protagonists, antagonists, storyteller, and audience are communicatively enacted. Moreover, intergenerational narratives influence identity development, through building the individual’s narrative capabilities for autobiographical memory and a group’s collective memory (Nelson and Fivush 2020). In addition, we may tell stories *for their own sake*; narratives not only support diverse human cognitive and communicative accomplishments, but are, in themselves, fundamental to what it means to be human. The expansion from fact to fiction and from the “real” to the imaginary is what distinguishes modern narratives from the protonarratives of §2.4 though even there, the hunter pantomiming the behavior of himself and his prey may have yielded to the temptation to embellish his prowess (though general notions of “prowess” and “embellishment” lay far in the future).

The complexification of human society both motivated and depended upon the complexification of language, and in particular the evolution of the capacity for the *flexible*

construal (complementing *context extension*) that, in describing a scene or situation, may “lift” selective attention from vision to language, a sophisticated descendant of gestural pointing as a complement to protolanguage. For example, we can employ the active versus passive voice to direct attention to the subject, distinguishing two descriptions of the same scene, as in

The boy kicked the ball

versus

The ball was kicked by the boy.

While formal semantics would regard the two sentences as having the same meaning, cognitive approaches to grammar by Langacker (1987) and semantics by Talmy (2000) build on the way in which such expressions refer to the same event in different ways, foregrounding different participants. Here, meaning is a matter of *linguistic conceptualization*, in which the speaker (the conceptualizer) selects a particular *construal* of the referential situation. We thus emphasize the active role and the perspective of the subject (Verhagen 2010) in semantic analysis. Another form of flexible *construal* (Sinha 2010; Talmy 1978) highlights spatial Figure-Ground relations:

The desk (Figure) is **under** the lamp (Ground)
The lamp (Figure) is **on** the desk (Ground).

The Figure is the entity that is most perceptually salient. The Ground is, in many cases, part of the background knowledge that is already shared by speaker and hearer, from which the speaker selects in order to provide the hearer with relevant new information. The entity that the expression is primarily about is designated, in functional analyses of language, the *topic*. The new information provided by the speaker is designated the *comment*. Topic and comment are the basic constituents of the *informational structure* that motivates grammatical choices (Chafe 1987; Krifka 2008). Metaphor and flexible *construal* can both be related to the human capacity for *conceptual blending* (Fauconnier and Turner 1998), a cognitive process that selectively combines elements from two or more mental spaces to create an emergent “blended” space with structure and meaning not present beforehand. We see this already foreshadowed in the notion (§3.1) that subgroups of a group of early humans may have developed their own micro-protolanguages whose blending could create a dynamically changing protolanguage that created an enlarged space of meaning accessible, more or less, to the larger group.

Grammatical constructions in evolutionarily modern languages reflect a complex interplay between perceptual salience, speaker’s subjective perspective, and the information structure of an utterance situated in a particular discourse context. Grammatical complexity, far from being an autonomous module caused by a singular evolutionary event, is a consequence of, and contributor to, the intertwined, gradual cognitive, communicative, and social evolutionary pathways of our species, supporting the coordination of responses to various cognitive and functional motivations.

5. EMPATHY, SOCIALITY, AND NARRATIVE

Whether the content of a narrative is fictional or factual, narrative is attested in every known human culture (Labov and Waletzky 1967; Maestre et al. 2023). When performative, narratives enable individual speakers and hearers to bond via attention to each other’s conversational contributions, and may cement community cooperation and serve intergenerational memory. Narratives, in their evolved EML forms, convert experience into memory, and physical distance into emotional intimacy (Nair 2003a, 2003b, 2026). After linking narrative to our account of language evolution, we explore further how humans employ narrative in increasingly complex societies and for inhabiting increasingly complex imaginative worlds. §5.3 closes by linking narrative to social construction and the social imaginary.

5.1. THE PATH FROM PROTONARRATIVE TO “MODERN” NARRATIVE

Barnard (2013) argues that the evolution of narrative, and especially of myth, pushed the evolution of linguistic complexity, with language coevolving with mythology in symbolic frameworks that extended the capacity for verbal expression. Our only demurrer is that narrative can perform valuable social functions without (or prior to) the emergence of myth. The crucial point here is the claim that narrative and language constitute another co-evolutionary *expanding spiral*.

Modern studies of pantomimed narrative (taking advantage of present-day drawings and videos) show why new grammatical devices had to be added to language as narrative evolved to recount stories far removed from the present environment, thus removing the efficacy of scaffolding provided by the immediate surroundings of the narrative and the availability of pointing. Sibierska et al. (2022) found that, even for modern humans skilled in language and pictorial representation, pantomime is an unreliable medium. In their study, even a pantomime of a single scene posed challenges for identifying which characters were involved and then who does what to whom. Contrast our hunting protonarrative where the two protagonists and the general nature of their encounter are already known to the observers, and so attribution of actions to agents is unlikely to be ambiguous. Here, the communication needed for a hunting expedition can contribute to embedding the pantomime of a skill into a narrative. We can relate this to the notion of micro-protolanguages as being limited to, and thus most effective, in limited domains.

But what then of the comprehension of modern temporally extended narratives rather than single scenes? In the study by Zlatev et al. (2024), a “Director” is presented with three simple sentences (in Polish) that together tell a simple story such as (in translation) “A man reached for an apple on a tree. The apple was bad. The man was sad.” The Director then reenacts the story relying on whole-body movements and not using words in a pre-designated space in front of a “Matcher.” The “Matcher” is then asked to match what they saw with one of four comic strips, each with three panels. The correct choice was made for 92% when the nar-

rative was in chronological order, but only 31% for achronological narratives such as “A man was sad. He reached for an apple on a tree. The apple was bad.” This seems to support the claim that “pure” pantomime may readily express some simple chronological narratives (at least if constrained to a highly focused topic) but that we need more abstract protosigns for more advanced techniques of storytelling. For example, given a sequence of pantomime-complexes S1, S2, S3 that convey three episodes of a story in chronological order, the suggestion is that a formula, call it F, was needed in the protolanguage so that its role in S3, F, S1, S2 would be something like “Because” with S1 and S2 then providing the background to S3.

Protonarratives would then increase in subtlety as protolanguage developed (i) a vocabulary of protosigns large enough to allow the recognition of the As, Xs, and Bs of *A does X to B* without strong reliance on a limited context and, crucially, (ii) new protosigns that could support non-chronological narratives of increasing complexity and support causal cognition. One such set of protosigns would provide the equivalent of pronouns to remind us that the agent or object now described has already been introduced earlier in the narrative (and note that this is earlier in the telling of the story, whether or not the story is presented in chronological order). We may have a “pantomimic fossil” here for pronouns. In a signed language, one need not use an explicit sign for a pronoun, but can instead use the position in signing space where the agent or object was first introduced to anchor later signing that invokes it. Intriguingly, Almor et al. (2007) used functional MRI to show that reading repeated names elicited more activation than pronouns in the middle and inferior temporal gyri and intraparietal sulcus and argue that the temporal lobe activation suggests that repeated names but not pronouns evoke multiple representations that then have to be integrated, whereas (a signature of our “fossil”?) the intraparietal sulcus activation may reflect the role in this integration of brain regions used for spatial attention and perceptual integration. In other words, as we comprehend a narrative, the “characters” may be “imagined” in a conceptual space that is externalized in the narratives of Deaf signers.

5.2. NARRATIVE, EMPATHY, AND INDIVIDUAL UNDERSTANDING

Going beyond §1.3, we now consider ways in which emotion may relate to the possible origins and ramifications of storytelling. Narratives are empathy machines. By enabling us to step into each other’s minds, these language structures promote the interactive growth of the emotions in ways that lay the foundations for community building—the mindreading of Ferretti’s (2025) triad (see §6.1.4). Stories simultaneously enable community-internal identifications (“we are what we are”) as well as stretch our imaginative faculties to wondrously include people, species and even imaginative constructions such as Martians, monsters, and machines (“we are what we are not”), which brings in the observation that language not only can support social coordination in the here-and-next but can also

support counterfactuals. Human counterfactual reasoning builds to a large extent on our causal cognition, being able to reason both from cause to effect and vice versa. Moreover, narrative is a language format that both requires and reinforces mental time travel and mental navigation (navigation through places both real and imaginary), bringing in the other two pillars of Ferretti’s triad, which we place within an evolutionary framework in §6. Reinforcing §§4.2 and 5.1, we stress that all these abilities require extensions of both lexicon and grammar.

Today, making complex inferences in order to live with others in familial and cultural settings is taken for granted as part of a lifelong learning curve that is, typically, narratively mediated and niche adjusted. Narratives, indeed, do more than promote empathy. Language underwrites the capacity for full-blown “displacement” or the ability to summon up scenarios that do not exist in the here-and-now/next. It gives us access to memories, fantasies, and plans for the future. Again, the scenario building depends crucially on our causal cognition (see §6.1.1, and the empathy-related discussion of mindreading in §6.1.2). Early cave art is spatial in structure but, in some cases at least, its representational drive is narrative. The ability to ask questions and deny as well as assert statements certainly fueled much of human biocultural niche expansion, at least since the emergence of expanded protolanguages. This motivates a characterization of humans as both *Homo narrans*, the humans that tell stories and *Homo quaerens*, the humans that question (and come up with answers, whether right or wrong).¹⁵ Combining these two cultural developments, narrative introduces (or, at least, greatly expands) the ability to ask “why?” and answer “because.”

We see here another expanding spiral in the coevolution of language and causal cognition, building on the Great Move. All animals have some capacity for reinforcement learning, to change details of the inner workings of their neural networks, driven by positive and negative reinforcement (reward and punishment) to increase the likelihood of achieving various goals when faced with different perceptual conditions (Sutton and Barto 2018). This is based on rules at the subneural level of synaptic plasticity, not on any appeal to symbolic representations as such. After the Great Move, it becomes possible to support causal cognition—understanding and putting into words the notion that “A causes C” and to organize our thinking in terms of events, which form the backbone of a narrative. There is no claim that such statements express inferences in scientific theories, only that (as when a teacher seeks to motivate a step in behavior that seems pointless to the student) prior experience indicates that this practice seems to succeed.

Stories may be the main vehicle in all cultures that train us emotionally while at the same time coaching us in rationality and hypothesis formation. When people in contemporary cultures describe themselves in relation to the emotions they experience, they seem to use varied descriptive categories such as traits, intuitions, desires, intentions, moods, sensations, qualia, and so on (for a full chart, see Nair 2003b, 2011b). Since the combinations thus engendered

are endless, we can have very large or small variations of perceived or self-described personality from individual to individual. More generally, Nair (2009, 2011a, 2014) suggests that narrative intentionality is oftentimes *performatively* driven in tandem with the motivation to *emotionally* engage an audience. Moreover, we can have extremely rich and teeming interior worlds or relatively limited and impoverished ones according to the specific ways in which we make inter-connections between all these cognitive traits in our narratives. Intentional behavior existed long before there were humans or even primates and certainly before there was conscious expression of intentions in (proto)language. As noted in §3.1, intentions are already present in demonstration. Hence, elevating intention into language (Nair 2003a) requires words to mediate the aboutness of future actions while pairing intentions with actions/things in the world (going to forage for food, marrying people, etc.). In current use of language, our intention to achieve some goal may further implicate the understanding of temporal order and event sequencing (as indicated in §5.1).

One fundamental way in which our internal representations appear to differ from those of animals is that, as a species, we constantly make connections between fictional worlds and the mental universe of the emotions. Animals do not, as far as we know, create fictions. The reason may be that they do not organize their cognitive worlds in events. And if they do not, it follows, animals do not use fictions or art as an important learning or cultural device, whereas humans consistently do. There is no society in the world today that does not tell stories as a basic communicative tool and no culture where children are not introduced to the world through stories, drawing, and music (Nair 2003a, 2011a, 2014, 2015).

As a species, we habitually attribute intentionality to utterances in context. One has to intend to *do* something like read *Das Kapital*, or get oneself vaccinated, or change the world. Interacting with “the other” requires us to assess moment to moment, however imperfectly, their intentions (more on this in the discussion of “mindreading” in §6.1.2). Even a child must judge when scolded whether a parent is “really” angry or upset or not and behave accordingly. However, just as we see protonarrative preceding narratives, so do we see protoconversations preceding conversation, long before the emergence of the *syntactically complex verbal codes* that we charted in §3.3 and §4. Further, the roots for persuasive reciprocity may be found in the sorts of protoconversations needed for people to coordinate their behaviors in a limited set of skills—a propensity that we offered (§3.1) as a source of diverse micro-protolanguages that might arise within a larger community.

Material symbols can range from amulets and candles to dragons, swords, and sharks, and as these symbols travel across cultural geographies in narrative formats, they are often “flexibly” converted into “niche” local contexts—candles can “become” Indian “diyas,” swords shape-change into scimitars, Buddha and Jesus are absorbed into the Hindu pantheon, etc. These symbols very often comprise sources of emotional arousal (fear, wonder, excitement, an-

ger). Narratives, too, are symbolic artifacts constituting, in both form and function, catchment areas for symbol storage and transfer. Indeed, symbols gain meaning only through the narratives or social practices in which they are embedded. Just as all language-users rely on grammar to build sentences that “construct” the world for them, so do we as modern symbol-using creatures rely on narratives to teach us to relate causes to effects, tutoring us in sense making. “The king died and then the queen died of grief,” (Forster 1927) illustrates such strong causality.

Such stories provide explanatory accounts (Nair 2026). Rudyard Kipling’s knowingly false Just-So-Story “How the leopard got its spots” published in 1902 has been succeeded by modern evolutionary explanations (e.g., Allen et al. 2011). POLE (§§2 and 3) rests on just-so stories, too, in its search for steps towards explanation of language origins. This justifies us in thinking of narrative (well, some narratives) as prototheory. Human beings would probably find it impossible to manage their lives without storing and reliving their salient experiences by way of such proto-theoretical accounts (Nair 2012). A story for children may have lessons within it to help the child develop as a member of the social group. For adults, some stories are told to simply pass the time, whereas others seek to explore cosmic questions (“where do we come from?” “what might we become?”). The latter requires a sophistication of language far beyond protolanguages to present particular hypotheses about phenomena in the world, especially the social world, and a paradigm to explain them. We will return to the role of narrative in social construction in §5.3. Bouizegarene et al. (2024) propose that narratives are “tools for mutual prediction in social contexts.” While this is not true of all narratives, it does show that narratives need not just be recollections of the past but active inferential frameworks that minimize uncertainty and guide future behavior. Thus, one way narrative is still evolving is to enhance “adaptive” intentionality, helping “agents” anticipate others’ intentions and coordinate plans.

Innocenti and Rozzi (2025), in their evolutionary analysis of narrative selection, argue that agents must choose between competing narratives to interpret the world (see Nair 2014; 2015, anticipating this very point). “Conformist narratives”, they suggest, may possess an evolutionary advantage in this regard. In short, narratives of this kind themselves undergo selection pressures thus arguably shaping, in time, the collective “intentionality” of societies. Baba et al. (2025) further maintain that intentionality is not fixed but “emergent” and actors’ goals evolve as they share meanings with others in their community. Intentionality is, so to speak, narratively scaffolded; human agents do not just have intentions, they culturally inherit and revise them through stories. Q. Li et al. (2026) consider experimental evidence in the same arena, suggesting that narrative framing significantly increases intrinsic motivation and engagement. According to them, narrative provides a “motivational architecture” for human communication, organizing future action by embedding goals in meaningful sequences. Yao et al. (2026) also do a meta-analysis of narra-

tive “efficacy,” underwriting the argument that narratives do not just describe motivation; rather, they produce and reshape “motivations” in a way that significantly improves self-efficacy.

This explanatory function of stories is evident in the Indian epics the *Ramayana* and the *Mahabharata*. These are narrated orally to pre-school and young schoolgoing children as well as in nuclear family settings, often in intergenerational contexts. They also take form as material objects in the form of wood and cloth puppets and toys for children and soft toys for infants (recall §3.3 on linking language-readiness to material praxis); in school texts as poems or stories, and as multilanguage “story box” discourses (Nair 2003b, 2011b, 2015). Both these texts, strikingly, make it a point to self-referentially narrate their origins; they insist on telling their audience *why* they came to be composed. Two types of complexity are involved—both the richness of the lexicon and grammar, and the diverse strategies of putting sentences together over an extended narrative in a way that not only moves forward the interaction of diverse characters but also places demands on the ability of listeners to keep track of what these characters have done and to anticipate (not always correctly) what their fates may be. We see here that cultural evolution has supported the expanded strategies of storytelling as well as the expanded capacity not only to understand more complex stories but also to become dissatisfied with overly simple stories (unless they are the lyrics of a song).

Indeed, making even the most elementary sense of the complex premises of these texts demands of an audience a great deal of reasoning acumen. It compels them (in India, as quite young children) to recognize that linguistic skill can evoke an emotional response to the sorrow of “the other” even if this other belongs to another species. Empathy and narrative creativity are inextricably interconnected (see the discussion of mindreading in §6.1.4). This invokes, and can build upon, a notion of multimodal narrative related to pantomime and protosign as well as early art. Clearly, such complexity comes long “after” the base story of human origins set forth in §§2 and 3 but does set forth ways a culture can build on them.

Narratives provide us with relatively low-cost means of taking mental risks. Consider, for example, our physiological reactions when we read or listen to a ghost story or adventure fiction. Our palms sweat, our pulse-rates shoot up, and our throats go dry, even as we “know” that the words on the page are pure make-believe. We don’t actually have to go and climb that cliff or love that boulder in order to realize the dangers and aches inherent in these situations (Nair 2003b, 2011a, 2026). As suggested earlier, narrative can serve as proto-theory, presenting particular hypotheses about phenomena in the world and a paradigm to explain them. This requires at least an early language that includes both questions and the ability to question whether something is not the case, and to ask what might be true instead. Yet, at the same time, most stories and fairy tales we tell children are false—they may serve to help children learn words and concepts (or they may help instill moral

principles and include scary situations and how to avoid them) in a non-factual way (e.g., the tales of Little Red Riding Hood or The Three Pigs).

As will be seen in §5.3, narratives—coupled with changes in the physical environment and technosphere—can cohere to create the social schemas that construct different social realities. But we also note that narratives come in many genres. There may well have been an evolutionary phase that grounded a few fundamental discourse genres still around today. Drama, for example, is supported by the human capacity for conversation with its continually alternating listening-and-talking. As suggested in §3.4, metaphor may have first originated when the need arose to find a word expressing some feature of a novel context and “recycling” a word from another context that fitted some but not all of those features—a far cry from the conscious deployment of metaphor today. Narrative would have flourished alongside these base genres. Poetry, finally, would have perhaps co-evolved with song as early humans became aware of meter as shared between language and music, augmented by the repetitive structures of clapping, foot tapping, and other rhythmic forms of body language, while the conscious deployment of rhyme and alliteration may have been relatively late arrivals—not just finding tools to share meanings in narrative or conversation, but to provide increasingly elaborate decoration. Indeed, these key genres of language use evolve to serve purposes beyond the simply functional or transactional, to sustain a certain quality of mental life among human communities. Moreover, communal deployment of music and dancing would have contributed to communal bonding (Cross et al. 2013; Lewis 2013), complementing the importance of narratives shared around the campfire. That is why narrative and its various genealogically ancient cousins still appear to be well and thriving today.

5.3. SOCIAL CONSTRUCTION OF REALITY

The processes charted in §4 provide a language with the power to express cognitive distinctions in a clear but economical way. The Speaker’s perspective on the referential situation also plays a role. Single lexical items often foreground a specific location within a larger semantic frame, or local universe of discourse: “Consider the meaning of the word *uncle*, which presupposes a background network of kinship relations, and foregrounds one particular node in it” (Verhagen 2010: 50). While in English we use general constructions to talk about kinship, there are languages in which kinship relations are themselves grammaticalized (Evans 2003; Evans and Levinson 2009). The institution of kinship in many extant societies embraces not only human kin, but other entities (some that “we” regard as animate, such as animals and others that “we” regard as inanimate—such as mountains and lakes, and some that “we” regard as supernatural). To inhabit a world populated by kin, structured by kin relationships, is to live in a storied world that is governed by shared imagination as much as by “veridical” perception.

In this light, consider the importance of other “social facts,” such as money, religious beliefs, or socio-political roles, that seem to depend upon a kind of social contract. An emphasis on collective agreement, however, has sometimes led to a tendency to treat social facts as purely fictive, involving beliefs about unreal entities (be they gods, dollars, or alien intelligences). However, Durkheim (2013: 21), who introduced the term in 1895, rightly emphasized that social facts are “invested with a coercive power.” They are the key to the *construction of shared and binding social realities* (Arbib and Hesse 1986; Searle 1995). The very language we speak is a *social schema* that organizes diverse other social schemas that in turn organize how we behave in a community, including morality, religion, ideology, and economic roles. A somewhat different take on collectively shared symbolic realities is offered by the notion of the *social imaginary* (Castoriadis 1987). We see it as dependent not only on symbolic capacities in general, but specifically upon the power that evolutionarily modern languages have endowed humans with, to engage in flexible construal and imaginative narrative thinking (Bruner 1991). The hunter’s tale (§2.4) showed how even pantomime can allow the protonarrator to characterize roles, but the flexibility of construal is limited without the resources of ELs/EMLs. A future challenge is to revisit this within our evolutionary framework, exploring how, as protolanguages and languages emerge, so does our power for narrative expand far beyond that expressible by early pantomime, supporting novel forms of grammaticalization, flexible construal, and imaginative narrative thinking in ways that yield their flowering in all known evolutionarily modern languages. Our account implies, in fact, that narrative and conversation were important drivers and beneficiaries of these processes.

In his “The Narrative Construction of Reality,” Bruner (1991) considers 10 properties of narrative. We consider two. First is *canonicity and breach*.

“Not every sequence of events recounted constitutes a narrative. Some happenings do not warrant telling about and accounts of them are said to be “pointless” rather than story like. A Schank-Abelson script [(Schank & Abelson, 1977)] is one such case: it is a prescription for canonical behavior in a culturally defined situation, how to behave in a restaurant, say. Narratives require such scripts as necessary background, but ... to be worth telling, a tale must be about how an implicit canonical script has been *breached, violated, or deviated from* ...” (Bruner, 1991:11) [Our italics]

Here we are concerned with the way in which a society may become shaped by a more or less coherent set of stories. In this regard, Bruner speaks of *narrative accrual*, the “cobbling together” of stories to bind them into some sort of whole.

“Whereas sciences can go “cumulative” employing procedures for verifying the predictions of general principles [(Hesse, 1980)], “verificationist criteria have limited applicability where human intentional states are con-

cerned ... Yet narratives do accrue, and, as anthropologists insist, the accruals eventually create something variously called a “culture” or a “history” or, more loosely, a “tradition.” ... Institutions, too, as we know from the innovative work of Eric Hobsbawm [see (Hobsbawm & Ranger, 1983)], “invent” traditions out of previously ordinary happenings and then endow them with privileged status.” (Bruner 1991: 18)

As §5.2 demonstrates, children arrive at a foundational knowledge of their cultures via stories and myths. In that they tie in causal, logical explanations of actual concrete “events” with emotional affect in a powerful way, the “lessons” taught to us as children, and then throughout our lives through stories, lay the corner-stones for most of our crucial beliefs (Nair 2017, 2018). Cognitively, that is, narrative seems to have evolved culturally to probe experience and create competing theories about the world, including the world of emotional experience, which is so crucial to cultural survival. In this sense, the expansion of narrative as a mental structure provided tremendous selective advantages to our species.

However, it would be a mistake to see language alone as supporting the kind of social complexity that characterizes all known “modern” human cultures, as §3.2 and the examples from Hindu culture in §5.2 have emphasized. Consider, too, a religion like Christianity that is grounded in the narratives (multiple, not always consistent) of the Bible, but central to which are patterns of ritual behavior in dedicated buildings involving objects deemed sacred by the religion, as well as paintings and music. In other words, such a social schema extends well beyond language into the reconstructed physical environment and the praxis it makes possible, creating a complex and distinctive niche in the social imaginary.

In the spirit of HOLO 1 and 2 we stress the interaction of semiosphere and technosphere, and in the spirit of EvoDevoSocio we stress that individuals growing up in similar environments may differ as to exactly what *collective agreement*, accepting a reality structured by social schemas, they have interiorized in guiding their own behavior. Indeed, at times individuals rebel (Palocz-Horvath 1964). As a result they may become outcasts, or may help define a new social schema, as when the conflict between papal authority and the lusts of a king catalyzed the rise of the Church of England. To this, we must add that the natural world provides implacable constraints on narrative constructions. The narrative that “climate change is a hoax” may have captured many, but in the end the reality of increased floods and fires, and of mass migration induced by the loss of arable land, will for many people lead to a rejection of this “collective fiction.”

6. THE COGNITIVE ARCHITECTURE OF LANGUAGE ORIGINS

We first summarize Ferretti’s triadic cognitive architecture of mental time travel, mind reading, and navigation, while adding causal cognition to the triad and notes on the missing evolutionary dimension. This lays the basis for an at-

tempt to offer a more comprehensive *evolutionary* cognitive architecture within our EvoDevoSocio framework, integrating the biocultural evolution of POLE with the Post-POLE progression from protolanguage to evolutionarily modern languages. It is crucial to our evolutionary framework that pantomime provides a bridge between practical skills and communicative skills.

The recent success of Large Language Models and other AI systems has demonstrated the immense power of large networks of adaptable artificial neurons to process language (language prompt in → language response out), but these are a far cry from models of how the human brain evolved to become language-ready, or how cultural evolution developed social groups in which individuals can gain mastery of the complex linkages of semiosphere and technosphere within the physical environment. Our task is now to sketch the evolution of cognitive architectures adequate to this task. As background for this effort, we add four observations on brain evolution.

1. The brain is not organized in terms of separable circuits devoted to separate (e.g., cognitive) functions—rather such functions are mediated by the competition and cooperation of activity in diverse regions. Yes, brains have regions whose lesion can destroy specific abilities but this does not imply that they alone support the ability, and we can find regions with “primary functions” such as the role of primary visual cortex—but even this region can respond to auditory signals, and visual experience depends on its interaction with diverse other regions.
2. Brain regions that evolved to help support ancestral functions can be rewired through experience (possibly socially mediated) to support new functions. This underlies our notion that it took primarily cultural evolution from a *language-ready* brain to a brain that would become *language-using* in children who developed in a language-using community *and that the fine structure, at least, of the brain is transformed in the process.*
3. The English neurologist Hughlings Jackson (1878–79) viewed the brain in terms of levels of increasing evolutionary complexity. Observing his neurological patients, he argued that damage to a “higher” level of the brain caused the patient to use “older” brain regions in a way disinhibited from controls evolved later to reveal behaviors that were more primitive in evolutionary terms. As the brain evolves, it exhibits new patterns of neural activity which can provide new “neuro-ecological niches” for reorganizing neural circuitry to exploit those patterns. Once such new circuitry evolves, there is a new “information environment” for the earlier circuitry—so it may evolve in turn to exploit these new patterns (Arbib 1991).
4. Genes can change the growth potential for cell morphology and connectivity, but these changes may affect whole regions of the brain rather than a single region. For example, Barton (2012) argues that genetic specification of cerebellum and neocortex are closely coupled, and so we may link expansion in cerebral

functions to the co-evolved expansion of cerebellum to provide the resources for the effective tuning and coordination of these capabilities.

The Great Move as charted in §1.4 posited that a major step in the evolution of the language-ready brain was the emergence of brain regions capable of representing symbolic actions, but with distributed regions interacting with each other and with regions more directly engaged with sensory, motor, and emotion representations and related brain mechanisms shared in great part with nonhuman species, and not limited to primates. We suggest that this ability to build representations exploiting other representations, and the ability of vast regions of “neural real estate” open to developmental reshaping, made possible ratcheting (Tennie et al. 2009) as well as graded embodiment and towers of abstraction (Arbib and Cuccio 2026; Arbib et al. 2014).

These general principles inform our overall perspective, but details of brain evolution are outside our remit.¹⁶

6.1. FERRETTI'S COGNITIVE ARCHITECTURE

In responding to Ferretti's argument that “stories came first,” we stressed that stories *as we know them today* did not come before language. Our purpose in this section is to place the three elements of Ferretti's cognitive architecture – mental time travel, mental space travel, and mindreading—in an evolutionary framework. POLE holds that the language-ready brain could from the start support simple protolanguages and protonarratives but that tens of millennia were required before the emergence of early languages via culturally-led biocultural evolution, with evolutionarily modern languages being “recent” on our evolutionary timescale (§7). In the same spirit, we will argue that the language-ready brain of early humans could support only a primitive form of Ferretti's triad, and that, following our EvoDevoSocio framework, it required enculturation in increasingly language-rich and narrative-rich societies to reshape human brains to support the triad as humans know it today. But first we add causal cognition to the triad.

6.1.1. Causal Cognition

Ferretti and Adornetti (2021) ask,

“How ... can one explain, without reference to cognitive systems, the ability to judge the coherence of what the interlocutor is saying in relation to what he or she has said before, or the ability to grasp the interlocutor's intentions in order to understand where his or her speech is going? Evaluating the coherence of a speech requires the ability to move forward and backward in time, ... [while] grasping the interlocutor's intentions calls into play mindreading devices ...”

This concern for coherence leads us to add *causal cognition* to the triad. Non-human animals understand the direct effects of their own bodily actions, but when it comes to reasoning about causes that are detached from their bodies, they show little understanding of the intermediate and hidden “forces,” whether physical forces or intentional

mental states, that are central in human thinking (Povinelli and Penn 2011; Tomasello 2010). Gärdenfors and Lombard (2020, 2024) argue that there has been coevolution between the manufacturing and use of tools and advanced forms of causal cognition in the *Homo* lineage. With the aid of tools such as throwing spears, bow and arrow, and poisoned arrows, humans can extend the effects of their actions in space and time.

Human causal cognition forms the basis for our understanding of events (and thus enriches the three components of Ferretti's triad). Causal cognition of events allows us to speculate about potential outcomes of actions, test and re-adjust our imaginative hypotheses, and shift attention from one target to another. It thereby allows new types of generalization in comparing events (Gärdenfors et al. 2018). An anthropological example (Liebenberg 1990) is that when hunting with poisoned arrows, Kalahari San people engage in "speculative tracking," using working hypotheses gained from the signs left by an animal, socially and experientially gained knowledge about the behaviors of the animal, and of the landscape in which the tracking is taking place. Based on these imaginative constructions, the hunter creates predictions in ever-changing circumstances involving a continuous cognitive process. The upshot is that such event cognition allows for more complex causal thinking to evolve.

6.1.2. Navigation in Space

Any animal that needs to forage for its food, find places to sleep, or know where it may be safe from predators must be capable of navigation in space. This involves, at least in simple form, having brain structures (and not just a hippocampus: Arbib 2021: Section 6.3) that support a cognitive map, knowing an increasing number of significant places in its environment and the means to get from one to another in ways that depend on the animal's current motivation. Navigation may involve both the memory of habitual routes and the ability to get to a suitable place via a detour when the habitual route is disrupted (for example, by the presence of a predator). In addition, we know that such a cognitive map is malleable. But without at least protolanguage to name places or provide criteria to recognize them, and to indicate how one gets to the next, one cannot communicate about a new and complex route that may go beyond what can be seen as the route is explained—and even less can one build narratives describing new landscapes, whether real or imagined, and the adventures that might be had as one traveled there. Contrast the waggle dance of bees (von Frisch 1954, 1967) that enables bees to share the location of a new pollen source they have discovered. The waggle must have a short-term memory of the direction and distance of the food source from the hive and encode that in the dance after returning to the hive. The bee's "language" is a genetically-encoded system restricted to communication encoding the location of pollen. Research in animal communication continues to surprise us with other communication systems, some completely alien to the hu-

man use of manual and vocal gestures. However, they all lack the open-endedness offered by pantomime.

6.1.3. Mental Time Travel

In discussing POLE-4a and POLE-4b, we saw basic *pantomime* corresponding to one action-object pair as providing a unit within what may be a larger performance of pantomime—as might occur in teaching the various stages of a skill, in developing a protonarrative, or in social coordination of multifaceted behaviors. The example of a hunter returning from a hunt showed that even a protonarrative can be based on the recall of past events. Similarly, we envisioned pantomimes that invite group members to join the mimer in hunting for prey. Both examples involve some element of "mental time travel" (Cheng et al. 2016; Suddendorf and Corballis 2007; Suddendorf et al. 2009)—the ability to recall past episodes and imagine novel ones, at least in making plans for future action. However, the "mental time travel" of our hunter is vastly simpler than the mental time travel of today's language-using humans. We thus add to the POLE sequence the proviso that a language-ready brain must already have had the ability to recall past episodes and imagine novel ones. Moreover, the examples in §5 show that human narrative extends far beyond the physically experienced past and the here-and-next into the realms of imagined realities.

Mental time travel and navigation in space are crucially interlinked. Certain animals can cache food at specific places during one season and recover them at another season, and this cognitive map is updated each time the animal empties one of its caches or visits the cache and discovers that another animal has removed the food (see Olton et al. 1980 for a related lab study of rats in mazes, while C.C. Smith and Reichman 1984 review studies of the evolution of food caching by birds and mammals). However, remembering where one cached food some months ago is very different from remembering when it was cached. Certainly, animal behavior is regulated internally by their circadian rhythm and externally by the ability—in many cases thanks to genetic underpinnings—to adapt to the changing seasons. But this "pre-Great Move" capability is very different from the sort of time humans have now become enculturated to, in which the time of year is specified by the date on a calendar, and the time of day is specified down to the minute or even more precisely by a clock.

The calendar and the clock are human inventions post-dating the rise of EMLs, but we hypothesize that the human ability to tell *autobiographical narratives* is both unique to the *Homo* lineage, and universal in *sapiens* populations, with the development of autobiographical memory and narrative forming yet another expanding spiral. Recall from §5.1 the study by Zlatev et al. (2024) of the limitation of even modern pantomime in expressing achronological narratives, even though breaking chronological order is a key property of narrative. Causal cognition is one of the abilities underlying the ability to over-ride chronological order. Similarly, autobiographical memory may be enriched

when a new episode evokes associatively some earlier episode that then triggers contemplation of a relationship between those episodes that one had not registered before. All this may occur when one tells a story about one's past. Perhaps more tellingly, mental time travel in modern humans supports not only narratives grounded in memory of past episodes but also can escape the past and the known environment to support those feats of imagination that can make narratives so appealing (Addis and Schacter 2012; Addis et al. 2007; Schacter and Addis 2020).

6.1.4. Mindreading

How does someone tell a story or carry a conversation that the hearer will find interesting? The secret here depends on the individual development of the storyteller informed more by experience of hearing others tell stories than by the general structure and evolution of their cognitive architecture, but also on the capacity for empathy stressed in §5.2. Here we relate empathy to mindreading.

Most animals have evolved responses (sometimes shaped by learning from experience) to tell predator from prey and to use observations of the action of the prey or predator to interpret behavior by attributing *behavioral dispositions* to others. Turning to conspecifics, recall Charles Darwin's (1872) comparison of human emotional expression with that in animals §1.3). It places the *external role* of emotion within a comparative, and thus implicitly evolutionary, framework. Animals could adjust their behavior in relation to a conspecific based on the other's facial (or other) "emotional" expression as a cue to behavioral disposition, without any attributions regarding the inner state of the conspecific. But responding to such cues is a long way from a "folk psychology" enabling the interpretation of the actions of other people in terms of posited (but not directly observable) mental states.

This capacity for "mindreading," that is inferring the intentional states of others from their behaviors and utterances, is often designated Theory of Mind. However, our use of the terminology of "mindreading" does not signal adherence to its explanation in terms of "theory theory" (Leslie 1987) or "simulation theory" (Gallese and Goldman 1998). We agree, rather, with Gallagher and Hutto (2008) that we come to understand the minds of others through intersubjective perceptual processes, pragmatically contextualized comprehension, and, ultimately, linguistically organized narrative competence. Indeed, when Gallagher and Hutto say "In seeing the actions and expressive movements of the other person one already sees their meaning; no inference to a hidden set of mental states (beliefs, desires, etc.) is necessary," they later concede that such primary intersubjectivity (Trevorthen 1979) underpins developmentally later practices that may involve explaining or predicting mental states in others—citing Stern's (1985) "layered model" in which developmentally primary understandings come to operate in parallel with more advanced ones. Consider that we may directly "see the meaning" of a smile in some cases and respond nonconsciously, but when someone we have learned to mistrust smiles at

us, we may well seek consciously to evaluate the "mental state" behind that smile—are they smiling "with me" or "at me"? This, of course, is the stuff of narrative. Indeed, the "Narrative Practice Hypothesis" claims that "children normally achieve folk psychological understanding by engaging in story-telling practices, with the support of others. The stories about those who act for reasons—i.e., folk psychological narratives—are the foci of this practice. Stories of this special kind provide the crucial training set needed for understanding reasons" (Hutto 2007). Again, we see the relation of causal cognition to narrative, and the suggestion that the search for reasons (*Homo quaerens* and *narrans*) co-evolves with the expanding power of languages. Gärdenfors (2024) sees mindreading as an important extension of human causal cognition, raising the issue of just what mechanisms related to causal cognition must be included in the definition of the language-ready brain.

The stories we hear (and the accompanying pictures) can expose us to a range of human and nonhuman behaviors and an otherwise scarcely accessible understanding of "what the other person is thinking and how that will affect their behavior." We have come a long way to develop languages and cultures rich enough to support a folk psychology in which mental states can play a central role. We can see evolutionary precursors of this in the mother caring for an infant in many mammalian and avian species, but linguistically supported human intersubjective understanding comes at the end of a long road of biocultural coevolution of social interaction and communication, supporting a rich understanding of the mental states of others, expressing that in language, and building imaginary scenarios of the mental states of protagonists in complex narratives. All this supports the view that mindreading, too, is a co-evolute of language for narrative. Mindreading applies even more strongly to conversation, in which speakers combine their past knowledge of an interlocutor, and of previous communicative exchanges and the state of a current conversation to build a complex discourse model of what other speakers know, how they are feeling, and what they care about.

6.2. EVOLUTION OF THE HUMAN COGNITIVE ARCHITECTURE

Having placed Ferretti's cognitive architecture in an evolutionary frame, we now provide a perspective on how cognitive architecture evolved to support narrative and conversation as well as teaching and social coordination, starting with pantomime. In doing so, we make explicit the cognitive architecture implicit in earlier sections. The aim here is to assess what was in place before and after the Great Move, incorporating the lessons from POLE, while stressing that what is unique about human cognitive architecture is that it depends not only on brain and body of the individual but also, in terms of the EvoDevoSocio framework, is crucially reshaped by the biocultural niche (including technosphere and semiosphere) in which the individual develops.

6.2.1. Before the Great Move

Moving beyond the frogs of §1.4 but starting before LCA-m and LCA-ga, we note the rich evolutionary basis shared by diverse mammals, rooting the biocultural evolution of language in the diverse sensory, motor, motivational, and other systems that serve (not only) mammalian social interaction and survival.

- **CA1) Generic Mammalian Cognitive Architecture.** All mammals have sophisticated sensory, motor and motivational systems for interacting with the environment in ways that involve both individual action and social interaction to meet needs for nutrition, threat avoidance, and reproduction. The ability to recognize and interact with prey, predators, conspecifics, and potential mates are all provided, but different species will have different investments in caregiving. They can navigate in space, and have diverse abilities for learning, including some procedural learning and episodic memory.

In other words, many creatures have the “pre-Ferretti triad.” They have (limited) episodic memory, spatial navigation, and some ability to recognize the behavioral dispositions of others. Mammals have enlarged cerebral cortices compared to other species, and primates are notable for their enlarged forebrains. This trend continues in the *Homo* lineage (though Neanderthals had somewhat larger brains on average than *sapiens*) but in relation to the Great Move (§1.4), this is not an add-on of a general-purpose computer but the emergence of a network of “semi-general” brain regions with differential connectivity and neuron distributions, with the power of the brain increasing both with number of neurons and the integrated diversity of these regions and the socio-ecological niche in which each animal develops.

- **CA2) Distinctive Primate Cognitive Architecture centered on manual dexterity, as of LCA-ga (POLE-1 and 2).** LCA-ga primates (on the evolutionary path to great apes, *Australopithecus*, and the *Homo* lineage) had, crucially, manual dexterity, an ability to recognize the actions of conspecifics, and the ability to learn new manual skills through emulation that rested on imitation by behavior parsing—recognizing subgoals in the observed skills but with little or no attention to the trajectories of movements observed to bridge these subgoals. Communication involved a limited set of species-specific calls plus a small innate repertoire of manual gestures augmented by a few gestures developed through ontogenetic ritualization by dyads. Different species exhibited a range of group structures and dominance patterns.

Such manual dexterity is well established and engages some simple tool use, but the motor control of vocalizations is limited. The path from LCA-ga to *Homo* involved diverse changes, including the development of opposable thumbs, bipedalism, and reshaping of the skull but we emphasize manual dexterity while briefly noting the significance of changes in vocal musculature and neural control that made

human speech possible.

Starting from this basis, our account integrates the technosphere (material praxis and creature-induced changes to the physical environment) and the semiosphere (including creature-induced changes to the social environment) in both the evolution of the language-ready brain and subsequent evolution of languages, emphasizing the ability of children to acquire the language of their community as their interactions within the technosphere and semiosphere both expand.

By taking manual dexterity as a crucial evolutionary component we thus go far beyond theories that focus purely on the emergence of grammar or narrative, or those focused solely on spoken languages.

With this we turn to post-LCA-ga features of Cognitive Architecture that emerged on the path to various species of *Homo*.

- **CA3) Complex action recognition and complex imitation (POLE-3).** This is relevant to the faster transmission and improved accuracy of practical skills as recognition of subgoals is augmented by observing and incorporating *various movement details* of the actions involved.

This is not only beneficial relatively early in the path from LCA-ga, witness the mastery of Oldowan stone-tool-making, but also creates the capability of modern children for acquiring the meanings (communicative goals) and the phonology (movement details) of their (spoken or signed) language. It also provides a substrate for the development of teaching by demonstration and pantomime.

6.2.2. The Great Move

- **CA4) Pantomime for teaching, social coordination, protoconversation, and protonarratives (POLE-4a, 4b, and 5).** Hominins complemented complex action recognition and imitation with the development of *teaching relationships* wherein teachers came to control the acquisition of certain skills via demonstration and pantomime. Simultaneously, pantomimes were increasingly used to support social coordination and could serve simple forms of both protoconversation and protonarratives. Crucially, pantomime offers an open-ended system for communication, understood as such by both mimer and observer.

Linking back to HOLO 1 (d’Errico et al. 2027 [this issue]), we stress that this affects the technosphere as well as the semiosphere. Increased means for teaching and social coordination more generally meant that more complex skills and the consequent changes in material objects and practices could emerge that could in turn support the development of novel pantomimes and means to employ them.

The key transition from demonstration (where the teacher simply executes a skill, albeit with some pauses and exaggerated movement to enable another to possibly-acquire the skill through complex imitation) to pantomime (and other forms of social coordination) depends on an open-ended mechanism whereby a widening circle of ac-

tions (and related objects) gain representations outside the basic sensorimotor circuitry that supports communication. Such pantomime involved a new social understanding of both the producers and interpreters of the pantomimes. As a result, a person could perform the bodily movement of an action in the absence of objects with the expectation that someone else could interpret this as communicating the need for an object or some associated behavior (an interpretation heavily dependent on context), rather than a failed attempt at some practical action. We also noted that one member of a dyad might come to understand a ritualized version of such a pantomime, but we distinguished this from conventionalization.

- **CA5) The Great Move continued—from pantomimed actions to conventionalized protowords and protolanguages (POLE-7).** The pantomimes of CA4 were still within the range of bodily actions (even if somewhat ritualized), while in the vocal domain some ritualized imitations of sounds might come to indicate certain courses of action via sound symbolism. However, in each case, the systems involved were expressive of how early humans might have reacted to seeing or hearing some changes in the physical and social world around them, and determining how to react on possibly incomplete cues. But with CA5, a great change occurs. As conventionalization builds on ritualization, members of a large group recognize that they are not dealing with a fixed set of communicative gestures but rather that they have access to an open-ended set of protowords that can be created and shared by the community. These can be introduced to better describe the world around them and refine their behavior accordingly (even though as yet they lacked the ability to put such realization into words). We use the term *protolanguage* for this open-ended set of protowords as shared by a community, noting that the same mechanisms that might at first yield micro-protolanguages for subgroups with particular concerns (foraging, hunting, child rearing, tool making, and so on) could later reorganize these into a shared (and dynamic) protolanguage shared by the community.

In particular, conventionalization supports the increasing use of “protowords” that can bear little trace of their “etymology” in embodied action and perception. In particular, different protowords may emerge for conveying an overall behavior or situation as well as key components. At this stage, we can begin to see ratcheting take hold, as conventionalized communication can support rapid instruction about how to assemble known skills in facing new challenges, once someone has figured out an assemblage or script that works.

The Great Move (§3.1) was supported by the expansion of neocortex in primates, but its scope expanded with CA4 and CA5. At this stage, the biocultural evolution that supports modern humans came to play less of a role since it had now created far more “neural real estate” than needed at stage CA5—but exploiting this neural real estate in social systems of increasing complexity laid the basis for acceler-

ating patterns of cultural evolution.

Indeed, changes from the brain of *Homo erectus* (cranial capacity: ~600–1,100cm³) to *Homo sapiens* (cranial capacity: 1,300–1,400cm³) (Leigh 1992; Smithsonian-Institution n.d.) may have converted a “ready-for-complex-protolanguages” brain to a “ready-for-evolutionarily-modern-languages” brain. This evolution may have supported the cognitive changes of §6.1, with the *H. erectus* brain being language-ready but limited in its capacity to extend the semiosphere and technosphere beyond the reach of complex protolanguages (but note too the changes in brain size from early to late *H. erectus*). This would have yielded a *sapiens* brain with an expanded capacity and the capability to provide the cognitive foundations necessary for cultural evolution to begin to *predominate* over biological (gene-based) evolution, allowing successive generations to develop languages of increasing complexity, eventually leading to evolutionarily modern languages.¹⁷ This process, however, presumably also depended on social and motivational factors—such as increased tolerance, shared attention, and cooperative teaching—which scaffolded the effective transmission and stabilization of communicative systems. We say “predominate over” rather than “obviate” because Baldwinian evolution (Baldwin 1896; Waddington 1953) may have played a role in favoring genetic changes that supported performance of certain cultural innovations (as may have occurred in supporting neural mechanisms of Speech-1 that increased efficacy in Speech-2).

6.2.3. Exploiting the Great Move

Our claim is that once humans had brains and bodies situated in a technosphere and semiosphere at the level of CA5 (POLE-7), their brains were at least complex-protolanguage-ready but that increase in brain capacity rather than “language genes” yielded brains that supported the cultural evolution (perhaps with an assist from Baldwinian evolution) that led from protolanguages to modern language, together with the partnership between caregivers and children enabling children to become part of the (proto)language community in which they were raised.

- **Post-POLE:** Cultural evolution came to predominate over biological evolution as language emerged from protolanguage, with a widening of the lexicon accompanied by a *compositional semantics* supported by construction grammars for combining word meanings to express and comprehend an expanding range of novel meanings. These expansions in the semiosphere co-evolved with expansions in the technosphere as well as supporting the increased sophistication of social schemas that supported and constrained the emergence of complex societies in which the here-and-next of behavior became increasingly intertwined with the worlds of imagination. We posit that by this stage, the EvoDevoSocio-supported cognitive architecture of humans included causal cognition, mental time travel, mental space travel, and the ability to attribute mental states to others in shared, intersubjective universes of discourse. Having completed our analysis of the expansion of

cognitive architectures supporting POLE-1 through POLE-7, we turn to the Post-POLE stages of the polycentric evolution of Simple Protolanguages (PL) → Expanded Protolanguages (EPL) → Early Languages (EL) → Evolutionarily Modern Languages (EML). We see this process as directional, but we emphasize that the stages that we postulate are not defined by fixed boundaries.

- **CA6) Simple Protolanguages (PL) → Expanded Protolanguages (EPL).** *Simple protolanguages* enabled teaching and social coordination based on conventionalized protowords, whether signed or spoken. *Expanded protolanguages* emerged with increased displacement in the use of protowords and the emergence of simple combinatorics to support limited co-ordinate constructions and basic narratives that convey event sequences. In the process, protowords became words or yielded words by processes such as fractionation to become identified by their roles in certain constructions.
- **CA7) Expanded Protolanguages (EPL) → Early Languages (EL).** With progressive lexicalization and grammaticalization but no sharp transition, expanded protolanguages shaded into *early languages* with nominal modification and phrasal recursion and the development of narratives expressing causality, motivation, and individual and group identity.
- **CA8) Early Languages (EL) → Evolutionarily Modern Languages (EML).** *Evolutionarily Modern Languages* (EML) emerge through differentiated socio-cultural organization that allows for, and is supported by, complex multimodal narratives, blending everyday experiences with the social imaginary and facilitating the construction of complex symbolic identities.

In terms of Ferretti's triad (§6.1), we insist on a coevolutionary spiral linking (proto)language to the increasing emergence of autobiographical memory as mental time travel emerged from memory of single behavior-relevant episodes in concert with the development of folk psychology and mind reading and the emergence of the ability to extend the scope of navigation from "real" to imaginary worlds. In the process, the subtlety of causal cognition expanded with the availability of languages with which to discuss it as humans became both *Homo quaerens* and *Homo narrans*.

- **CA9) The technosphere increasingly supports the semiosphere.** This catch-all builds on the observation that language is only one, if the most socio-cognitively flexible, part of the semiosphere. Drawing, writing, photography, and video and, increasingly, computers all support an expanding semiosphere.

We saw in HOLO 1 that diverse artifactual and natural objects, constructed spaces, and enacted rituals structure human behavior, whether or not their meanings are also expressed in words. As the semiosphere expands, along with innovations in the technosphere, languages change to include technical registers and jargons to express data and arguments in the sciences, and the languages of law, theology, philosophy, and more. We explored relevant artifacts in §3.2, where we noted how archewriting may have

supported the "recent" emergence of writing, and the later emergence of number systems, and "languages" for the specialized expression of, e.g., mathematical reasoning and chemical notation.

7. TOWARDS A NEW TIMELINE

Our two papers together offer a comprehensive account of the open-endedness of the linked evolution of semiosphere and technosphere. The top panel of Figure 2 offers a speculative timeline for the HOLO 2 account of language origins. We align this protolanguage-to-language timeline with four timelines based primarily on HOLO 1 (d'Errico et al. 2027 [this issue]) for selected population varieties in the human lineage, major geographic expansions of hominins starting from Africa, material sign forms and types, and tool technologies.

We emphasize again the inherently polycentric nature of this process. Evolutionary steps occurred asynchronously across different regions and populations, reflecting diverse ecological and cultural contexts and ensuing needs that may or may not have been addressed by a complexification in material culture and communication systems. Some traits may have spread through cultural contact, while others emerged independently in separate processes. Innovations were lost in some populations and reacquired later, further complicating the evolutionary sequence. Nonetheless, the cumulative effect is that, despite individual variations, all contemporary humans have the cognitive architecture charted in §6, no matter what their material and linguistic circumstances.

7.1. LCA-m AND LCA-ga

Monkeys have continued evolving since LCA-m and great apes have continued to evolve since LCA-ga but we used shared abilities of monkeys and great apes, respectively, to hypothesize the following:

- **POLE-1. LCA-m (around 25 Mya):** LCA-m had manual dexterity, an ability to recognize the actions of conspecifics, and the ability to learn new manual skills, but little in the way of imitation. Communication involved a limited set of species-specific calls and facial gestures.
- **POLE-2. LCA-ga (around 7 Mya):** LCA-ga preserved the ability for dexterous manual actions and for recognition of actions by others, plus a *simple imitation system* for emulation of manual actions based on trying to replicate achievement of subgoals than the means used by another to achieve them. A small innate repertoire of manual gestures could be augmented by novel communicative gestures developed within a dyad or group.

7.2. 3,000,000 TO 800,000 YEARS AGO

Primitive versions of the stages we chart next may be seen in present-day monkeys and great apes, but pointing behavior, for example, is unknown in these primates in the wild. Again, monkey calls that draw attention to threats are very different from drawing attention to details of ongoing behavior let alone developing and living within the complex physical and symbolic worlds of modern humans.

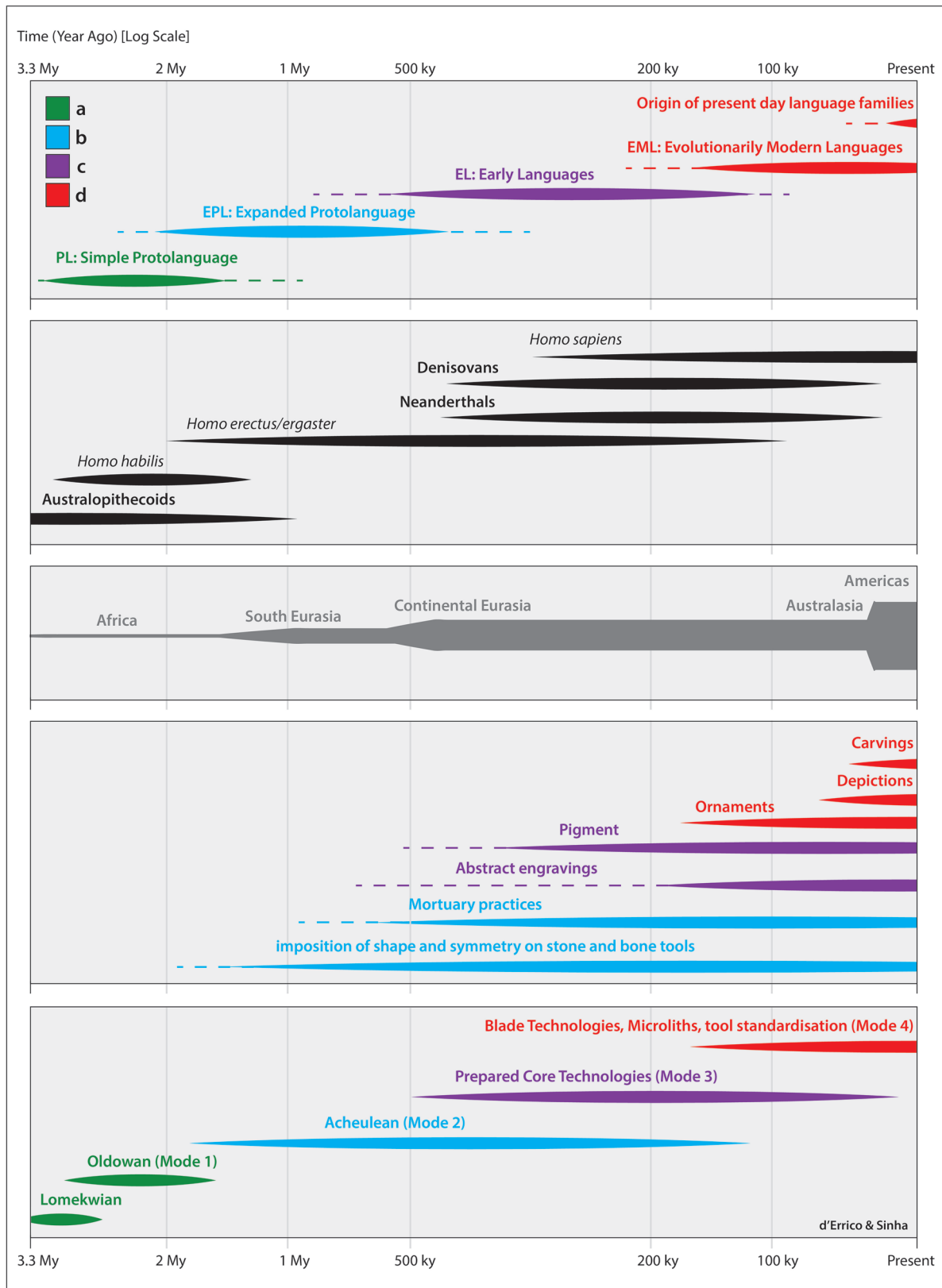


Figure 2. Timelines for hypothesized Stages in Language Origins in HOLO 2 (top panel) in comparison with timelines assessed in HOLO 1 (lower panels). A broken and faint segment indicates the earliest plausible date for an innovation or taxonomic variety while a solid segment marks (in all but the top panel) the earliest securely attributed appearance of an innovation or taxonomic variety, given the available data. The (a, b, c, d) color coding of the top panel offers tentative suggestions about the level of (proto)language needed to support the activities charted in the lower panels (figure by Francesco d'Errico and Chris Sinha).

POLE-3 charted the emergence post LCA-ga of the ability for both *complex action recognition*, the ability to recognize in the behavior of others *various movement details* of actions within it (with limitations based on the familiarity of the component actions and the complexity of the assemblage), and the ability to exploit this to support *complex imitation* (and not just emulation) of observed skills, based on trying to replicate not only achievement of a goal but also the means used by another to achieve it. This was perhaps not fully developed in the Oldowan, but was probably crucial in the Acheulean and certainly in the advanced Acheulean, after 600-400 Ka (see HOLO 1, as well as Sanchez-Dehesa Galan and Pelegrin 2024; Shipton 2019).

However, imitation is not the same as teaching, whose emergence from demonstration and then pantomime was posited by POLE-4. We noted that such demonstration can precede the use of symbols understood as such, but may rely on teachers' awareness that they are showing learners how to perform a task or solve a problem. The intention of teachers is that their learners exhibit the appropriate movements for each action, and perform these actions, not a fixed ordering but reflecting the available affordances for the achievement of current subgoals. Both demonstration and pantomime must be combined with drawing attention. This does not need a protolanguage at early stages, but some indication from the practitioner to the apprentice.

Leadbeater et al. (2006) suggest that in the context of the evolution of teaching behavior, "teaching" should include not only the transfer of skills but also the transfer of concepts, rules, strategies and narratives, but we locate these much later in the timeline, perhaps less than 400,000 years ago. The levels involve increasing demands on intentions and communication and involve different degrees of cultural transmission in their temporal and social dimension.

Some early Oldowan assemblages exhibit systematically biased patterns of flake detachment that are underdetermined by the morphological variability of Oldowan cores, suggesting (Stout et al. 2010) a learned tendency to select targets in relation to the position of previous detachments. But there are long periods of relative stasis in our early history. Oldowan technology lasted for over 1 million years, overlapping the start of the Acheulean 1.6 mya, a technological niche that persisted in some regions until around 130 ka, with a 0.76 Ma stasis in the early Acheulean. The relative stasis of Oldowan tool-making exhibits transmission of a few basic discoveries rather than emergence of a general skill for tool making, suggesting that while complex action recognition may have been in place, teaching may have been very limited compared to that which supported late Acheulean developments.

7.3. 800,000 TO 400,000 YEARS AGO

Pantomime may have been instrumental in conveying how to properly grasp and use stones as hunting weapons, process and prepare animal skins for clothing, or produce fire. In lithic technology, the later phases of refined Acheulean biface production included the careful thinning of edges

as a means to achieving symmetrical forms, and preparing striking platforms. These steps require a high degree of precision and an understanding of the mechanical properties of stone, making pantomime an effective medium for conveying subtle techniques. Subsequent conventionalization of pantomimes may have assisted the emergence of shared symbols by a group. The ability to represent some actions symbolically (perhaps in a manner that integrates both action and affordance) through gestures and sequences removed from direct interaction with the objects themselves, would have been crucial.

Pantomime of recent events could provide an early form of protonarrative. Pantomime of possible forthcoming events could be the basis for coordinated planning of group activities. But pantomimes can be ambiguous and limited in the types of meanings they can convey. They became conventionalized and then varied in arbitrary ways to convey new meanings. Protowords, whether spoken or signed, began to cohere into larger systems of communication, starting perhaps with limited micro-protolanguages. POLE-7 is the crucial point in the evolution of symbol-readiness—the Great Move of being able to freely create novel communicative acts while meeting the parity condition that at least some observers can "get the message" and, if desired, master the use of this protoword in appropriate circumstances.

As we have noted, the relative timing of flexible vocal control and learning and the development of protosign remains a topic of continuing discussion. Whatever the resolution of this debate, we argue that communicative gestures based on protosign were gradually accompanied by increasing use of vocalizations, with a progressive shift toward protospeech. HOLO 1 documents evidence that the separation between *H. sapiens* and Neanderthals may have occurred around 500 Ka to 700 Ka and that between Neanderthals and Denisovans around 400 Ka to 500 Ka. This is consistent with the notion that late *Homo erectus* populations had language-ready brains by 400 Ka. However, the brain of *H. erectus* differs greatly from that of *sapiens*, and so we posit that by perhaps 150 Ka, capacities for culturally evolved behaviors that had increased through Baldwinian evolution (Baldwin 1896; Waddington 1953) and the contribution of other mutations to human variation may have converted a cognitive architecture that was ready for some complexification of protolanguages to one that could support the ready sharing of innovations that could lead to early languages.

That said, we do not assume that a language-ready brain emerged abruptly. We suggest that key ingredients—such as enhanced manual control, action recognition, vocal flexibility, and shared intentionality—were already present to at least some extent well before the split and were gradually assembled over a longer evolutionary timescale. In this perspective, the period between ~600 and 400 Ka (see HOLO 1) may mark not the origin of language-readiness, but a phase in which these capacities became sufficiently integrated and stabilized to support sustained cultural communication. At present, the archaeological record, while

suggestive—see, for example, early construction behaviors such as the notched logs dated to ~476 Ka (Barham et al. 2023) and the incised shells from the *Homo erectus* site of Trinil, Java, dated to c. 540–430 Ka (Joordens et al. 2015) and discussed in HOLO 1—does not allow us to precisely identify when all components were first in place.

Our claim is that these elements provided the cognitive foundations necessary for cultural evolution to begin to predominate over biological (gene-based) evolution, allowing successive generations to develop protolanguages of increasing complexity, eventually leading by 400 kya to hominins who could use expanded protolanguages.

7.4. 400,000 TO 150,000 YEARS AGO

Middle Stone Age and Middle Paleolithic technologies arose by 400 Ka in Africa and Eurasia. This involved a shift from shaped bifacial tools to the removal from a prepared core of flakes to yield a predetermined morphology to be used as such or retouched to create a variety of tools used for cutting, scraping, and sawing. This happened in some areas and not in others. These tools were used to make other tools such as wooden artifacts. By c. 500 Ka there is evidence that stone tools were hafted and by 110 Ka that bone tools were fully shaped with techniques specific to the material, such as scraping, grinding, and incising. Protolanguages became more powerful, expanding in complexity of lexicon and accompanying constructions, thus further diminishing the role for pantomimes. However, pantomimes continue to be used to the present day, whether explicitly (when a word does not come to mind, a pantomime may be at hand) or in the form of co-speech (or co-sign) gestures. This may have been when expanded protolanguages increasingly combined symbols using simple constructions to link a few protowords. This might have already supported enriched narratives and these might have involved further components of the language-ready (and, we have argued in §3.2, culture-ready) brain. As suggested in §6.1, this might have provided the means to extend mental time travel, mindreading, and empathy, with hominins progressing to early languages.

7.5. 160,000 TO 60,000 YEARS AGO

In parallel with the increasing sophistication of tools, the use of mineral pigments gradually increased in the archaeological record and by 140 Ka was used to decorate shell beads, providing the earliest evidence for the emergence of human symbolic niches. The archaeological evidence reviewed in HOLO 1 supports the production of symbolic material culture (complex mortuary practices, pigment use, ornaments, abstract incisions, abstract drawings on objects and cave walls, creation of underground non-utilitarian structures) between 160 Ka and 60 Ka among populations living in Africa and Eurasia. We interpret this as evidence that various populations had the ability to share symbolically-mediated meanings. Recalling the stages of §4, we suggest that as expanded protolanguages further expanded in complexity they acquired the features that began, in different places at different times, the transition into early

modern languages. These languages supported the symbolic communication of shared meanings and expression of novel meanings, independently of pantomimes. Such developments would support caregivers and teachers as they began not merely to direct attention, but to enable learners to grasp relationships between concepts, explaining relationships between concepts with displaced reference, and the expression of causal relations between abstracted entities. Teaching by explaining such relationships involves displaced communication and causality. It relies on attributions to others of mental states, including the ability to ask questions to ascertain the knowledge states of others, and is facilitated by communicative systems that support questioning, negation, and the flexible combination of elements. The emergence of such capacities would also enable new forms of social interaction extending beyond immediate practical concerns—asking questions that may not have immediate answers (*Homo quaeerens*), as well as early forms of narrative that provide group-acceptable accounts, constituting the seeds of myth. Increasing time spent in verbal interaction—chatting, gossiping, complaining, storytelling—would further stabilize and elaborate these communicative practices. What characterizes this stage is not the first appearance of these capacities, but their progressive conventionalization, integration, and open-ended recombination within increasingly structured communicative systems. This stage was most likely achieved by Neanderthals and early *Homo sapiens*, and possibly by Denisovans, although the trajectories of these developments likely varied across populations.

7.6. 60,000 YEARS AGO TO PRESENT

HOLO 1 reviewed evidence of innovations in both technosphere and semiosphere before the Later Stone Age from Africa and Upper Paleolithic from Eurasia. This period in tool manufacturing persists until the advent of agriculture around 12 Ka. In Eurasia, the Upper Paleolithic extends from approximately 45 ka to the end of the Pleistocene; the African Later Stone Age has a different, regionally variable chronology. Bow and arrow technologies are probably present in southern Africa by 60 Ka. Eyed needles appear c. 40 Ka in Asia and 26 Ka in Europe. Microblade technologies to manufacture composite points for spears, burins for working antler, bone, and ivory, bone harpoons and spear throwers made of wood, bone, or antler appear during the Eurasian Upper Paleolithic.

This period is also marked by an unprecedented expansion of symbolic innovations, including a wide variety of personal ornaments identifying distinct social groups, complex burials with grave goods and evidence of ritual practices, extensive cave art featuring naturalistic depictions and abstract motifs, portable art objects, and the use of pigments and engravings to communicate and express identity and beliefs. Furthermore, variations in these symbolic and technological expressions suggest cultural regionalization that may reflect ethnolinguistic diversity, pointing to a complex mosaic of early human societies with distinct identities and languages. As proposed in Merlin

Donald's (1991) stage 3, humans began to construct artificial memory systems, devices specifically conceived to store, use, and recover information, and for Donald, especially spatial information, outside the physical body.

The earliest known representational paintings that can plausibly be interpreted as depicting a scene from a narrative are dated to around 50 Ka (Brumm et al. 2021; Oktaviana et al. 2024). It is likely, then, that the core constructional subsystems of evolutionarily modern languages had emerged by 60 Ka, and perhaps considerably earlier. Like all the transitions depicted in Figure 2, that from early languages to evolutionarily modern languages was polycentric and involved no clear or universal discontinuity. Our current understanding of grammaticalization (§4.1) suggests that later steps involved elaboration of clause-combining, modifier-modified and subordinating constructions. The comparative horizon for the phylogenetic reconstruction, using Bayesian methods, of the origins of today's language families is around 10,000 years (Barbieri et al. 2022; Hoffmann et al. 2021). The protolanguages (in the sense of historical linguistics) of extant language families are all dated to within this time depth. There is no reason to suppose that these were systemically less complex than their descendant languages.

7.7. RECENT ACCELERATION AND ON TO THE ANTHROPOCENE

As we look beyond the scope of the HOLO project, we cannot but be struck by the acceleration of cultural and socio-technological change, and the impact of the dramatic increase of human population and energy use per capita that has led recent centuries to be seen as the start of a new epoch dubbed the Anthropocene. Demographic and social scaling had important consequences for the normative role of language in the social construction of communities, while mechanisms ensuring intercommunity comprehension and trading must have operated for tens of millennia in small-scale societies. The emergence of larger, more interconnected communities introduced new pressures toward stabilization, standardization, and the regulation of linguistic practices, adding layers of constraint that were increasingly decoupled from immediate face-to-face interaction. The Neolithic revolution, beginning around 12 Ka, brought in agriculture and the domestication of animals (Watkins 2024) and formed the basis for the first cities to be formed in Southwest Asia, with cities appearing at different times in different parts of the world. However, some communities developed rich languages and cultural practices without passing through a "Neolithization"—as in the case of the continent-wide culture of Aboriginal Australia (Dixon 1984). On the other hand, the bringing together of large populations in cities led to a form of "energized crowding" (M.E. Smith 2023) that quickened the pace of innovation.

Writing was a late arrival that provides a prime example of polycentrism in that writing systems occurred more or less independently in the peri-Mediterranean region, in China, and in Mesoamerica. By externalizing language and making it durable, writing further enabled the explicit

codification, preservation, and institutional enforcement of linguistic conventions across time and space. In modern times, systems that increase the extent and speed of communication have led to an unprecedented increase in the tempo of technological change. Printing, which goes back 0.6 Ka in Europe, entrenched religious conflicts, transformed the economic practices of "print capitalism" (B. Anderson 2016), and drove the rise of nationalism and of standard national languages, even as it permitted the internationalization of science and the spread of news, whether "real" or "fake." The telegraph and the telephone were innovations of the 19th century. Modern computers began to spread beyond a few research centers in the 1960s, and the internet and the World Wide Web began to transform worldwide human interaction in the 1990s. The first iPhones were released only in 2007 and smartphones have since transformed human interaction patterns for billions of people. These successive and cumulative innovations have increased the complexity and density of human societies, facilitated the exchange of knowledge across vast distances and diverse cultures, and accelerated the pace of innovation. Enhanced connectivity on a global scale has allowed humans to manipulate and adapt to their environment in unprecedented ways, thus redefining the boundaries of our ecological and cultural niche. In some sense, we may say that, while certain innovations may be *polycentric* (arising independently in separated geographical regions), increasingly humanity is being transformed by *polynodal* innovations developed through the integrated activity of globally distributed experts and then widely disseminated through commercial means. All this paved the way for the "cybernetic revolution" that greatly expanded the machine component of "human-machine symbiosis" (Arbib 1976; Licklider 1960), as so dramatically exemplified in the generative AI in the 2020s, continuing to blur the distinction between technosphere and semiosphere.

ENDNOTES

¹Historical linguists use the term *protolanguage* for the shared ancestor of a family of languages, so that Latin is a protolanguage for the Romance languages. Other authors use the term for any reduced version of modern languages, such as "child language," pidgins, and the language of aphasics. We use the term only in the sense of a possible precursor in the biocultural evolution of languages.

²At San Marco in Venice, placing a round dome on a square platform created the need to fill in the gaps with surfaces (*spandrels*) that were then adopted as a place for decorative embellishment. Similarly, behavior will find new ways to exploit innovations that arose for other purposes (Gould and Lewontin 1979).

³Ironically, while Goldin-Meadow is a co-author of the recent review "What enables human language? A biocultural framework" (Arnon et al. 2025), its focus is almost completely on spoken languages—despite her expertise on signed languages.

⁴Note that this is not Saussure's (1916) dichotomy between *langue* as the rules and conventions of a signifying system independent of the individual user and *parole*, which is more than Speech 2 in that it includes spoken and written uses of *langue*.

⁵There is a subtle change here. Elsewhere, we have referred to LCA-m and LCA-ga as the last common *ancestor* (singular) of modern humans with monkeys and great apes, respectively. We now use *ancestors* (plural) to stress that there was no single primate that served as our

last common ancestor. Rather, varied species of primate interacted and evolved across millions of years to yield today's monkeys and great apes. The discussion in HOLO 1 of Neanderthal and Denisovan genes in today's *Homo sapiens* reinforces this point.

⁶Blasi et al. (2019) have argued that the phonetics of spoken languages have been reshaped by changes in the human bite configuration due to changes in dietary and behavioral practices since the Neolithic. However, such considerations are outside our scope.

⁷Żywiczyński et al. (2024) provide an edited collection of recent papers on diverse roles of pantomime from prehistory to the present day. Several of these papers are cited here.

⁸Boesch et al. (2020) studied termite-fishing in 17 chimpanzee communities and conclude that termite fishing in wild chimpanzees shows some elements of cumulative cultural diversity. However, this is a far cry from the extent of human cultural diversity.

⁹The capitalization in Deaf people refers to deaf people who are part of a signed language community.

¹⁰Anne Russon (2018; Russon and Andrews 2010) has demonstrated simple forms of pantomime in orangutans, but these seem to be based on interaction with humans combined with natural curiosity rather than taking us all the way to POLE-4, and do not involve POLE-3.

¹¹Some of the material here is developed more fully by Arbib (2024) where the distinction between ad hoc pantomime, basic pantomime (he called it ur-pantomime), and modern theatrical pantomime (Brown 2024) is introduced, as is the relation of pantomime to protonarrative. Other chapters in the book (Żywiczyński et al. 2024) provide related contributions.

¹²In ASL, which is a modern human language, this distinction works not only for the AIRPLANE/FLY pair, but is a general “morphological” principle—nouns are often denoted with constrained back and forth movements, while verbs are denoted by more continuous movements. However, it must be understood that at the stage of language evolution we consider here, there were no general notions of noun or verb or other syntactic categories, and that at first each conventionalization would have been based on the need to resolve an oft-occurring ambiguity. Later, of course, certain patterns would be noticed that could help structure innovation and learning.

¹³An important challenge for conversation theory is *turn taking*. It is outside our scope, but Rossano (2018) and Burkart et al. (2018) present relevant findings for a comparative primatology of turn taking and what is different about humans.

¹⁴An important qualification is that many “recent” technological innovations would have been impossible using only the languages of “everyday interaction.” The formulas of mathematics and the diagrams of chemistry are among the ways that the syntax of language has been expanded to unleash the increasing pace of technological change. Number systems seem to be an earlier form of “language plug-in.”

¹⁵The term *Homo quarens* has been earlier used by Feldstein (1978) in his book *Homo quarens, the Seeker and the Sought*, but his focus is very different from ours.

¹⁶*The Evolution of Nervous Systems*, now in its third edition (Kaas anderculano-Houzel 2026), has four volumes, with volume 3 devoted to non-human primates and volume 4 examining the evolution of nervous systems of humans and other hominoids.

¹⁷This distinction between a “language-ready” and an “even-more language-ready” brain was not part of MSH.

ACKNOWLEDGEMENTS

The paper is an outgrowth of the conference on “Human Origins and Language Origins” organized by Chris Sinha and held 18th–20th October 2019 at Hunan University, Changsha, China. His co-authors thank Chris for his guiding vision that shaped the conference and brought us all together to develop the ideas presented in this paper and in HOLO 1. We thank Prof. Liu Zhengguang for his help in securing financial support for the meeting from Hunan University, and Prof. Liu, as well as numerous colleagues

and students from the School of Foreign Languages, Hunan University, for their hospitality. We are most appreciative for the care of the editor and the two reviewers for their exceptionally thoughtful and constructive comments that have led us to strengthen and extend our arguments and improve the readability of the whole. Francesco D’Errico’s research was supported by the following agencies: Research Council of Norway, Centres of Excellence (SFF); Centre for Early Sapiens Behaviour, SapienCE grant# 262618; Initiative d’Excellence IdEx, University of Bordeaux, Talent program grant# 191022-001; French government in the framework of the University of Bordeaux’s IdEx “Investments for the Future” program/ GPR “Human Past”; and European Research Council Synergy Grant for the project Evolution of Cognitive Tools for Quantification (QUANTA), No. 951388, and the Association For Trans-Eurasia Exchange and Silk Road Civilization Development (ATES). Bernd Heine’s research was sponsored in part by the Korean Ministry of Education, Science and Technology within its World Class University program, Prof. Haiping Long and Sun Yatsen University, Guangzhou, China, and the University of Cape Town, South Africa.

AUTHORS’ CONTRIBUTIONS

Each author contributed drafts of key sections and Arbib edited the various drafts into a coherent document. All authors enriched successive drafts. Figure 1 was developed by Arbib, d’Errico and Sinha. Figure 2 was developed by d’Errico and Sinha.



This work is distributed under the terms of a [Creative Commons Attribution-NonCommercial 4.0 Unported License](https://creativecommons.org/licenses/by-nc/4.0/).

REFERENCES

- Aboitiz, F., 2013. How did vocal behavior “take over” the gestural communication system? *Lang. Commun.* 5, 167–176.
- Addis, D.R., Schacter, D., 2012. The hippocampus and imagining the future: where do we stand? *Front. Hum. Neurosci.* 5(173). <https://doi.org/10.3389/fnhum.2011.00173>
- Addis, D.R., Wong, A.T., Schacter, D.L., 2007. Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia* 45(7), 1363–1377. <https://doi.org/10.1016/j.neuropsychologia.2006.10.016>
- Adone, D., Gabel, A. (Eds.), 2025. *The Evolution, Acquisition and Development of Syntax: Insights from Creole Languages and Beyond*. Cambridge University Press, Cambridge.
- Allen, W.L., Cuthill, I.C., Scott-Samuel, N.E., Baddeley, R., 2011. Why the leopard got its spots: relating pattern development to ecology in felids. *Proc. Royal Soc. B: Biol. Sci.* 278(1710), 1373–1380. <https://doi.org/10.1098/rspb.2010.1734>
- Almor, A., Smith, D.V., Bonilha, L., Fridriksson, J., Rorden, C., 2007. What is in a name? Spatial brain circuits are used to track discourse references. *NeuroReport* 18,

- 1215–1219.
- Altavilla, D., 2025. How psychological and descriptive narratives modulate neural responses to emotional expressions. *Front. Hum. Neurosci.* 19, Article 12043528.
- Anderson, B., 2016. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso, London.
- Anderson, M.L., 2010. Neural reuse: a fundamental organizational principle of the brain. *Behav. Brain Sci.* 33(4), 245–266. <https://doi.org/10.1017/S0140525X10000853>
- Arbib, M.A., 1976. Artificial Intelligence: Cooperative Computation and Man-machine Symbiosis. *IEEE Trans. on Computers C-25*, 1346–1352.
- Arbib, M.A., 1991. Neurons, schemas, persons, and society. In: Earley, J.E. (Ed.), *Individuality and Cooperative Action*. Georgetown University Press, Georgetown, pp. 63–86 (see also General Discussion, pp. 179–180).
- Arbib, M.A., 2002. Grounding the Mirror System Hypothesis for the evolution of the language-ready brain. In: Cangelosi, A., Parisi, D. (Eds.), *Simulating the Evolution of Language*. Springer London, London, pp. 229–254.
- Arbib, M.A., 2005a. From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav. Brain Sci.* 28(2), 105–124; response 125–167. <http://www.ncbi.nlm.nih.gov/pubmed/16201457>
- Arbib, M.A., 2005b. Interweaving protosign and protospeech: further developments beyond the mirror. *Interact. Stud.* 6, 145–171.
- Arbib, M.A., 2012. *How the Brain Got Language: The Mirror System Hypothesis*. Oxford University Press, New York & Oxford.
- Arbib, M.A., 2021. *When Brains Meet Buildings: A Conversation Between Neuroscience and Architecture*. Oxford University Press, New York.
- Arbib, M.A., 2024. Pantomime within and beyond the evolution of language. In: Żywiczyński, P., Blomberg, J., Boruta-Żywiczyńska, M. (Eds.), *Perspectives on Pantomime*. John Benjamins Publishing Company, Philadelphia/Amsterdam, pp. 16–57.
- Arbib, M.A., 2025. From the protolanguage spectrum to the underlying bases of language. In: Adone, D., Gabel, A. (Eds.), *The Evolution, Acquisition and Development of Syntax: Insights from Creole Languages and Beyond*. Cambridge University Press, Cambridge, pp. 11–33.
- Arbib, M.A. (Ed.), 2020. *How the Brain Got Language: Towards a New Road Map*. John Benjamins Publishing Company, Amsterdam/Philadelphia.
- Arbib, M.A., Aboitiz, F., Burkart, J., Corballis, M., Coudé, G., Hecht, E., Liebal, K., Myowa-Yamakoshi, M., Pustejovsky, J., Putt, S., Rossano, F., Russon, A.E., Schoenemann, P.T., Seifert, U., Semendeferi, K., Sinha, C., Stout, D., Volterra, V., Waciewicz, S., Wilson, B., 2018. The comparative neuroprimatology 2018 (CNP-2018) road map for research on how the brain got language. *Interact. Stud.* 19(1–2), 370–387.
- Arbib, M.A., Bickerton, D. (Eds.), 2010. *The Emergence of Protolanguage: Holophrasis vs Compositionality*. John Benjamins Publishing Company, Philadelphia/Amsterdam.
- Arbib, M.A., Cuccio, V., 2026. The body in language, the language beyond the body: embrainment and graded embodiment in the evolution and use of language. *Front. Psychol.* 17. <https://doi.org/10.3389/fpsyg.2026.1774505>
- Arbib, M.A., Fellous, J.-M., 2004. Emotions: from brain to robot. *Trends Cogn. Sci.* 8(12), 554–561. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15556025
- Arbib, M.A., Frigaszy, D.M., Healy, S.D., Stout, D., 2023. Tooling and construction: from nut-cracking and stone-tool making to bird nests and language. *Curr. Res. Behav. Sci.* 5, 100121. <https://doi.org/10.1016/j.crbeha.2023.100121>
- Arbib, M.A., Gasser, B., Barrès, V., 2014. Language is handy but is it embodied? *Neuropsychologia* 55, 57–70.
- Arbib, M.A., Hesse, M.B., 1986. *The Construction of Reality*. Cambridge University Press, Cambridge.
- Arbib, M.A., Liaw, J.-S., 1995. Sensorimotor transformations in the worlds of frogs and robots. *Artif. Intell.*, 72, 53–79.
- Arbib, M.A., Liebal, K., Pika, S., 2008. Primate vocalization, gesture, and the evolution of human language. *Curr. Anthropol.* 49(6), 1053–1076. <https://doi.org/10.1086/593015>
- Armstrong, D.F., Stokoe, W.C., Wilcox, S.E., 1994. Signs of the origin of syntax. *Curr. Anthropol.* 35(4), 349–368.
- Arnon, I., Carmel, L., Claidière, N., Fitch, W.T., Goldin-Meadow, S., Kirby, S., Okanoya, K., Raviv, L., Wolters, L., Fisher, S.E., 2025. What enables human language? A biocultural framework. *Science* 390(6775), eadq8303. <https://doi.org/10.1126/science.adq8303>
- Baba, S., Hafsi, T., Hemissi, O., 2025. Intentionality and attentionality dynamics in an institutional change process. *J. Manag. Stud.* 63(6), 2700–2731.
- Barbieri, C., Blasi, D.E., Arango-Isaza, E., Sotiropoulos, A.G., Hammarström, H., Wichmann, S., Greenhill, S.J., Gray, R.D., Forkel, R., Bickel, B., Shimizu, K.K., 2022. A global analysis of matches and mismatches between human genetic and linguistic histories. *Proc. Nat. Acad. Sci. U.S.A.* 119(47), e2122084119. <https://doi.org/10.1073/pnas.2122084119>
- Barham, L., Duller, G.A.T., Candy, I., Scott, C., Cartwright, C.R., Peterson, J.R., Kabukcu, C., Chapot, M.S., Melia, F., Rots, V., George, N., Taipale, N., Gethin, P., Nkombwe, P., 2023. Evidence for the earliest structural use of wood at least 476,000 years ago. *Nature* 622(7981), 107–111. <https://doi.org/10.1038/s41586-023-06557-9>
- Barham, L., Everett, D., 2021. Semiotics and the origin of language in the Lower Palaeolithic. *J. Archaeol. Method Theory* 28(2), 535–579. <https://doi.org/10.1007/s10816-020-09480-9>
- Barnard, A., 2013. Cognitive and social aspects of language origins. In: Lefebvre, C., Comrie, B., Cohen, H. (Eds.), *New Perspectives on the Origins of Language*. John Benjamins Publishing Company, Amsterdam/Philadelphia, pp. 53–71.

- Barney, A., Martelli, S., Serrurier, A., Steele, J., 2012. Articulatory capacity of Neanderthals, a very recent and human-like fossil hominin. *Philos. Trans. R. Soc. Lond. B* 367(1585), 88–102. <https://doi.org/10.1098/rstb.2011.0259>
- Barrès, V., Lee, J.Y., 2014. Template construction grammar: from visual scene description to language comprehension and agrammatism. *Neuroinformatics* 12(1), 181–208. <https://doi.org/10.1007/s12021-013-9197-y>
- Barrett, L.F., 2017. *How Emotions Are Made. The Secret Life of the Brain*. Macmillan, London.
- Barton, R.A., 2012. Embodied cognitive evolution and the cerebellum. *Phil. Trans. R. Soc. B* 367(1599), 2097–2107. <https://doi.org/10.1098/rstb.2012.0112>
- Beniamine, S., Bonami, O., McDonough, J., 2017. When segmentation helps. Implicative structure and morph boundaries in the Navajo verb. First International Symposium on Morphology, Lille, France. <https://shs.hal.science/halshs-01955118>
- Berwick, R.C., Chomsky, N., 2015. *Why Only Us: Language and Evolution*. MIT Press, Cambridge, MA.
- Bickerton, D., 1995. *Language and Human Behavior*. University of Washington Press, Seattle.
- Bickerton, D., 2009. *Adam's Tongue. How Humans Made Language, How Language Made Humans*. Hill & Wang, New York.
- Bickerton, D., 2014. *More than Nature Needs: Language, Mind, and Evolution*. Harvard University Press, Cambridge, MA.
- Black, M., 1962. *Models and Metaphors*. Cornell University Press, Ithaca.
- Blasi, D.E., Moran, S., Moisik, S.R., Widmer, P., Dediu, D., Bickel, B., 2019. Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science* 363(6432), 1192–1192. <https://doi.org/10.1126/science.aav3218>
- Boë, L.J., Heim, L.J., Honda, K., Maeda, S., 2002. The potential Neandertal vowel space was as large as that of modern humans. *J. Phon.* 30, 465–484.
- Boesch, C., Bombjaková, D., Meier, A., Mundry, R., 2019. Learning curves and teaching when acquiring nut-cracking in humans and chimpanzees. *Sci. Rep.* 9(1), 1–14.
- Boesch, C., Kalan, A.K., Mundry, R., Arandjelovic, M., Pika, S., Dieguez, P., Ayimisin, E.A., Barciela, A., Coupland, C., Egbe, V.E., Eno-Nku, M., Fay, J.M., Fine, D., Hernandez-Aguilar, R.A., Hermans, V., Kadam, P., Kambi, M., Llana, M., Maretti, Morgan, D., Murai, M., Neil, E., Nicholl, S., Ormsby, L.J., Orume, R., Pacheco, L., Piel, A., Sanz, C., Sciaky, L., Stewart, F.A., Tagg, N., Wessling, E.G., Willie, J., Kühl, H.S., 2020. Chimpanzee ethnography reveals unexpected cultural diversity. *Nat. Hum. Behav.* 4(9), 910–916. <https://doi.org/10.1038/s41562-020-0890-1>
- Bouizegarene, N., Ramstead, M.J.D., Constant, A., Friston, K.J., Kirmayer, L.J., 2024. Narrative as active inference: an integrative account of cognitive and social functions in adaptation. *Front. Psychol.* 15, 1345480. <https://doi.org/10.3389/fpsyg.2024.1345480>
- Brentano, F., 1995. *Psychology from an Empirical Standpoint* (translated by A.C. Rancurello, D.B. Terrell, L.L. McAlister). Routledge, London.
- Brown, S., 2024. The pantomimic origins of the narrative arts. In: Żywicznyński, P., Waciewicz, S., Boruta-Żywicznyńska, M., Blomberg, J. (Eds.), *Perspectives on Pantomime: Evolution, Development, Interaction*. John Benjamins Publishing Company, Philadelphia/Amsterdam, pp. 139–158.
- Brown, S., Mittermaier, E., Kher, T., Arnold, P., 2019. How pantomime works: implications for theories of language origin. *Front. Commun.* 4, 9. <https://doi.org/10.3389/fcomm.2019.00009>
- Brumm, A., Oktaviana, A.A., Burhan, B., Hakim, B., Lebe, R., Zhao, J.X., Sulistyarto, P.H., Ririmasse, M., Adhityatama, S., Sumantri, I., Aubert, M., 2021. Oldest cave art found in Sulawesi. *Sci. Adv.* 7(3), eabd4648.
- Bruner, J., 1991. The narrative construction of reality. *Crit. Inq.* 18(1), 1–21. <https://doi.org/10.1086/448619>
- Burkart, J.M., Guerreiro Martins, E.M., Miss, F., Zuercher, Y., 2018. From sharing food to sharing information. Co-operative breeding and language evolution. *Interact. Stud.* 19(1–2).
- Byrne, A., 2006. Intentionality. In: Sarkar, S., Pfeifer, J. (Eds.), *The Philosophy of Science: An Encyclopedia*. Routledge, New York/London, pp. 405–410.
- Byrne, R.W., 2003. Imitation as behaviour parsing. *Philos. Trans. R. Soc. B* 358(1431), 529–536. <https://doi.org/10.1098/rstb.2002.1219>
- Byrne, R.W., Russon, A.E., 1998. Learning by imitation: a hierarchical approach. *Behav. Brain Sci.* 21(5), 667–684; discussion pp. 684–721. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10097023
- Caldwell, C.A., Smith, K., 2012. Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. *PLoS One* 7(8), e43807. <https://doi.org/10.1371/journal.pone.0043807>
- Capirci, O., Volterra, V., 2008. Gesture and speech. The emergence and development of a strong and changing partnership. *Gesture* 8(1), 22–44
- Castoriadis, C., 1987. *The Imaginary Institution of Society*. Polity Press, Cambridge.
- Castro, L., Castro-Cañadilla, D., Castro-Nogueira, M.Á., Toro, M.Á., 2025. Cultural accumulation or stasis? The impact of imitation and early teaching on hominin cultural evolution. *BioSystems* 258, 105641.
- Chafe, W., 1987. Cognitive constraints on information flow. In: Tomlin, R.S. (Ed.), *Coherence and Grounding in Discourse* 11. John Benjamins, Philadelphia and Amsterdam, pp. 21–51.
- Cheng, S., Werning, M., Suddendorf, T., 2016. Dissociating memory traces and scenario construction in mental time travel. *Neurosci. Biobehav. Rev.* 60, 82–89. <https://doi.org/10.1016/j.neubiorev.2015.11.011>
- Chomsky, N., 1995. *The Minimalist Program*. MIT Press, Cambridge, MA.

- Cisek, P., 2022. Evolution of behavioural control from chordates to primates. *Phil. Trans. R. Soc. B* 377(1844), 20200522. <https://doi.org/10.1098/rstb.2020.0522>
- Colagè, I., d'Errico, F., 2020. Culture: the driving force of human cognition. *Top. Cogn. Sci.* 12(2), 654–672. <https://doi.org/10.1111/tops.12372>
- Colagè, I., d'Errico, F., 2025. An empirically based scenario for the evolution of cultural transmission in the human lineage during the last 3.3 million years. *PLoS One* 20(6), e0325059. <https://doi.org/10.1371/journal.pone.0325059>
- Conde-Valverde, M., Martínez, I., Quam, R.M., Rosa, M., Velez, A.D., Lorenzo, C., Jarabo, P., Bermúdez de Castro, J.M., Carbonell, E., Arsuaga, J.L., 2021. Neanderthals and *Homo sapiens* had similar auditory and speech capacities. *Nat. Ecol. Evol.* 5(5), 609–615. <https://doi.org/10.1038/s41559-021-01391-6>
- Corballis, M.C., 2002. *From Hand to Mouth, The Origins of Language*. Princeton University Press, Princeton, NJ.
- Corballis, M.C., 2019. Language, memory, and mental time travel: an evolutionary perspective. *Front. Hum. Neurosci.* 13, 217. <https://doi.org/10.3389/fnhum.2019.00217>
- Corina, D.P., Poizner, H., Bellugi, U., Feinberg, T., Dowd, D., O'Grady-Batch, L., 1992. Dissociation between linguistic and nonlinguistic gestural systems: a case for compositionality. *Brain Lang.* 43(3), 414–447.
- Croft, W., 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press, Oxford.
- Cross, I., Fitch, T., Aboitiz, F., Iriki, A., Jarvis, E., Lewis, J., Liebal, K., Merker, B., Stout, D., Trehub, S., 2013. Culture and evolution. In: *Arbib, M.A. (Ed.), Language, Music, and the Brain: A Mysterious Relationship*. MIT Press, Cambridge, MA, pp. 541–562.
- d'Errico, F., Banks, W.E., 2015. The archaeology of teaching: a conceptual framework. *Cambr. Archaeol. J.* 25(4), 859–866. <https://doi.org/10.1017/S0959774315000384>
- d'Errico, F., Colagè, I., 2018. Cultural exaptation and cultural neural reuse: a mechanism for the emergence of modern culture and behavior. *Biol. Theory* 13(4), 213–227. <https://doi.org/10.1007/s13752-018-0306-x>
- d'Errico, F., Doyon, L., Colagè, I., Queffelec, A., Le Vraux, E., Giacobini, G., Vandermeersch, B., Maureille, B., 2018. From number sense to number symbols. An archaeological perspective. *Phil. Trans. R. Soc. B* 373(1740), 20160518. <https://doi.org/10.1098/rstb.2016.0518>
- d'Errico, F., Stringer, C.B., 2011. Evolution, revolution or saltation scenario for the emergence of modern cultures? *Phil. Trans. R. Soc. B* 366(1567), 1060.
- d'Errico, F., Sinha, C., Arbib, M.A., Gärdenfors, P., Heine, B., Kuteva, T., Nair, R.B., 2026. Human origins and language origins 1: a polycentric scenario for biocultural niche expansion. *PaleoAnthropology* 2027:1.
- Darwin, C., 1871. *The Descent of Man and Selection in Relation to Sex*. John Murray, London.
- Darwin, C., 1872. *The Expression of the Emotions in Man and Animals* (republished in 1965). University of Chicago Press, Chicago.
- Davidson, I., 2020. Marks, pictures and art: their contribution to revolutions in communication. *J. Archaeol. Method Theory* 27, 745–770.
- Davidson, I., Nowell, A., 2021. Scenes and non-scenes in rock art. In: *Davidson, I., Nowell, A. (Eds.), Making Scenes: Global Perspectives on Scenes in Rock Art*. Berghahn Books, Oxford, UK, pp. 16–31.
- de Voogt, A., 2024. The evolution of writing systems: an introduction. In: *Lock, A., Sinha, C., Gontier, N. (Eds.), The Oxford Handbook of Symbolic Evolution*. Oxford University Press, Oxford, pp. 893–914.
- De Waal, F., 2016. *Are We Smart Enough to Know How Smart Animals are?* WW Norton & Company, New York.
- Deacon, T.W., 1997. *The Symbolic Species: The Co-Evolution of Language and the Brain*. WW Norton, New York.
- Dediu, D., Levinson, S.C., 2013. On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. *Front. Psychol.* 4, 397. <https://doi.org/10.3389/fpsyg.2013.00397>
- Dediu, D., Levinson, S.C., 2018. Neanderthal language revisited: not only us. *Curr. Opin. Behav. Sci.* 21, 49–55. <https://doi.org/10.1016/j.cobeha.2018.01.001>
- Dediu, D., Moisik, S.R., Baetsen, W.A., Bosman, A.M., Waters-Rist, A.L., 2021. The vocal tract as a time machine: inferences about past speech and language from the anatomy of the speech organs. *Phil. Trans. R. Soc. B* 376(1824), 20200192. <https://doi.org/10.1098/rstb.2020.0192>
- Dehaene, S., Pegado, F., Braga, L.W., Ventura, P., Filho, G.N., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., Cohen, L., 2010. How learning to read changes the cortical networks for vision and language. *Science* 330(6009), 1359–1364. <https://doi.org/10.1126/science.1194140>
- Denton, D.A., McKinley, M.J., Farrell, M., Egan, G.F., 2009. The role of primordial emotions in the evolutionary origin of consciousness. *Conscious Cogn.* 18(2), 500–514. <http://www.sciencedirect.com/science/article/B6WD0-4T6CF30-1/2/422bd72b40b127e776891ce23f8d9dc7>
- Dessalles, J.-L., 2007. *Why We Talk: The Evolutionary Origins of Language*. Oxford University Press, Oxford, UK.
- di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G., 1992. Understanding motor events: a neurophysiological study. *Exp. Brain Res.* 91(1), 176–180. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1301372
- Dingemanse, M., Torreira, F., Enfield, N.J., 2013. Is “huh?” a universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PLoS One* 8(11), e78273.
- Dixon, R.M.W., 1984. *Searching for Aboriginal Languages: Memoirs of a Field Worker*. Chicago University Press, Chicago.
- Donald, M., 1991. *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Harvard

University Press, Cambridge, MA.

- Donald, M., 1998. Mimesis and the executive suite: missing links in language evolution. In: Hurford, J.R., Studdert-Kennedy, M., Knight, C. (Eds.), *Approaches to the Evolution of Language: Social and Cognitive Bases*. Cambridge University Press, Cambridge, pp. 44–67.
- Donald, M., 1999. Preconditions for the evolution of protolanguages. In: Corballis, M.C., Stephen, E.G. (Eds.), *The Descent of Mind: Psychological Perspectives on Hominid Evolution*. Oxford University Press, Oxford, pp. 138–154.
- Donald, M., 2012. Evolutionary origins of autobiographical memory: a retrieval hypothesis. In: Berntsen, D., Rubin, D.C. (Eds.), *Understanding Autobiographical Memory: Theories and Approaches*. Cambridge University Press, Cambridge, pp. 269–289.
- Dor, D., Jablonka, E., 2014. Why we need to move from gene-culture co-evolution to culturally-driven co-evolution. In: Dor, D., Knight, C., Lewis, J. (Eds.), *The Social Origins of Language*. Oxford University Press, Oxford, pp. 15–30.
- Draper, B.A., Collins, R.T., Brolio, J., Hanson, A.R., Rise-man, E.M., 1989. The schema system. *Int. J. Comput. Vis.* 2, 209–250.
- Dunbar, R., 1993. Co-evolution of neocortex size, group size and language in humans. *Behav. Brain Sci.* 16, 681–735.
- Durkheim, E., 2013. *The Rules of Sociological Method: And Selected Texts on Sociology and Its Method*, 2nd edition (ed. Steven Lukes, trans. W.D. Halls). Palgrave Macmillan, Basingstoke.
- Ekman, P., 1998. Afterword: Universality of emotional expression? A personal history of the dispute. In: *The Expression of the Emotions in Man and Animals* by Charles Darwin, 3rd edition. Oxford University Press, New York, pp. 363–393.
- Elbers, L., 1988. New names from old words: related aspects of children's metaphors and word compounds. *J. Child Lang.* 15, 591–617.
- Ellison, T.M., Reinöhl, U., 2024. Compositionality, metaphor, and the evolution of language. *Int. J. Primatol.* 45(3), 703–719. <https://doi.org/10.1007/s10764-022-00315-w>
- Evans, N., 2003. Culture and structuration in the languages of Australia. *Ann. Rev. Anthropol.* 32, 13–40.
- Evans, N., 2017. Did language evolve in multilingual settings? *Biol. Philos.* 32(6), 905–933. <https://doi.org/10.1007/s10539-018-9609-3>
- Evans, N., Levinson, S.C., 2009. The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* 32(05), 429–448. <https://doi.org/10.1017/S0140525X0999094X>
- Fauconnier, G., Turner, M., 1998. Conceptual integration networks. *Cogn. Sci.* 22(2), 133–187.
- Fay, N., Lim, S., 2012. From hand to mouth: an experimental simulation of language origin. In: Smith, A.D., Schouwstra, M. de Boer, B., Smith, K. (Eds.), *The Evolution of Language*. World Scientific, London, pp. 401–402. https://doi.org/10.1142/9789814295222_9789814290065
- Fay, N., Walker, B., 2024. Iconicity and its contribution to gesture-first theories of language origin. In: Fischer, O., Akita, K., Perniss, P. (Eds.), *Oxford Handbook of Iconicity in Language*. Oxford University Press, Oxford, pp. 697–709.
- Ferretti, F., 2021. The narrative origins of language. In: Gontier, N., Lock, A., Sinha, C. (Eds.), *Oxford Handbook of Human Symbolic Evolution*. Oxford University Press, Oxford, pp. 679–700. <https://doi.org/10.1093/oxford-hb/9780198813781.9780198813013.9780198813733>
- Ferretti, F., 2024. Agonistic conversation: a cognitive-interactive perspective on the origin of grammar. In: Adornetti, I., Ferretti, F. (Eds.), *Introducing Evolutionary Pragmatics: How Language Emerges from Use*. Routledge, New York-London, pp. 124–149.
- Ferretti, F., 2025. Stories come first. The origins of human communication from a naturalised perspective of language. *Riv. Internazionale Filos. Psicol.* 16(1), 1–16.
- Ferretti, F., Adornetti, I., 2021. Persuasive conversation as a new form of communication in “*Homo sapiens*”. *Phil. Trans. R. Soc. B* 376(1824), 20200196. <https://doi.org/10.1098/rstb.2020.0196>
- Ferretti, F., Adornetti, I., 2024. Narrative pantomime at the origin of language. In: Żywiczyński, P., Blomberg, J., Boruta-Żywiczyńska, M. (Eds.), *Perspectives on Pantomime*. John Benjamins Publishing Company, Philadelphia/Amsterdam, pp. 78–99.
- Fillmore, C.J., Kay, P., 1993. *Construction Grammar Coursebook (Reading Materials for Ling. X20)*. University of California, Berkeley.
- Fodor, J.A., 1975. *The Language of Thought*. Harvard University Press, Cambridge, MA.
- Fogassi, L., Coudé, G., Ferrari, P.F., 2013. The extended features of mirror neurons and the voluntary control of vocalization in the pathway to language. *Lang. Cogn.* 5, 145–155.
- Forster, E.M. 1927. *Aspects of the Novel*. Penguin Classics, London.
- Fragaszy, D.M., Eshchar, Y., Visalberghi, E., Resende, B., Laity, K., Izar, P., 2017. Synchronized practice helps bearded capuchin monkeys learn to extend attention while learning a tradition. *Proc. Nat. Acad. Sci. U.S.A.* 114(30), 7798–7805. <https://doi.org/10.1073/pnas.1621071114>
- Gallagher, S., Hutto, D., 2008. Understanding others through primary interaction and narrative practice. In: Zlatev, J., Racine, T.P., Sinha, C., Itkonen, E. (Eds.), *The Shared Mind: Perspectives on Intersubjectivity*. John Benjamins Publishing Company, Amsterdam/Philadelphia, pp. 17–38.
- Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G., 1996. Action recognition in the premotor cortex. *Brain* 119, 593–609.
- Gallese, V., Goldman, A., 1998. Mirror neurons and the simulation theory of mind-reading. *Trends Cognit. Sci.* 2, 493–501.

- Gamble, C., 1994. *Timewalkers: The Prehistory of Global Colonization*. Harvard University Press, Cambridge, MA.
- Gärdenfors, P., 1995. Cued and detached representations in animal cognition. *Behav. Process.* 35(1–3), 263–273.
- Gärdenfors, P., 2017. Demonstration and pantomime in the evolution of teaching. *Front. Psychol.* 8, 415. <https://doi.org/10.3389/fpsyg.2017.00415>
- Gärdenfors, P., 2021. Demonstration and pantomime in the evolution of teaching and communication. *Lang. Commun.* 80, 71–79. <https://doi.org/10.1016/j.langcom.2021.06.001>
- Gärdenfors, P., 2024. The relations of demonstration and pantomime to causal reasoning and event cognition. In: Żywiczyński, P., Blomberg, J., Boruta-Żywiczyńska, M. (Eds.), *Perspectives on Pantomime*. John Benjamins Publishing Company, Philadelphia/Amsterdam, pp. 58–77.
- Gärdenfors, P., Ekström, A., 2026. Sound iconicity in the transition from protosign to protospeech. *Front. Lang. Sci.* 5, 1864222. <https://doi.org/10.3389/flang.2026.1864222>
- Gärdenfors, P., Högborg, A., 2017. The Archaeology of Teaching and the Evolution of *Homo docens*. *Curr. Anthropol.* 58(2), 188–208. <https://doi.org/10.1086/691178>
- Gärdenfors, P., Jost, J., Warglien, M., 2018. From actions to effects: three constraints on event mappings. *Front. Psychol.* 9, 1391. <https://doi.org/10.3389/fpsyg.2018.01391>
- Gärdenfors, P., Lombard, M., 2020. Technology led to more abstract causal reasoning. *Biol. Philos.* 35(4), 40.
- Gärdenfors, P., Lombard, M., 2024. The evolution of human causal cognition. In: Wynn, T., Overmann, K.A., Coolidge, F.L. (Eds.), *The Oxford Handbook of Cognitive Archaeology*. Oxford University Press, Oxford, pp. 277–298.
- Goldberg, A.E., 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. The University of Chicago Press, Chicago.
- Goldin-Meadow, S., 2024. Gesture is an intrinsic part of modern-day human communication and may always have been so. In: Gontier, N., Lock, A., Sinha, C. (Eds.), *The Oxford Handbook of Human Symbolic Evolution*. Oxford University Press, Oxford, pp. 759–784.
- Goldman-Rakic, P., 1991. Parallel systems in the cerebral cortex: the topography of cognition. In: Arbib, M.A., Robinson, J.A. (Eds.), *Natural and Artificial Computation*. MIT Press, Cambridge, MA, pp. 155–176.
- Gould, S.J., Lewontin, R.C., 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B* 205, 581–598.
- Graham, K.E., Rossano, F., Moore, R.T., 2024. The origin of great ape gestural forms. *Biol. Rev.* 100(1), 190–204. <https://doi.org/10.1111/brv.13136>
- Gunasekaram, C., Battiston, F., Sadekar, O., Padilla-Iglesias, C., van Noordwijk, M.A., Furrer, R., Manica, A., Bertranpetit, J., Whiten, A., van Schaik, C.P., Viničius, L., Migliano, A.B., 2024. Population connectivity shapes the distribution and complexity of chimpanzee cumulative culture. *Science* 386(6724), 920–925. <https://doi.org/10.1126/science.adk3381>
- Haspelmath, M., Dryer, M.S., Gil, D., Comrie, B., 2005. *The World Atlas of Language Structures*. Oxford University Press, Oxford.
- Heine, B., 1997. *Cognitive Foundations of Grammar*. Oxford University Press, Oxford.
- Heine, B., 2002. On the role of context in grammaticalization. In: Wischer, I., Diewald, G. (Eds.), *New Reflections on Grammaticalization*. Typological Studies in Language 49. John Benjamins Publishing Company, Philadelphia/Amsterdam, pp. 83–101.
- Heine, B., 2023. *The Grammar of Interactives*. Oxford University Press, Oxford.
- Heine, B., Kaltenböck, G., Kuteva, T., 2013. On the origin of grammar. In: Lefebvre, C., Comrie, B., Cohen, H. (Eds.), *New Perspectives on the Origins of Language*. John Benjamins Publishing Company, Philadelphia/Amsterdam, pp. 379–405.
- Heine, B., Kuteva, T., 2002. *World Lexicon of Grammaticalization*. Cambridge University Press, Cambridge.
- Heine, B., Kuteva, T., 2007. *The Genesis of Grammar: A Reconstruction*. Oxford University Press, Oxford.
- Heine, B., Kuteva, T., 2011. Grammaticalization theory as a tool for reconstructing language evolution. In: Tallerman, M., Gibson, K. (Eds.), *The Oxford Handbook of Language Evolution*. Oxford University Press, Oxford, pp. 563–573.
- Hesse, M.B., 1980. *Revolutions and Reconstructions in the Philosophy of Science*. Indiana University Press, Bloomington, IN.
- Hesse, M.B., 1988. The cognitive claims of metaphor. *J. Specul. Philos.* 2, 1–16.
- Hewes, G.W., 1976. The current status of the gestural theory of language origin. In: Harnad, S., Steklis, H.D., Lancaster, J. (Eds.), *Origins and Evolution of Language and Speech*. New York Academy of Sciences, New York, pp. 482–504.
- Hobaiter, C., Byrne, R.W., 2014. The meanings of chimpanzee gestures. *Curr. Biol.* 24(14), 1596–1600. <http://dx.doi.org/10.1016/j.cub.2014.05.066>
- Hobsbawm, E., Ranger, T. (Eds.), 1983. *The Invention of Tradition*. Cambridge University Press, Cambridge.
- Hoffmann, K., Bouckaert, R., Greenhill, S.J., Kühnert, D., 2021. Bayesian phylogenetic analysis of linguistic data using BEAST. *J. Lang. Evol.* 6(2), 119–135. <https://doi.org/10.1093/jole/lzab005>
- Hoppitt, W.J.E., Brown, G.R., Kendal, R., Rendell, L., Thornton, A., Webster, M.M., Laland, K.N., 2008. Lessons from animal teaching. *Trends Ecol. Evol.* 23, 486–493.
- Hutto, D., 2000. The narrative practice hypothesis: origins and applications of folk psychology. *Royal Inst. Philos. Suppl.* 60, 43–68.
- Indurkha, B., 1992. *Metaphor and Cognition*. Kluwer Academic Publishers, Dordrecht.
- Innocenti, F., Rozzi, R., 2025. An evolutionary analysis of narrative selection. arXiv preprint: arXiv:2508.03540.
- Iriki, A., Taoka, M., 2012. Triadic (ecological, neural, cog-

- nitive) niche construction: a scenario of human brain evolution extrapolating tool use and language from the control of reaching actions. *Phil. Trans. R. Soc. B* 367, 10–23.
- Iverson, J.M., Thelen, E., 1999. Hand, mouth and brain: the dynamic emergence of speech and gesture. *J. Conscious. Stud.* 6(11-12), 19–40.
- Jablonka, E., Lamb, M.J., 2005. *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life.* Bradford Books/MIT Press, Cambridge, MA.
- Jackendoff, R., 1999. Possible stages in the evolution of the language capacity. *Trends Cogn. Sci.* 3(7), 272–279.
- Jackendoff, R., 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford University Press, Oxford and New York.
- Jackson, J.H., 1878–79. On affections of speech from disease of the brain. *Brain* 1, 304–330.
- Joordens, J.C.A., d’Errico, F., Wesselingh, F.P., Munro, S., de Vos, J., Wallinga, J., Ankjærgaard, C., Reimann, T., Wijbrans, J.R., Kuiper, K.F., Mûcher, H.J., Coqueugniot, H., Prié, V., Joosten, I., van Os, B., Schulp, A.S., Panuel, M., van der Haas, V., Lustenhouwer, W., Reijmer, J.J.G., Roebroeks, W., 2015. *Homo erectus* at Trinil on Java used shells for tool production and engraving. *Nature* 518(7538), 228–231. <https://doi.org/10.1038/nature13962>
- Kaas, J.H., Herculano-Houzel, S. (Eds.), 2026. *Evolution of Nervous Systems* (3rd edition). Academic Press, Cambridge, MA.
- Kegl, J., Senghas, A., Coppola, M., 1999. Creation through contact: sign language emergence and sign language change in Nicaragua. In: DeGraff, M. (Ed.), *Comparative Grammatical Change: The Intersection of Language Acquisition, Creole Genesis, and Diachronic Syntax.* MIT Press, Cambridge, MA, pp. 179–237.
- Kemmerer, D., 2005. Against innate grammatical categories. *Behav. Brain Sci.* <http://www.bbsonline.org/Preprints/Arbib-05012002/Supplemental/>
- Kimura, D., 1993. *Neuromotor mechanisms in human communication.* Oxford University Press, New York, Oxford.
- Klima, E., Bellugi, U., 1979. *The Signs of Language.* Harvard University Press, Cambridge, MA.
- Kline, M.A., 2015. How to learn about teaching: an evolutionary framework for the study of teaching behavior in humans and other animals. *Behav. Brain Sci.* 38, e14000031. <https://doi.org/10.1017/S0140525X14000090>
- Knight, S., Liu, L., Kornish, L.J., 2025. Building persuasive stories with emotion sequences. *Marketing Science Institute Working Paper* (No. 25-124). <http://dx.doi.org/10.2139/ssrn.5489708>
- Koelsch, S., Jacobs, A.M., Menninghaus, W., Liebal, K., Klann-Delius, G., von Scheve, C., Gebauer, G., 2015. The quartet theory of human emotions: an integrative and neurofunctional model. *Phys. Life Rev.* 13, 1–27. <https://doi.org/10.1016/j.pprev.2015.03.001>
- Krifka, M., 2008. Basic notions of information structure. *Acta Linguist. Hung.* 55(3-4), 243–276.
- Labov, W., Waletzky, J., 1967. Narrative analysis: oral versions of personal experience. In: Helm, J. (Ed.), *Essays on the Verbal and Visual Arts.* University of Washington Press, Seattle, pp. 1912–1924.
- Lakoff, G., Johnson, M., 1980. *Metaphors We Live By.* University Of Chicago Press, Chicago.
- Laland, K.N., Odling-Smee, J., Feldman, M.W., 2000. Niche construction, biological evolution, and cultural change. *Behav. Brain Sci.* 23(01), 131–146. <https://doi.org/10.1017/S0140525X00002417>
- Langacker, R.W., 1987. *Foundations of Cognitive Grammar, volume 1. Theoretical Prerequisites.* Stanford University Press, Stanford.
- Leavens, D.A., Tagliabue, J.P., Hopkins, W.D., 2014. From grasping to grooming to gossip: innovative use of chimpanzee signals in novel environments supports both vocal and gestural theories of language origins. In: Pina, M., Gontier, N. (Eds.), *The Evolution of Social Communication in Primates: A Multidisciplinary Approach.* Springer International Publishing, Cham, pp. 179–194.
- LeDoux, J., 2012. Rethinking the emotional brain. *Neuron* 73(4), 653–676. <http://dx.doi.org/10.1016/j.neuron.2012.02.004>
- Leigh, S.R., 1992. Cranial capacity evolution in *Homo erectus* and early *Homo sapiens*. *Am. J. Phys. Anthropol.* 87(1), 1–13. <https://doi.org/10.1002/ajpa.1330870102>
- Leslie, A.M., 1987. Pretense and representation in infancy: the origins of “theory of mind”. *Psychol. Rev.* 94, 84–106.
- Lewis, J., 2013. A cross-cultural perspective on the significance of music and dance to culture and society insight from BaYaka pygmies. In: Arbib, M.A. (Ed.), *Language, Music, and the Brain: A Mysterious Relationship.* MIT Press, Cambridge, MA, pp. 45–65.
- Li, L., Comi, T.J., Bierman, R.F., Akey, J.M., 2024. Recurrent gene flow between Neanderthals and modern humans over the past 200,000 years. *Science* 385(6705), eadi1768.
- Li, Q., Chen, Z., Wu, T., Shen, C., Yang, Y., Zhao, X., 2026. Exploring the effects of narrative augmented reality in cultural heritage: a study on emotional arousal, intrinsic motivation, cognitive load, and learning. *Int. J. Hum. Comput. Interact.* 42(1), 16–26. <https://doi.org/10.1080/10447318.2025.2504185>
- Licklider, J., 1960. Man-machine symbiosis. *IRE Trans. Hum. Factors Electron.* 1, 4–11.
- Liebenberg, L., 1990. *The Art of Tracking the Origin of Science.* David Philip, Claremont, South Africa.
- Lotman, Y., 1990. *Universe of the Mind: A Semiotic Theory of Culture.* (Transl. Ann Shukman). I.B. Tauris and Co. Ltd., London.
- Lyn, H., 2024. Kanzi or can’t he? Animal language studies. In: Lock, A., Sinha, C., Gontier, N. (Eds.), *The Oxford Handbook of Symbolic Evolution.* Oxford University Press, Oxford, pp. 611–643.
- MacNeilage, P.F., Davis, B.L., 2005. The frame/content theory of evolution of speech: a comparison with a gestural-origins alternative. *Interact. Stud.* 6(2), 173–199.

- <http://www.sign-lang.uni-hamburg.de/BibWeb/LiDat.acgi?ID=63566>
- Maestre, M.M., Vicente, M., Lloret, E., Suárez Cueto, A., 2023. Extracting narrative patterns in different textual genres: a multilevel feature discourse analysis. *Inf. Manag.* 14, 28.
- Mandler, G., 1984. *Mind and Body: The Psychology of Emotion and Stress*. Norton, New York.
- Mangiaterra, V., Pietro, C.B. d. S., Canal, P., Bambini, V., 2026. Metaphors' journeys across time and genre: tracking the evolution of literary metaphors with temporal embeddings. *arXiv preprint, arXiv:2602.13701*.
- Marshall, J., Atkinson, J., Smulovitch, E., Thacker, A., Woll, B., 2004. Aphasia in a user of British Sign Language: dissociation between sign and gesture. *Cogn. Neuropsychol.* 21(5), 537–554. <https://doi.org/10.1080/02643290342000249>
- Mineiro, A., Moita, M., 2024. The pantomime roots of Sao Tome and Principe Sign Language. In: Żywicznyński, P., Blomberg, J., Boruta-Żywicznyńska, M. (Eds.), *Perspectives on Pantomime*. John Benjamins Publishing Company, Philadelphia/Amsterdam.
- Mithen, S., 2024. *The Language Puzzle: Piecing Together the Six-Million-Year Story of How Words Evolved*. Hachette, London, UK.
- Morgan, T.J.H., Feldman, M.W., 2024. Human culture is uniquely open-ended rather than uniquely cumulative. *Nat. Hum. Behav.* 9(1), 28–42. <https://doi.org/10.1038/s41562-024-02035-y>
- Morgan, T.J.H., Uomini, N.T., Rendell, L.E., Chouinard-Thuly, L., Street, S.E., Lewis, H.M., Cross, C.P., Evans, C., Kearney, R., de la Torre, I., Whiten, A., Laland, K.N., 2015. Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nat. Commun.* 6(1), 6029. <https://doi.org/10.1038/ncomms7029>
- Müller, G.B., 2007. Evo-devo: extending the evolutionary synthesis. *Nat. Rev. Genet.* 8(12), 943–949. <https://doi.org/10.1038/nrg2219>
- Nair, R.B., 2003a. The Aesthetics of the Ordinary: Evoking Narrative Wonder Within the Linear Grammar of Modernity. *Forum on Indian Representations 1 & 2*, Evam, pp. 266–288.
- Nair, R.B., 2003b. *Narrative Gravity: Conversation, Cognition, Culture*. Routledge, London.
- Nair, R.B., 2009. *Poetry in a Time of Terror: Essays in the Postcolonial Preternatural*. Oxford University Press, Oxford.
- Nair, R.B., 2011a. The Nature of Narrative: Schemes, Genes, Memes, Dreams and Screams! In: Geertz, A.W., Jensen, J.S. (Eds.), *Religious Narrative, Cognition and Culture: Image and Word in the Mind of Narrative*. Equinox Series in Religion, Cognition and Culture, London, pp. 117–146.
- Nair, R.B., 2011b. Thinking out of the story box: creative writing and narrative culture in South Asia. *Text* 10, 1–22.
- Nair, R.B., 2012. Philological angst: or how the narrative of census, caste and race in India still informs the discourse of the 21st century. In: Messling, M., Ette, O., Fink, W. (Eds.), *WortMachtStamm: Rassismus und Determinismus in der Philologie 18/19 Jh.* Wilhelm Fink, Munich, pp. 55–87.
- Nair, R.B., 2014. Narrative as a mode of explanation: evolution & emergence. In: Lissack, M., Garber, A. (Eds.), *Modes of Explanation: Affordances for Action and Prediction*. Palgrave Macmillan, London, pp. 140–154.
- Nair, R.B., 2015. Virtue, virtuosity and the virtual: experiments in the contemporary Indian English novel. In: Anjaria, U. (Ed.), *The History of the Indian Novel in English*. Cambridge University Press, Cambridge, pp. 51–66.
- Nair, R.B., 2017. Localopolis and cosmopolis: an Indian narrative. In: Coste, D. (Ed.), *Cosmopolitanism and Cross-cultural Negotiation*. Special Issue of *EU-topias* 14, 147–159.
- Nair, R.B., 2018. Putting Oneself in Someone Else's Shoes. *Innovate SDGs, Issue 1 on the Role of Creativity and Innovation in Implementing SDG*. Publication of the Permanent Mission of St Vincent and the Grenadines to the United Nations, New York.
- Nair, R.B., 2022. Body of knowledge and knowledge of the body. In: Bade, D., Love, N. (Eds.), *The Early Development Of Metaphor In Bilingual Children (Special Issue in Honor of Talbot J. Taylor in Language and Communication)*. Elsevier, Amsterdam, pp. 87–102.
- Nair, R.B., 2024. The symbolic evolution and writing of speech. In: Lock, A., Sinha, C., Gontier, N. (Eds.), *The Oxford Handbook of Symbolic Evolution*. Oxford University Press, Oxford, pp. 893–914.
- Nair, R.B., 2026. *Deception: Mind, Metaphor, Memes, and Mimicry Machines*. Bloomsbury Academic, London.
- Nelson, K., Fivush, R., 2020. The development of autobiographical memory, autobiographical narratives, and autobiographical consciousness. *Psychol. Rep.* 123(1), 71–96. <https://doi.org/10.1177/0033294119852574>
- Newell, A., 1990. *Unified Theories of Cognition*. Harvard University Press, Cambridge, MA.
- Newell, A., 1992. Précis of unified theories of cognition. *Behav. Brain Sci.* 15(3), 425–437. <https://doi.org/10.1017/S0140525X00069478>
- Newell, A., Shaw, J.C., Simon, H.A., 1959. Report on a general problem-solving program. Paper presented at the International Conference on Information Processing (sponsored by UNESCO), Paris, France.
- Newell, A., Simon, H.A., 1972. *Human Problem Solving*. Prentice-Hall, Englewood Cliffs, NJ.
- Nishimura, T., Tokuda, I.T., Miyachi, S., Dunn, J.C., Herbst, C.T., Ishimura, K., Kaneko, A., Kinoshita, Y., Koda, H., Saers, J.P.P., Imai, H., Matsuda, T., Larsen, O.N., Jürgens, U., Hirabayashi, H., Kojima, S., Fitch, W.T., 2022. Evolutionary loss of complexity in human vocal anatomy as an adaptation for speech. *Science* 377(6607), 760–763. <https://doi.org/10.1126/science.abm1574>
- Oktaviana, A.A., Joannes-Boyau, R., Hakim, B., Burhan, B., Sardi, R., Adhityatama, S., Hamrullah, Sumantri,

- I., Tang, M., Lebe, R., Ilyas, I., Abbas, A., Jusdi, A., Mahardian, D.E., Noerwidi, S., Ririmasse, M.N.R., Mahmud, I., Duli, A., Aksa, L.M., McGahan, D., Setiawan, P., Brumm A., Aubert, M., 2024. Narrative cave art in Indonesia by 51,200 years ago. *Nature* 631, 714–718. <https://doi.org/10.1038/s41586-024-07541-7>
- Olton, D.S., Becker, J.T., Handelmann, G.E., 1980. Hippocampal function: working memory or cognitive mapping? *Physiol. Psychol.* 8, 239–246.
- Ortony, A., 1979. *Metaphor and Thought*. Cambridge University Press, Cambridge.
- Paloczi-Horvath, G., 1964. *The Facts Rebel*. Secker and Warburg, London.
- Pika, S., Fröhlich, M., 2019. Gestural acquisition in great apes: the Social Negotiation Hypothesis. *Anim. Cogn.* 22(4), 551–565. <https://doi.org/10.1007/s10071-017-1159-6>
- Pinson, A., Xing, L., Namba, T., Kalebic, N., Peters, J., Oegema, C.E., Traikov, S., Reppe, K., Riesenberger, S., Maricic, T., Derihaci, R., Wimberger, P., Pääbo, S., Huttner, W.B., 2022. Human TKTL1 implies greater neurogenesis in frontal neocortex of modern humans than Neanderthals. *Science* 377(6611), eabl6422. <https://doi.org/10.1126/science.abl6422>
- Poizner, H., Klima, E., Bellugi, U., 1987. *What the Hands Reveal About the Brain*. MIT Press, Cambridge, MA.
- Polich, L., 2005. The Emergence of the Deaf Community in Nicaragua: “With Sign Language You Can Learn So Much”. Gallaudet University Press, Washington, D.C.
- Povinelli, D., Penn, D., 2011. Through a floppy tool darkly. In: McCormack, T., Hoerl, C., Butterfill, S. (Eds.), *Tool Use and Causal Cognition*. Oxford University Press, Oxford, pp. 69–88.
- Rendell, L., Boyd, R., Enquist, M., Feldman, M.W., Fogarty, L., Laland, K.N., 2011. How copying affects the amount, evenness and persistence of cultural knowledge: insights from the social learning strategies tournament. *Phil. Trans. R. Soc. B* 366(1567), 1118–1128. <https://doi.org/10.1098/rstb.2010.0376>
- Rizzolatti, G., Arbib, M.A., 1998. Language within our grasp. *Trends Neurosci.* 21(5), 188–194. <http://www.ncbi.nlm.nih.gov/pubmed/9610880>
- Roberts, S., Levinson, S.C., 2017. Conversation, cognition and cultural evolution: a model of the cultural evolution of word order through pressures imposed from turn taking in conversation. *Interact. Stud.* 18(3), 402–442.
- Rolls, E.T., 2011. Functions of human emotional memory: the brain and emotion. In: Nalbantian, S., Matthews, P.M., McClelland, J.L. (Eds.), *The Memory Process: Neuroscientific and Humanistic Perspectives*. MIT Press, Cambridge, MA, pp. 173–191.
- Rossano, F., 2018. Social manipulation, turn-taking and cooperation in apes: implications for the evolution of language-based interaction in humans. *Interact. Stud.* 19(1–2), 151–169.
- Russon, A., 2018. Pantomime and imitation in great apes: implications for reconstructing the evolution of language. *Interact. Stud.* 19(1–2), 200–215.
- Russon, A., Andrews, K., 2010. Orangutan pantomime: elaborating the message. *Biol. Lett.* 7(4), 627–630. <https://doi.org/10.1098/rsbl.2010.0564>
- Sanchez-Dehesa Galan, S., Pelegrin, J., 2024. Identifier différents niveaux de savoir-faire dans l’Acheuléen de Garba I (Éthiopie), hiatus, lacunes, absences : identifier et interpreter les vides archéologiques. In: *Actes du 29e Congrès préhistorique de France (31 mai–4 juin 2021, Toulouse)*. Société préhistorique française, Paris, pp. 65–84.
- Sandler, W. 201). *Vive la différence: sign language and spoken language in language evolution*. *Lang. Cogn.* 5(2–3), 189–203.
- Sandler, W., Aronoff, M., 2007. Is phonology necessary for language? Paper presented at the Emergence of Language Structures Workshop, Unpublished talk, University of California at San Diego (February 6, 2007).
- Saussure, F., 1916. *Cours de linguistique générale* (ed. by C. Bally and A. Sechehaye, with the collaboration of A. Riedlinger). Payot, Lausanne and Paris. (English translation by W. Baskin, *Course in General Linguistics*, Glasgow: Fontana/Collins, 1977).
- Schacter, D.L., Addis, D.R., 2020. Memory and imagination: perspectives on constructive episodic simulation. In: Abraham, A. (Ed.), *The Cambridge Handbook of the Imagination*. Cambridge University Press, Cambridge, pp. 111–131.
- Schank, R., Abelson, R., 1977. *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Lawrence Erlbaum Associates, Mahwah, NJ.
- Scherer, K.R., 2013. Emotion in action, interaction, music, and speech. In: Arbib, M.A. (Ed.), *Language, Music, and the Brain: A Mysterious Relationship*. Strüngmann Forum Reports, vol. 10. MIT Press, Cambridge, MA, pp. 107–139.
- Searle, J.R., 1979. *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge University Press, Cambridge.
- Searle, J.R. (1995). *The Construction of Social Reality*. Free Press, New York.
- Senghas, A., 2003. Intergenerational influence and ontogenetic development in the emergence of spatial grammar in Nicaraguan Sign Language. *Cogn. Dev.* 18, 511–531.
- Seyfarth, R.M., Cheney, D.L., Marler, P., 1980a. Monkey responses to three different alarm calls: evidence for predator classification and semantic communication. *Science* 210, 801–803.
- Seyfarth, R.M., Cheney, D.L., Marler, P., 1980b. Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Anim. Behav.* 28, 1070–1094.
- Shipton, C., 2019. The evolution of social transmission in the Acheulean. In: Overmann, K.A., Coolidge, F.L. (Eds.), *Squeezing Minds from Stones: Cognitive Archaeology and the Evolution of the Human Mind*. Oxford University Press, Oxford, pp. 332–354.
- Sibierska, M., Boruta-Żywiczyńska, M., Żywiczyński, P.,

- Waciewicz, S., 2022. What's in a mime? An exploratory analysis of predictors of communicative success of pantomime. *Interact. Stud.* 23(2), 289–321.
- Sinha, C., 2009. Language as a biocultural niche and social institution. In: Evans, V., Pourcel, S. (Eds.), *New Directions in Cognitive Linguistics*. John Benjamins Publishing Company, Philadelphia/Amsterdam, pp. 289–310.
- Sinha, C., 2010. Cognitive linguistics, psychology and cognitive science. In: Geeraerts, D., Cuyckens, H. (Eds.), *The Oxford Handbook of Cognitive Linguistics*. Oxford University Press, Oxford, pp. 1266–1294.
- Sinha, C., 2015a. Language and other artifacts: socio-cultural dynamics of niche construction. *Front. Psychol.* 6, 1601. <https://doi.org/10.3389/fpsyg.2015.01601>
- Sinha, C., 2015b. Ontogenesis, semiosis and the epigenetic dynamics of biocultural niche construction. *Cogn. Dev.* 36, 202–209. <https://doi.org/10.1016/j.cogdev.2015.09.006>
- Skipper, J.I., Goldin-Meadow, S., Nusbaum, H.C., Small, S.L., 2009. Gestures orchestrate brain networks for language understanding. *Curr. Biol.* 19(8), 661–667. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19327997
- Smith, C.C., Reichman, O.J., 1984. The evolution of food caching by birds and mammals. *Ann. Rev. Ecol. Syst.* 15, 329–351. <http://www.jstor.org/stable/2096952>
- Smith, M.E., 2023. *Urban Life in the Distant Past: The Prehistory of Energized Crowding*. Cambridge University Press, Cambridge.
- Smithsonian-Institution, n.d. Brains. What Does It Mean to Be Human? Retrieved October 4, 2026, from <https://humanorigins.si.edu/human-characteristics/brains>
- Sterelny, K., 2012. *The Evolved Apprentice*. MIT Press, Cambridge, MA.
- Sterelny, K., 2020. Innovation, life history and social networks in human evolution. *Phil. Trans. R. Soc. B* 375(1803), 20190497.
- Stern, D., 1985. *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology*. Basic Books, New York.
- Stibbard-Hawkes, D.N.E., 2025. Reconsidering the link between past material culture and cognition in light of contemporary hunter–gatherer material use. *Behav. Brain Sci.* 48, e1. <https://doi.org/10.1017/S0140525X24000062>
- Stokoe, W.C., 2001. *Language in Hand: Why Sign Came Before Speech*. Gallaudet University Press, Washington, D.C.
- Stout, D., 2018. Archaeology and the evolutionary neuroscience of language: the technological pedagogy hypothesis. *Interact. Stud.* 19(1–2), 256–271
- Suddendorf, T., Addis, D.R., Corballis, M.C., 2009. Mental time travel and the shaping of the human mind. *Phil. Trans. R. Soc. B* 364(1521), 1317–1324. <https://doi.org/10.1098/rstb.2008.0301>
- Suddendorf, T., Corballis, M.C., 2007. The evolution of foresight: what is mental time travel, and is it unique to humans? *Behav. Brain Sci.* 30, 299–351.
- Sutton, R.S., Barto, A.G., 2018. *Reinforcement Learning: An Introduction*, second edition (first edition, 1998). MIT Press, Cambridge, MA.
- Talmy, L., 1978. Figure and ground in complex sentences. In: Greenberg, J. (Ed.), *Universals of Human Language* vol. 4, Syntax. Stanford University Press, Stanford, CA, pp. 625–649.
- Talmy, L., 2000. *Towards a Cognitive Semantics*. MIT Press, Cambridge, MA.
- Tennie, C., Call, J., Tomasello, M., 2009. Ratcheting up the ratchet: on the evolution of cumulative culture. *Philos. Trans. R. Soc. Lond. B* 364(1528), 2405–2415.
- Thornton, A., Raihani, N.J., 2008. The evolution of teaching. *Anim. Behav.* 75, 1823–1836.
- Tomasello, M., 1999. *The Cultural Origins of Human Cognition*. Harvard University Press, Cambridge, MA.
- Tomasello, M., 2010. *Origins of Human Communication*. MIT Press, Cambridge, MA.
- Tomasello, M., Call, J., Warren, J., Frost, T., Carpenter, M., Nagell, K., 1997. The ontogeny of chimpanzee gestural signals. In: Wilcox, S., King, B., Steels, L. (Eds.), *Evolution of Communication*. John Benjamins Publishing Company, Amsterdam/ Philadelphia, pp. 224–259.
- Tomasello, M., Kruger, A.C., Ratner, H.H., 1993. Cultural learning. *Behav. Brain Sci.* 16, 495–552.
- Tomasello, M., Vaish, A., 2013. Origins of human cooperation and morality. *Ann. Rev. Psychol.* 64(1), 231–255. <https://doi.org/10.1146/annurev-psych-113011-143812>
- Trevarthen, C.B., 1979. Communication and co-operation in early infancy: a description of primary intersubjectivity. In: Bullowa, M. (Ed.), *Before Speech*. Cambridge University Press, Cambridge, pp. 321–347.
- van Mazijk, C., 2024. Iconic origins of language? An essay review of Steven Mithen's *The Language Puzzle* (2024). *Biol. Philos.* 39(4), 16.
- van Schaik, C.P., Pradhan, G.R., Tennie, C., 2019. Teaching and curiosity: sequential drivers of cumulative cultural evolution in the hominin lineage. *Behav. Ecol. Socio-biol.* 73(1), 2. <https://doi.org/10.1007/s00265-018-2610-7>
- Verhagen, A., 2010. Construal and perspectivization. In: Geeraerts, D., Cuyckens, H. (Eds.), *The Oxford Handbook of Cognitive Linguistics*. Oxford University Press, Oxford, pp. 48–81.
- Volterra, V., Capirci, O., Rinaldi, P., Sparaci, L., 2018. From action to spoken and signed language through gesture: some basic issues for a discussion on the evolution of the human language-ready brain. *Interact. Stud.* 19(1–2), 216–238.
- von Frisch, K., 1954. *The Dancing Bees: An Account of the Life and Senses of the Honey Bee*. Translated by Dora Ilse. Springer-Verlag, Vienna.
- Von Frisch, K., 1967. *The Dance Language and Orientation of Bees*. Harvard University Press, Cambridge, MA.
- Waciewicz, S., Żywicznyński, P., 2018. Language origins: fitness consequences, platform of trust, cooperation, and turn-taking. *Interact. Stud.* 19(1–2), 167–182.
- Wen, H., Wang, D., Bi, Y., 2024. Processing language partly shares neural genetic basis with processing tools and body parts. *eNeuro* 11(8), ENEURO.0138–0124.2024.

- <https://doi.org/10.1523/eneuro.0138-24.2024>
- Whitehead, C., 2010. The culture ready brain. *Soc. Cogn. Affect. Neurosci.* 5(2–3), 168–179. <https://doi.org/10.1093/scan/nsq036>
- Whiten, A., Goodall, J., McGrew, C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C.E.S., Wrangham, R., Boesch, C., 1999. Cultures in chimpanzees. *Nature* 399, 682–685.
- Whiten, A., Hinde, R.A., Laland, K.N., Stringer, C.B., 2011. Culture evolves. *Phil. Trans. R. Soc. B* 366(1567), 938–948. <https://doi.org/10.1098/rstb.2010.0372>
- Wray, A., 2000. Holistic utterances in protolanguage: the link from primates to humans. In: Knight, C., Studdert-Kennedy, M., Hurford, J. (Eds.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*. Cambridge University Press, Cambridge, pp. 285–202.
- Xu, Y., Malt, B.C., Srinivasan, M., 2017. Evolution of word meanings through metaphorical mapping: systematicity over the past millennium. *Cogn. Psychol.* 96, 41–53. <https://doi.org/10.1016/j.cogpsych.2017.05.005>
- Yao, Z., Choi, E.P.H., Chow, D.D.W.N., Chen, H., Wang, L., Ho, M.H., 2026. Effects of narrative-based interventions on self-efficacy and self-management in chronic disease: a systematic review and meta-analysis. *Appl. Psychol.: Health Well-Being* 18(1), e70127.
- Zieba, M., 2025. Narrative identity and compassion. *Front. Psychol.* 16, 1476446.
- Zlatev, J., Persson, T., Gärdenfors, P., 2005. Bodily mimesis as “the missing link” in human cognitive evolution. *Lund U. Cogn. Stud.* 121, 1–45.
- Zukow-Goldring, P., 2006. Assisted imitation: affordances, effectivities, and the mirror system in early language development. In: Arbib, M.A. (Ed.), *Action to Language via the Mirror Neuron System*. Cambridge University Press, Cambridge, UK, pp. 469–500.
- Żywicznyński, P., Blomberg, J., Boruta-Żywicznyńska, M. (Eds.), 2024. *Perspectives on Pantomime*. John Benjamins Publishing Company, Philadelphia/Amsterdam.
- Żywicznyński, P., Waciewicz, S., Lister, C., 2021. Pantomimic fossils in modern human communication. *Phil. Trans. R. Soc. B* 376(1824), 20200204. <https://doi.org/doi:10.1098/rstb.2020.0204>
- Żywicznyński, P., Waciewicz, S., Sibierska, M., 2018. Defining pantomime for language evolution research. *Topoi* 37(2), 307–318. <https://doi.org/10.1007/s11245-016-9425-9>