

Before Babel: The Evolutionary Roots of Human Language

Piera Filippi

Abstract The aim of the present work is to identify the evolutionary origins of the ability to speak and understand a natural language. I will adopt Botha's "Windows Approach" (*Language and Communication*, 2006, 26, pp. 129–143) in order to justify the following two assumptions, which concern the evolutionary continuity between human language and animals' communication systems: (a) despite the uniqueness of human language in sharing and conveying utterances with an open-ended structure, some isolated components of our linguistic competence are shared with non-human primates, grounding a line of evolutionary continuity; (b) the very first "linguistic" utterances were holistic, that is, whole bunches of sounds able to convey information despite their lack of modern syntax. I will address such suppositions through the comparative analysis of three constitutive features of human language: syntax, the semantic value of utterances, and the ability to attribute mental states to conspecifics, i.e. the theory of mind.

Keywords Language evolution • Non-human primates • Holistic protolanguage • Semantics • Syntax • Theory of mind

The aspects of things that are most important for us are hidden because of their simplicity and familiarity. (One is unable to notice something – because it is always before one's eyes.) The real foundations of their inquiry do not strike people at all. Unless that fact has at some time struck them. – And this means: we fail to be struck by what, once seen is most striking and most powerful.

(Wittgenstein 1953 [2010, § 129]).

P. Filippi (✉)
University of Vienna, Vienna, Austria

Vrije Universiteit Brussel, Ixelles, Belgium
e-mail: pie.filippi@gmail.com

Homo loquens

Words, thoughts, and reasoning are all constitutive parts of human *natural* history. Humans' ordinary life is so permeated by them that, as fate would ironically have it, they are one of the most mysterious topics of studies accessible to the human mind; mysterious and difficult, for sure, but nonetheless extremely fascinating. I believe that one of the most efficient ways to explore the nature of such a complex phenomenon as language is studying its origins' dynamics, which can shed light on those features that distinguish it from other animals' systems of communication; in short, what makes human communication unique. Indeed, for many centuries theorists of language sciences have speculated on the evolution of language, but the impossibility to find direct evidence has repeatedly led to a state of impasse. In fact, unlike other phenomena addressed by evolutionary research, language cannot be studied through paleontological data, as it has never fossilized on rocky stratifications able to indicate an evolutionary path towards *species* and time. Back in 1866, the lack of any scientific progress in the study of this topic led the *Société linguistique de Paris* to publish an edit banning any communication related to the origins of language or to the existence of a universal language that all modern languages share. However, in the last century this topic has seen a considerable revival due to the emergence of a new fertile methodology of research, in which multiple disciplines related to language and biology interact with each other (cf. Fig. 1). Within this methodological frame, the aim of the present work is to explore the origins of language bridging research on linguistics and philosophy of language to the comparative investigation of animal communication.

In order to avoid any conceptual misinterpretation, I wish to initially point out a terminological distinction which is missing in numerous spoken languages: (1) the semantic distinction between the faculty of *language*, meant in a broad sense, as a general biological tool that allows communication, (2) the ability to speak and understand a natural language. The first meaning refers to the ability to produce a visual and/or acoustic sign in association with a specific referential object. Thus,

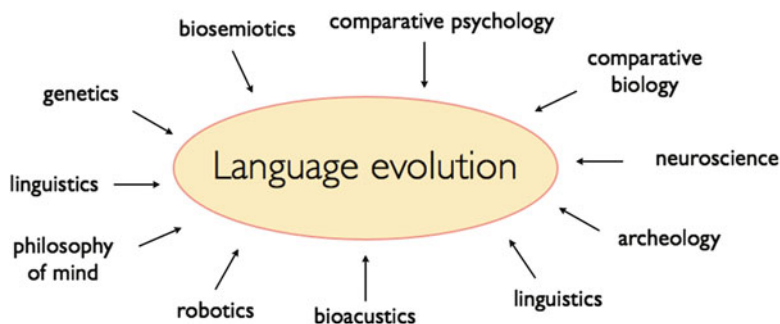


Fig. 1 The interdisciplinary approach on language evolution. In order to pinpoint the evolutionary dynamics of language, a coalition of multiple types of expertise is required. Observations from different fields are now encouraged to be integrated (modified from Christiansen and Kirby 2003)

non-human animals' cognitive and communication systems are part of this broad biological set. Notably, the faculty of language (broad sense) includes the second type of ability, i.e. to speak (or sign) a natural language such as Hindi, Chinese, or Italian following specific combinatorial and morfo-syntactical rules. This latter order of language is specific to humans, which typically employ it in a social group, conveying information or influencing behaviours. In order to clarify the distinction between these two orders of the faculty of language, it is worth taking into consideration Terrence W. Deacon's observations.¹ His hypothesis is that in animal communication systems, each sound or sequence of sounds relates to one referential object (indexical association). In contrast, what makes human language unique is that "the relationship that a word has to an object is a *function* of the relationship that word has to other linguistic units within the sentence".² This means that in human language the propositional system of linguistic units (be they morpho-syntactic elements, words, or entire sentences), which are ruled by combinatorial rules, guides the act of reference. Thus, the combinatorial dimension is one feature that makes human language unique. In fact, unlike animal referential calls, propositional languages have specific morfo-syntactic organizations. Specifically, a set of language-specific rules governs the combination of morphological and syntactic units, generating a potentially infinite set of utterances, and thus enabling much of the generative power specific to human language.³

In the present article, I will argue that in animals' communication systems, one can identify general language-related cognitive traits that were critical for the evolutionary path of propositional language. Thus, the underlying assumption is that although human language includes a set of intertwined morfo-syntactic and conceptual-intentional operations, some components of our linguistic competence, taken in isolation, are shared with other animals. Indeed, although much research has been dedicated to the individuation of one (monogenesis) or more (polygenesis) natural languages as a common root of modern spoken languages, here I wish to adopt an interdisciplinary, comparative approach with the aim to identify the biological constraints underlying the evolution of language in non-human animals' communication systems. In fact, comparative research on different species can help shed light on the biological constraints underlying the emergence of human language. I will apply a comparative framework with the aim to analyze the evolution of three constitutive components that are highly intertwined in human language, but that I will keep separate merely for methodological reasons: semantics, syntax and the ability to attribute mental states to conspecifics. Ultimately, this approach could help us grasp a better understanding of the communicative abilities shared across animals, thus shedding light on the cognitive features that make human verbal communication species-specific.

¹ Deacon 1998.

² *Ibid.*, p. 79.

³ Chomsky 1980; Yip 2006.

Methodology

In order to provide a scientifically valid contribution to the study of the evolution of language – a topic which itself tends to be an object of speculation – it is necessary to adopt a precise empirical methodology. First, it is opportune to adopt an interdisciplinary approach to the topic, linking the different theoretical observations and the empirical data within a coherent frame of concepts, which could lead to an increasing understanding of language and its evolutionary dynamics.

In this direction, the “Windows Approach to language evolution”, proposed by Rudolf P. Botha⁴ is a valid methodology, which in my opinion is able to root the research on language evolution on informative, empirically grounded theory. According to Botha, we should explore the evolution of language putting together empirical data from multiple research areas that are linked to this broad topic, for instance animal communication or archaeology. This methodological strategy enables the investigation of a phenomenon that is not directly observable empirically, as it is the case with the origins of language. To be scientifically valid, such an empirically informed theory – which the author refers to as the “window theory” – should be characterized by three basic features. First, it should be grounded on phenomena about which there is direct evidence.⁵ Second, it should be warranted, in the sense that it “has to take an empirical form which gives a systematic account of how properties of present forms of language and (properties of) stages in the emergence of language are interlinked”.⁶ Finally, the window theory should be pertinent: “[Window] inferences can be pertinent – that is, about the evolution of the ‘right entity’, namely language – only if they are underpinned by a restrictive theory of what language is”.⁷

The Nature of the First Human Utterances

Given such methodological assumptions, I will take the studies concerning monkeys and apes’ communicative systems as a conceptual window through which one can observe the phylogenetic path of human language. More specifically, I will use this conceptual frame of research in order to justify the evolutionary thesis according to which the very first “linguistic” utterances were holistic, that is to say, whole bunches of sounds able to convey information despite their lack of modern syntax.

A fertile question one could start the exploration of such issues with could be the following: shall we refer to the first *Homo* vocal units as mere representational

⁴ Botha 2006.

⁵ *Ibid.*, p. 134.

⁶ *Ibid.*, p. 137.

⁷ *Ibid.*, p. 139.

labels attached to objects in the surrounding environment, or is it not more correct to conceive of them rather as functionally referential units? There are two opposite theses that follow this later approach in the present debate about language evolution. On one hand, Dereck Bickerton's analytical model of explanation asserts that names were labels (mostly referring to environmental objects such as food or aggressors), whose increasing number and complexity consequentially gave rise to syntax.⁸

This idea collides with the holistic model of explanation, first proposed by Otto Jespersen in the early twentieth century,⁹ and recently revived by Alison Wray, according to which the first meaningful units were not mere labels, but had a complex intrinsic internal meaning: "In this holistic protolanguage the messages are semantically complex and agrammatical. [...] Simply, the whole thing means the whole thing".¹⁰ In particular, Wray's idea is that the first hominids may have communicated by means of random sequences of sound, to which they associated functionally referential meanings relying on the pragmatic context of use. The first expressions were, according to her ideas, formulaic and internally amorphous, though efficient in their performative, *manipulative* purposes.¹¹

Let us imagine a situation in which the protagonists are the very first hominids who become aware of an imminent attack from a dangerous predator, e.g. a leopard. Most likely, our very first ancestor would have given an alarm call, similar to that of the great apes. In this situation, would we translate such a vocalization not as a simple name, but rather as a more complex message with an intrinsic emotional connotation, which could lead to an appropriate reaction somehow achievable by the utterance: "I've just seen a leopard... Behave accordingly!", or "Warning, ground danger!"?

In order to address this question, I will review relevant research on the communication system of our non-human primate ancestor, with whom we share genetic traits inherited by a common ancestor. The idea is to examine three core abilities that might have grounded a line of phylogenetic continuity (and discontinuity at the same time) between monkeys' communication system and the human language: syntax, the semantic value of utterances, and the ability to attribute mental states to conspecifics, i.e. the theory of mind.

⁸ Bickerton 2002.

⁹ Jespersen 1922 [2013, Chapter XXI].

¹⁰ Wray 2002, p. 118. Cf. Fitch 2010, Chapter 14.

¹¹ Cf. Wray 1998, p. 51: "Protolanguage would, then, be a phonetically sophisticated set of formulaic utterances, with agreed function-specific meanings, that were a direct development from the earlier noises and gestures, and which had, like them, no internal structure. Each would be phonetically arbitrary, unrelated in sound to even those utterances that meant similar things".

Semantics

Regarding the semantic level, the meaning value of primate alarm calls refers to several different domains. For instance, eminent researchers on monkeys' communication system such as Robert M. Seyfarth and Dorothy L. Cheney have addressed this by defending the thesis according to which their signals are highly informative, given the agreed meaning of "information" as the reduction of uncertainty in the recipient.¹² Indeed, as they observe, the signal can be used by listeners to extrapolate information concerning the presence of food, the caller's identity, the kind of predator and the urgency of the danger. Concerning the signalling of the presence of food, it is worth noting that recent research conducted by Zanna Clay and Klaus Zuberbühler on bonobos has revealed that: "Captive bonobos at two locations produced five acoustically distinct call types when interacting with food: barks, peeps, peep-yelps, yelps and grunts. The production and distribution of these call types within a sequence was not random but was significantly associated with the preference score of the food".¹³

Similarly, alarm calls can indicate the presence of specific types of predators, and the related level of danger, eliciting the most appropriate behavioural response.¹⁴ Furthermore, by hearing the signals exchanged by two or more monkeys, the listeners can infer the kind of relationship and approach that exists between them, perceiving them as actors predisposed to behave according to specific social patterns, such as who is supposed to groom or threaten who on the basis of the affiliated dominance rank:

In groups of long-lived, highly social animals, communication and cognition are linked to fitness. To survive, avoid stress, reproduce, and raise offspring who are themselves successful, individuals need both a system of communication that allows them to influence other animals' behaviour and a system of mental representations that allows them to recognize and understand other animals' relationships. Because these mental representations concern animate creatures and are designed to predict behaviour, they include information (if rudimentary) about other individuals' mental states, and about the causal relations between one social event and another.¹⁵

These observations suggest that monkeys' vocalizations have a semantic value. At this point, however, we should address the question whether there is a strict link between the sound of the call itself and its meaning, or as it sometimes happens in human language (in the case of synonymy), whether the different calls could convey the same "meaning". Indeed, calls with similar acoustic features might elicit different responses. For instance, an eagle alarm call can lead a monkey placed on a tree to jump into a bush, while a monkey already located in a safe position does not react by moving to a different place. On the other hand, it is also true that calls with

¹²Seyfarth et al. 2010.

¹³Clay and Zuberbühler 2009, p. 1392. Cf. Hauser and Marler 1993.

¹⁴Struhsaker 1967.

¹⁵Cheney and Seyfarth 2008, p. 270.

different acoustic features elicit similar responses: a leopard growl and a monkey's alarm call elicit the same behavioural response, which is climbing up a tree. As Seyfarth and Cheney observe,¹⁶ this phenomenon tells us that the recipients' response depends either on the physical properties of the signal and on the specific information they acquire from it. Also, Zuberbühler and his colleagues¹⁷ provided evidence that female Diana monkeys do not respond to the shriek of an eagle if they are exposed to an alarm call emitted by a Diana monkey male five minutes earlier, even though these two types of signals are acoustically completely different. This suggests that Diana monkeys do not classify sound merely on the basis of their acoustic features, but also by the semantic meaning they convey. Such considerations support the hypothesis that monkeys are provided with a mental representation of the object linked to the conveyed signal.

Finally, for the purpose of our study, it is necessary to emphasise that one cannot refer to monkeys' vocalizations as to mere automatic innate reflexes:

Monkeys, then, seem genetically predisposed to give particular contexts. But this is not to say that their vocalizations are entirely reflexive and involuntary. Although their call *repertoire* may be relatively fixed, their choice of whether to call or to remain silent is more flexible. [...] There is no obligatory link between the sight of a predator and the production of an alarm.

[...] Primate vocalizations are not involuntary reflexes, impossible to suppress. They are, instead, much more like the other behaviours in which animals choose to engage. As they go about their daily lives, baboons decide whether or not to vocalize, just as they decide whether or not to groom, play or form alliance. Their behaviour depends on a complex combination of their own motivation, the particular situation at hand, and who else is involved. Primates can control whether they vocalize or not; what they cannot control are the detailed acoustic features of the calls they choose to produce.¹⁸

Thus, as clearly inferable through field observations, monkeys' vocalizations are linked to a mental representation of the referred object. In fact, it has been reported that vervet monkeys are able to suppress a vocalization, if a conspecific has previously emitted it in response to the same predator encounter. Moreover, acoustically similar vocalizations can lead to different responses, relying on the involved subjects and on the specific situation in which they occur. These data tell us that the potential meanings of monkey alarm calls are not strictly fixed to a mere genetic level, but are, in contrast, bearers of associations learned through experiences and interactions.

In addition, multiple studies have reported the use of informative calls in a wide range of animal species such as birds, frogs, rats, bats, chickens, bees.¹⁹ The pervasive presence of this core communicative feature in widely distant species indicates that the ability to convey information that favors survival in the environment, i.e. calls linked, for instance, to the presence of food, predators, sexual attraction

¹⁶Seyfarth et al. 2010.

¹⁷Zuberbühler et al. 1997 and 1999.

¹⁸Cheney and Seyfarth 2008, pp. 226–227, 233.

¹⁹Cf. Hauser 1996 for a detailed review.

or emotional state is a pivotal biological constraint shared across phylogenetically distant species.

Syntax

Recently, Peter Marler, a researcher on animal communication, revived Martinet's concept of duality of pattern, and applied it to the overall analysis of animal signals.²⁰ Specifically, he highlights the distinction between two levels of syntax:

1. The phonological syntax,²¹ which consists of the meaningless recombination of sounds into longer sequences. This syntactic level concerns the rules for the combinatorial structure of sounds;
2. The lexical syntax, whose rule of recombination concerns the generation of meaning within the sentence context.

For the purposes of this paper I will address the question whether there is any observable evidence that either of these steps, or at least some crucial aspects of them, are present in monkey communication systems, in order to find some evolutionary precursors of language. In order to avoid terminological confusion, it is worth emphasizing that with the term *syntax* I refer to the meaning modelled on the Greek word *syntaxis*, composed by “*syn*” (‘together’, ‘with’) – and “*taxis*” (‘order’, ‘connection’, ‘coordination of the parts according to structural rules’), which must be kept conceptually distinguished from the definition of the term *syntax* as, intrinsically tied to the semantic values of the lexical units occurring within the sentence context.

As to mere phonological syntax, we can find examples of sound sequences in animal vocal communication. Erroneously, indeed, it has historically been assumed that animal vocalizations are merely an acoustically graded continuum, in contrast to human utterances, which are perceived as differentiated into phonetic discrete units. By this regard, Cheney and Seyfarth assert:

Given the potential ambiguity inherent in a graded series of calls, and the importance of distinguishing both between different call types and between the call of different individuals, it appears that baboon listeners have been under strong selective pressure to detect subtle distinctions within a graded acoustic continuum and to link these differences in acoustic structure with differences in individual identities, social events, predators and so on.²²

In other words, monkeys are indeed able to categorize their communicative vocalizations into different acoustic features which convey different meanings, relying on contextual cues linked to the environment (presence of food or predators), to

²⁰ Marler 2000.

²¹ I use this term referring to the regulated combination of monkey vocalization sounds, which by themselves are not as differentiated as human languages phonemes.

²² Cheney and Seyfarth 2008, pp. 232–233.

the social relationship occurring between the vocalizing monkeys (in the case of vocal interactions), or to the emotional state of the caller. The inferred meaning of the vocalization relies either on the acoustic features of the signal, or on the information acquired on the basis of associations experienced in the past.

Concerning the second level of description adopted by Marler, the lexical syntax, recent studies suggest that the levels of syntactical complexity characterizing human verbal propositions are not widespread in animal communication systems.²³ Primate calls cannot be broken into meaningful units, and there are no parts comparable to words which can be combined in any rule-governed structure within a meaningful sentence, conveying a message which would be more than the sum of its parts. Nonetheless, recent field research has revealed the existence of a few important exceptions concerning rudimental cases of “vocal syntax” in non-human animals. Zuberbühler has observed that Campbell monkeys, a species living on the western Ivory coast, emit a pair of low “boom” calls before their alarm calls, in the presence of less dangerous situations such as a falling branch or upon hearing the predator alarm call of a distant group. As the author asserts, it seems that this acoustic component somehow affects the overall meaning of the call:

[“Boom” vocalization] is given in pairs separated by some seconds of silence and typically precedes an alarm call series by about 25s. Boom-introduced alarm call series are given to a number of disturbances, such as a falling tree or large breaking branch, the far-away alarm calls of a neighbouring group, or a distant predator. Common to these contexts is the lack of direct threat in each, unlike when callers are surprised by a close predator.²⁴

In this direction, a study conducted on the potty-nosed monkey reveals that this species uses two types of signals (pyows and hawks) and inverting them generates different meaning effects:

Series consisting of “pyows” are a common response to leopards, while “hacks” or “hacks” followed by “pyows” are regularly given to crowned eagles. Sometimes, males produce a further sequence, consisting of 1–4 “hacks”. These “pyows-hack” (P-H) sequences can occur alone, or they are inserted at or near the beginning of another call series. Regardless of the context, P-H sequences reliably predict forthcoming group progression. [...] *We conclude that, contrary to current theory, meaningful combinatorial signals have evolved in primate communication and future work may reveal further examples.*²⁵

Although these data confirm the ability, at least in some species of monkeys, to combine a few signals in a very rudimental way generating qualitatively different meanings, they lack the general capacity to apply combinatorial rules to produce an open-ended set of vocal productions, an ability that is typically human.

Importantly, evidence suggests that songbirds and whales also possess the ability of phonological syntax; in fact, a number of studies addressed have shown that these species are able to concatenate the notes of their songs following a hierarchical and non-random transitional structure.²⁶ Further, it has been shown that in chickadees,

²³ Cf. Collier et al. 2014 for a review.

²⁴ Zuberbühler 2002, p. 294. Cf. Ouattara et al. 2009.

²⁵ Arnold and Zuberbühler 2008, pp. 202–203; italics mine. – P.F.

²⁶ Okanoya 2004; Suzuki et al. 2006; Clarke et al. 2006; Jansen et al. 2012; Berwick et al. 2011.

experimental change to songs composition, rhythm, or component order tends to interfere with its communicative function.²⁷ Based on these data, we can identify in the ability to concatenate sounds within an utterance an “analogous” trait, i.e. a biological trait that has evolved independently in phylogenetically distant species, under the same selective forces. Importantly, studies suggest that this ability has evolved under the evolutionary pressures linked to sexual selection,²⁸ territory defense,²⁹ or group bonding.³⁰

Theory of Mind

A study concerning the evolutionary dynamics of language cannot disregard the research on the precursors of the capacity that had a key role in determining the specificity of human cognition: the ability to attribute mental states to conspecifics within a frame of shared intentions and joint actions.

It is worth asking, then, whether non-human animals are equipped with some equivalent ability. In order to address this question, it is necessary to distinguish the signaler’s perspective from the receiver’s one. As Seyfarth and Cheney assess, indeed, the formers are not aware of the state of knowledge of the receivers, neither do they communicate on the explicit goal to change it. Nonetheless, on the other hand, the achieved effect is to supply the listeners with useful information, or to cause an emotional and behavioural response:

[...] the co-evolution of caller and recipient has favored signalers who call strategically and listeners who acquire information from vocalizations, using this information to represent their environment. The inability of animals to recognize the mental states of others places important constraints on their communication and distinguishes animal communication most clearly from human language. With the possible exception of chimpanzees, animals cannot represent the mental state of another. As a result, whereas signalers may vocalize to change a listener’s behavior, they do not call with the specific goal of informing others or in response to the perception of ignorance in another. Similarly, whereas listeners extract subtle information from vocalizations, this does not include information about the signaler’s knowledge. Listeners acquire information from signalers who do not, in the human sense, intend to provide it.³¹

Interestingly, multiple studies suggest that a wide variety of species (phylogenetically both related and distant from humans) possess the ability to know what other individuals see.³² This might be considered an evolutionary precursor of the theory of mind. Importantly, although extensive research has been dedicated to

²⁷ Freeberg and Lucas 2002.

²⁸ Searcy and Andersson 1986.

²⁹ Holland et al. 2000.

³⁰ Boeckle and Bugnyar 2012; Doupe and Kuhl 1999; Treisman 1978.

³¹ Seyfarth and Cheney 2003, p. 168.

³² Bugnyar and Bernd 2006; Hare and Tomasello 1999; Hare et al. 2000 and 2003; Anderson et al. 1996; Ruiz et al. 2008.

animals' ability to infer others' states of mind, no common agreement on the interpretation of the resulting findings was achieved. In fact, much of the observed behaviours might be merely explained in terms of associative learning from previous experience. Thus, we can conclude that although the ability to attribute mental states to other individual (i.e. to understand the other's beliefs and desires in intentional terms and to use this knowledge to trigger specific behaviors) is uniquely human, certain evolutionary constraints underlying this ability are present also in non-human species.

Could We See a Holistic Protolanguage Through Monkeys' Communication System?

The data discussed above can be used as a window through which the evolution of language can be studied. According to the methodological criteria proposed by Botha, this approach satisfies the three conditions of groundedness, warrantedness, and pertinence. Indeed, an overall analysis of non-human animals' vocalization system has provided pivotal empirical data (although further investigations are still necessary). This allows us to recognise that the criterion of groundedness of the theory is satisfied. Moreover, the comparative approach I have adopted is empirically supported by the evolutionary data provided by studies on "homologs" – i.e. structurally similar traits that belong to phylogenetically close species and on "analogues" – i.e. functionally similar traits that phylogenetically distant species have acquired independently. Finally, the condition of pertinence is guaranteed by the identification of language with the ability to speak and understand a natural language, where meanings are: (1) syntactically structured, (2) acquired through social practises and (3) ontologically tied to the pragmatic and/or emotional situation in which they occur.

The data reviewed in the present study support the adoption of the holistic model proposed by Wray³³ as more plausible than the analytic one proposed by Bickerton. Indeed, even if the signalers are not able to communicate intentionally (that is, with a conscious, explicit aim to provide other individuals with specific information) – the listeners are nonetheless able to get from such unintentional utterances an arrangement of complex meanings, not reducible to mere lexical labeling. Regarding this last point, it is noteworthy to remark Cheney and Seyfarth's observations about primates' alarm calls: "Baboon alarm calls, like those of other primates, are thus holistic utterances, simultaneously both eventish and objectish because they incorporate both reference to an object and a disposition to behave toward that object in a particular way".³⁴

³³Wray 1998 and 2002.

³⁴Cheney and Seyfarth 2008, p. 256.

Conclusions

The assumption that the first human utterances were holistic is an important step in the study of the origin of language, and opens new questions to address. For instance, it would be interesting to study the specific dynamics concerning the evolution of the ability to know what other individuals see into the ability to infer what they know: a faculty that is closely related to the ability to share thoughts, attention targets, and goals. A second question concerns the pragmatic and cognitive process that, within an increasing complex frame of shared attention and actions, grounds the evolution of the holistic messages into syntactically structured sentences. These research questions might pave the way for an increasing understanding of the evolution of propositional language and to its links to animal communication.

Acknowledgments Funding was provided by a European Research Council Advanced Grant SOMACCA (No. 230604) awarded to WTF. The funders had no role in decision to publish, or preparation of the manuscript.

References

- Anderson, J. R., Montant, M., & Schmitt, D. (1996). Rhesus monkeys fail to use gaze direction as an experimenter-given cue in an object-choice task. *Behavioural Processes*, *37*(1), 47–55.
- Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology*, *18*(5), 202–203.
- Berwick, R. C., Okanoya, K., Beckers, G. J. L., & Bolhuis, J. J. (2011). Songs to syntax: The linguistics of birdsong. *Trends in Cognitive Sciences*, *15*(3), 113–121. Doi:[10.1016/j.tics.2011.01.002](https://doi.org/10.1016/j.tics.2011.01.002).
- Bickerton, D. (2002). Foraging versus social intelligence in the evolution of protolanguage. In A. Wray (Ed.), *The transition to language* (pp. 207–225). Oxford: Oxford University Press.
- Boeckle, M., & Bugnyar, T. (2012). Long-term memory for affiliates in ravens. *Current Biology*, *22*, 1–6. Doi:[10.1016/j.cub.2012.03.023](https://doi.org/10.1016/j.cub.2012.03.023).
- Botha, R. P. (2006). On the windows approach to language evolution. *Language and Communication*, *26*, 129–143.
- Bugnyar, T., & Bernd, H. (2006). Pilfering ravens, *Corvus corax*, adjust their behaviour to social context and identity of competitors. *Animal Cognition*, *9*(4), 369–376.
- Cheney, D. L., & Seyfarth, R. M. (2008). *Baboon metaphysics. The evolution of a social mind*. Chicago: The University of Chicago Press.
- Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, *3*(01), 1–15.
- Christiansen, M. H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in Cognitive Sciences*, *7*, 300–307.
- Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS ONE*, *1*(1), e73. Doi:[10.1371/journal.pone.0000073](https://doi.org/10.1371/journal.pone.0000073).
- Clay, Z., & Zuberbühler, K. (2009). Food-associated calling sequences in bonobos. *Animal Behaviour*, *77*(6), 1387–1396.
- Collier, K., Bickel, B., Schaik, C. P. P. van, Manser, M. B., & Townsend, S. W. (2014). Language evolution: Syntax before phonology? *Proceedings of the Royal Society B: Biological Sciences*, *281*(1788). Doi:[10.1098/rspb.2014.0263](https://doi.org/10.1098/rspb.2014.0263).

- Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain*. New York: WW Norton & Company.
- Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*, 22(1), 567–631.
- Fitch, W. T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.
- Freeberg, T. M., & Lucas, J. R. (2002). Receivers respond differently to chick-a-dee calls varying in note composition in Carolina chickadees, *Poecile carolinensis*. *Animal Behaviour*, 63(5), 837–845. doi:10.1006/anbe.2001.1981.
- Hare, B., & Tomasello, M. (1999). Domestic dogs (*Canis familiaris*) use human and conspecific social cues to locate hidden food. *Journal of Comparative Psychology*, 113(2), 173–177. doi:10.1037/0735-7036.113.2.173.
- Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, 59(4), 771–785.
- Hare, B., Addessi, E., Call, J., Tomasello, M., & Visalberghi, E. (2003). Do capuchin monkeys, *Cebus apella*, know what conspecifics do and do not see? *Animal Behaviour*, 65(1), 131–142.
- Hauser, M. D. (1996). *The evolution of communication*. Cambridge: MIT Press.
- Hauser, M. D., & Marler, P. (1993). Food-associated calls in rhesus macaques (*Macaca mulatta*). I. Socioecological factors. *Behavioral Ecology*, 4, 194–205.
- Holland, J., Dabelsteen, T., & Paris, A. L. (2000). Coding in the song of the wren: Importance of rhythmicity, syntax and element structure. *Animal Behaviour*, 60(4), 463–470.
- Jansen, D. A., Cant, M. A., & Manser, M. B. (2012). Segmental concatenation of individual signatures and context cues in banded mongoose (*Mungos mungo*) close calls. *BMC Biology*, 10(1), 97. doi:10.1186/1741-7007-10-97.
- Marler, P. (2000). Origins of music and speech: Insights from animals. In N. L. Walli, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 31–48). Cambridge: MIT Press.
- Okanoya, K. (2004). Song syntax in Bengalese finches: Proximate and ultimate analyses. *Advances in the Study of Behavior*, 34, 297–346.
- Jespersen, O. (1922 [2013]). *Language. Its nature, development and origin*. Oxford: Routledge.
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, 106, 22026–22031.
- Ruiz, A., Gómez, J. C., Roeder, J. J., & Byrne, R. W. (2008). Gaze following and gaze priming in lemurs. *Animal Cognition*, 12(3), 427–434. doi:10.1007/s10071-008-0202-z.
- Searcy, W. A., & Andersson, M. (1986). Sexual selection and the evolution of song. *Annual Review of Ecology and Systematics*, 17, 507–533.
- Seyfarth, R. M., Cheney, D. L., Bergman, T., Fischer, J., Zuberbühler, K., & Hammerschmidt, K. (2010). The central importance of information in studies of animal communication. *Animal Behaviour*, 80(1), 3–8. doi:10.1016/j.anbehav.2010.04.012.
- Seyfarth, R. M., & Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual Review of Psychology*, 54, 145–173.
- Struhsaker, T. T. (1967). Auditory communication among vervet monkeys (*Cercopithecus aethiops*). In S. A. Altmann (Ed.), *Social communication among primates* (pp. 281–324). Chicago: University of Chicago Press.
- Suzuki, R., Buck, J. R., & Tyack, P. L. (2006). Information entropy of humpback whale songs. *The Journal of the Acoustical Society of America*, 119(3), 1849. doi:10.1121/1.2161827.
- Treisman, M. (1978). Bird song dialects, repertoire size, and kin association. *Animal Behaviour*, 26, 814–817. doi:10.1016/0003-3472(78)90146-X.
- Wittgenstein, L. (1953 [2010]). *Philosophical investigations*. New York: Wiley.
- Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication*, 18, 47–67.
- Wray, A. (2002). Dual processing in protolanguage: Performance without competence. In A. Wray (Ed.), *The transition to language* (pp. 113–137). Oxford: Oxford University Press.

- Yip, M. J. (2006). The search for phonology in other species. *Trends in Cognitive Sciences*, 10(10), 442–446.
- Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour*, 63, 293–299.
- Zuberbühler, K., Noë, R., & Seyfarth, R. M. (1997). Diana monkey long-distance calls: Messages for conspecifics and predators. *Animal Behaviour*, 53(3), 589–604.
- Zuberbühler, K., Cheney, D. L., & Seyfarth, R. M. (1999). Conceptual semantics in nonhuman primates. *Journal of Comparative Psychology*, 113(1), 33–42.