

**Cite this article:** Levinson SC. 2022 The interaction engine: cuteness selection and the evolution of the interactional base for language. *Phil. Trans. R. Soc. B* **377**: 20210108. <https://doi.org/10.1098/rstb.2021.0108>

Received: 15 September 2021

Accepted: 27 January 2022

One contribution of 14 to a theme issue 'Revisiting the human 'interaction engine': comparative approaches to social action coordination'.

**Subject Areas:**

evolution, cognition, behaviour

**Keywords:**

interaction engine, language evolution, theory of mind, empathy, cuteness selection, alloparenting

**Author for correspondence:**


Stephen C. Levinson

e-mail: [stephen.levinson@mpi.nl](mailto:stephen.levinson@mpi.nl)

# The interaction engine: cuteness selection and the evolution of the interactional base for language

Stephen C. Levinson

Language and Cognition, Max Planck Institute for Psycholinguistics, Nijmegen, Gelderland, The Netherlands

 SCL, 0000-0001-8961-5316

The deep structural diversity of languages suggests that our language capacities are not based on any single template but rather on an underlying ability and motivation for infants to acquire a culturally transmitted system. The hypothesis is that this ability has an interactional base that has discernable precursors in other primates. In this paper, I explore a specific evolutionary route for the most puzzling aspect of this interactional base in humans, namely the development of an empathetic intentional stance. The route involves a generalization of mother–infant interaction patterns to all adults via a process (cuteness selection) analogous to, but distinct from, RA Fisher's runaway sexual selection. This provides a cornerstone for the carrying capacity for language.

This article is part of the theme issue 'Revisiting the human 'interaction engine': comparative approaches to social action coordination'.

## 1. On the puzzle of language evolution

As was much remarked upon in Darwin's time, there is an apparently unbridgeable gulf between human language and the communication systems of other species. Darwin [1] sought to overcome this by invoking distinct phases and mechanisms in the evolution of language (see [2]). The topic of language evolution has always been a controversial area and remains so [3,4], but new palaeontological data and the recovery of ancient DNA have considerably narrowed the scope of the controversy. It is clear that there are many biological adaptations for language including specialist enervation of the respiratory and vocal tract, tuning of hearing to the bandwidth of speech, the auditory-vocalization pathways for vocal imitation, neural adaptations like slight extension of the arcuate fasciculus [5], and genetic adaptations like variants of FOXP2 which facilitate language production through developmental pathways [6]. Yet it is also clear that the great bulk of language complexity, by contrast, must be attributed not to an innate source but to cultural evolution, for languages differ in their construction on every level, from the sounds to the syntax. Indeed, apart from the complexity of human communication systems, the thing that sets language apart is its sheer diversity across human groups—no other animal (as far as we know) has a communication system which can differ so fundamentally in both sound, form and meaning [7]. As Darwin [1, p. 56] remarked, 'language is an art made possible by an instinct to learn'.

This diversity raises a puzzle about how the physiological adaptations could have evolved to support such a moving target. One answer to this puzzle is to appreciate the central contribution that human interactional capacities make to language, since in contrast to the diversity of languages these interactional characteristics turn out to be nearly uniform across the species. For example, there are precise parallels across languages and cultures in the temporal properties of turn-taking in conversation [8], in the mechanisms that maintain

intersubjective understanding through ‘repair’ of unclear utterances [9], and in the organization of sequences of contingent responses [10]. There therefore appears to be a much greater constancy in the niche for language usage and language learning than in the organization of languages themselves. It is this interactional niche and the properties characterizing language use which has probably played an instrumental role as a crucial constant target for the biological adaptations underpinning language: human interactional abilities offer a ‘machine tool’ as it were for producing languages.

Darwin’s adversary Max Müller [11, p. 403] famously held that ‘language is our Rubicon, and no brute will dare to cross it’. However, when attention is turned to the underlying interactional abilities that make language possible, continuities between human and other primate communication systems become much more apparent. Human languages have evolved in a quite specific communicational niche—a highly intensive form of social interaction with special properties—the roots of which however can be seen in other primates (e.g. [12,13]).

## 2. The ‘interaction engine’

On this account then, human languages have evolved and diversified through cultural evolution, being learned by each successive generation. This makes the context for acquisition by children crucial. The prime ecological niche for language in general and language acquisition in particular is social interaction—this is the context in which languages are learnt by children, using native interactional capacities which allow them to bootstrap themselves into their local tongues. Some time ago, I outlined the case [14] for a specific type of human social interaction that is not itself linguistic, but which makes language possible. The capacities that enable this form of communicative interaction I dubbed the ‘interaction engine’, including crucially the attribution of intention to communicative acts involving a modelling of the other’s mentation, and even a modelling of the other’s model of ego’s intentions. This last is what makes it possible to invent a signal, as when I rub my chin to confidentially indicate you have egg on yours (my signal is based on a Gricean intention to get you to wonder what I must be trying to get you to think). In addition, I point to a bunch of more easily observable features that characterize human communicative interaction, including turn-taking, contingent sequences and multimodal signals.


In the succeeding years, a great deal of research has borne out those original suggestions. Large spoken corpora in numerous languages have established striking uniformities in conversational behaviour [8–10], starkly contrasting with further evidence for structural diversity across languages [15]. Experimental and neurocognitive work reveals the extraordinary work-arounds that make linguistic processing possible in the tight timing of conversational exchanges [16,17], while work on infant–caretaker interaction shows that ‘proto-conversation’ is early established in the first year of life [18–20], and is evidenced in the most remote societies [21]. Work on autism spectrum disorders provides evidence that many aspects of the interaction engine are systematically disrupted together [22]. One of the most exciting developments has been the exploration of many of these properties across the primate order, suggesting precursor abilities in many other species, e.g. communicative turn-taking [23–25].

Here we focus on some principal components of the ‘interaction engine’ ([26]; for a more comprehensive list of design features see [27]), namely: (i) orchestration of multimodal signals (gesture, vocalization, gaze, facial displays), (ii) highly controlled timing (e.g. tight response latencies), (iii) manipulation of specific contingencies between initiating signal and response, and (iv) aspects of what has been called ‘theory of mind’, specifically the ability to attribute detailed intentions to communicative acts.

Taking these four components in turn, we here sketch the human specificities of each. First, then, multimodality: given the pre-eminence of speech it is easy to overlook the fact that language is primarily produced within a multimodal display, with gaze to addressee, facial displays, and near obligate gesture of the hands [28]. There are very interesting questions, but few answers, about how these behavioural streams are coordinated and understood—parts of the behavioural stream that are understood to belong to one another are often not synchronous [29]. In general, the multimodal nature of animal communication is also often underplayed [30]. As Darwin [31] noted in the context of expressions of emotion, there are clear phylogenetic continuities here with other primates, but also (as more recently observed) some human specialisms like the white sclera that makes human gaze so trackable [32].

The second component is timing. Human communication is typically done in alternation, taking turns at short bursts of communication, averaging about 2 s long, with very rapid transitions of the order of 200 ms between turns [17]. These timings are very similar across languages and cultures [8]. The response timing remains more or less the same in sign languages [33], which is interesting because it shows that the alternation of turns is not motivated by avoiding masking of the auditory signal (signs are produced by the motor system but perceived by the visual system with no interference). There is great sensitivity in timing, as slight delays are interpreted as meaningful [34]. Again, phylogenetic precursors are clearly in evidence, as vocal turn-taking systems can be found scattered right across the primate order, usually with slower turn-transitions around 1 s long ([23]; fast turn-taking is especially prominent in the Callitrichidae, notable for their alloparenting, relevant below). The great apes seem to be the exceptions as vocal turn-taking is not in much evidence, but their more flexible gestural communicative systems are remarkable for showing timing characteristics (e.g. 200 ms response intervals) very close to the human norm [24]. Given the great ape evidence, a popular theory of language evolution suggests an early gestural phase, and this is consistent with palaeontological evidence which suggests that the breathing adaptations for speech post-date early forms of African *Homo erectus* [35].

The third component is the set of contingencies between a turn and its response. In language usage, this is the relationship between question and answer, request and compliance, greeting and reciprocal greeting and so on. In humans, these are highly elaborated in two directions: first, there is a huge inventory of such ‘speech acts’ and the potential to always invent more, and second, there is the indefinite recursive elaboration of these sequences. For example, after a question another question–answer pair may intervene before the answer is supplied, as in (from [36, p. 333]):

A: <i>May I have a bottle of Mich?</i>	<-Base-1st-Part	
B: <i>Are you twenty-one?</i>	<-Insert-1st-Part	
A: <i>No</i>	<-Insert- 2nd-Part	
B: <i>No</i>	<-Base-2nd-Part	

Indeed, since far deeper centre-embedding is found in these conversational sequences than in natural language grammar, this is a possible origin for the syntax of non-local dependencies [37]. In any animal communication system, contingencies between signal and response are the indication that some kind of communication has effectively taken place, and detailed contingencies have been observed both in monkey call exchanges (see [38]) and in ape gesture systems, e.g. with request-compliance exchanges between chimpanzee or orangutan infants and mothers [39].

The final component discussed here is the ability to accurately attribute intentions to communicative turns, involving at least some aspects of what has been called ‘theory of mind’. For the contingencies just outlined hold, not mechanically between the form of the utterances in human communication, but between their underlying intents, actions or purposes. For example, ‘It’s six o’clock’ could serve variously as an excuse to leave, a suggestion to go to the pub, the declaring of a meeting open, and so forth, according to sequence and context. It is this feature that above all gives human communication its flexibility. It presupposes basic theory of mind—that actions are explained by goal-driven purposeful behaviour. However, more than that, it requires grasping the idea that an action has been done specifically to have the intention recognized, for that is its sole or main purpose. I can push the wine bottle towards you, suggesting you have some more, or I can make as to fasten a button suggesting you do up yours. Actions that might look like practical actions can be done in such a way as to indicate they are actually communicative, as when I rub my chin vigorously to indicate you still have breakfast on yours. The successful attribution of communicative intentions is dependent on assumptions of veridical (or at least helpful) communication, cooperation and trust.

These four components (multimodality, timing, contingency, intentionality) are the critical design features that make human communicative interaction possible. Studies comparing their properties across a dozen or more human language and cultures, mostly unrelated, indicate that these features are universal, in the sense that they are relatively uniform across unrelated cultures.

Comparative ethology across humans and primates suggests that these components were added cumulatively over deep time, and before vocal language as we know it. Fully vocal language probably developed somewhere after early forms of *Homo erectus* (who seems to have lacked the voluntary breath control essential to fine speech production) but before the last common ancestor of modern humans and Neanderthals, so between say 1.4 Ma and 700 Ka [40]. However, early tool using (and certainly Mode II or Acheulian technologies) suggests some advanced form of communication afforded cultural and technological transmission far earlier [41]. If so, it must have been at least partly gestural given the lack of respiratory control in African *Homo erectus* [35], and was plausibly built on the antecedent ape gesture system, so inheriting the turn-taking timing shared between humans and chimpanzees. That language usage has developed within an antecedent turn timing straightjacket is suggested by the severe cognitive processing challenges that now have to be overcome both in child language acquisition and adult production, where complex syntax is squeezed rapidly into short turns [17]. Given such a gradual accumulation, it is legitimate to ask whether these four components (multimodality, timing, contingency, intentionality) form the

core of an integrated system. Here evidence from autism spectrum disorder (ADS) suggests they do, since ADS often exhibits simultaneous impairments of gaze, turn-taking, timing [42] and gesture [43], sequencing of contingent acts and ascription of intentions [44].

As pointed out, each of these interactional components have clear precursors in the communicative behaviour of other primate species [23,45], but the highly developed intentionality of human communication still seems quite exceptional. Much discussion of language treats it on a coding analogy: speakers code their thoughts in the conventions of their language, and hearers knowing the conventions decode the intended thoughts. This is how computers communicate with one another, but not how human language works. Human language capacities rest squarely on the ability to infer communicative intentions. Consider the following exchange where no compliment is actually encoded:

A: *I could eat the whole of that cake.*

B: *Thanks. It’s quite easy to make actually.*

Intention attribution also serves to unlock novel expressions like ‘nano bonus’ or ‘muscular speech’. Linguists point out that nearly every ordinary utterance relies heavily on disambiguation and pragmatic resolution based on inferences about speaker’s meanings or intents (‘that is a good point’ could be appraisal of the knife in your hand, or appreciation of the argument you just made). Even the simplest indexical pointing of the hand requires resolution (is what is indicated an object, its colour, the desire for it, a direction, or what?). All these inferences to the best interpretation rely on the assumption of cooperative communication, together with rules of thumb that follow from that assumption [46]. The language learning situation relies on the child grasping the intentionality behind utterances—even the learning of a name or referring expression requires inferring whether the indicated thing is a type or token, a quality or an entity. The point then is this: it is not language that has made human social interaction possible, it is rather our interactional abilities that have enabled language, and a crucial ingredient is this rich intentionality, this reflexive mind-reading. The question then arises what exactly is the evolutionary background behind this crucial ingredient in human communicative interaction. This question is the focus of rest of this paper.

### 3. The origins of intention recognition in communication: ‘cuteness selection’ and alloparenting

In exploring how the intention-recognition capacity of the ‘interaction engine’ may have evolved, one naturally looks for precursors among the other primates, specifically one looks for behavioural niches where empathy and ‘theory of mind’ might be exhibited. Although apes clearly grieve for their dead, there is no evidence that non-human primates show empathy in cases other than distress [47], and apes at least do not seem motivated by sympathy to perform pro-social acts [48].

An important and unresolved question is whether the non-human great apes exhibit any signs of complex reflexive intentionality. First-order intentionality (attempts to get the

other to do something) does not seem in doubt, but second-order intentionality (attempts to get the other to think or desire something) and beyond remains controversial, despite evidence for persistence, repair, upgrading and audience-design [30]. Critical will be evidence about how the gesture signals, the characteristic flexible communication systems of the great apes, actually get acquired. Some theories hold there is an innate repertoire [49], others that many gestures are interactionally learnt ([50]; see also [51,52]). The latter would open up the possibility of third-order intentionality (the attempt to get the other to think such-and-such by virtue of the recognition of that attempt, i.e. full Gricean 'meaning'), a possibility that Dennett [53] has argued should at least be entertained to see if it yields interesting predictions (and see [54]).

However, one domain where most mammals and not just the great apes tend to display interest in, and caring for the mental and physical states of others, is the maternal relationship between mothers and offspring. This is the locus for a restricted and specialized 'theory of mind', as required for intermittent on-demand feeding and other kinds of infant care. It is notable that this maternal relationship is the environment in which a flexible gesture system first gets established in the development of chimpanzees [24,50], just as it is the key setting for the first stages of language acquisition (in both cases of course other social partners play a key role in the full development of the communicative repertoire; [50]).

Now, humans are clearly an alloparenting species, that is, they engage carers other than the mother alone in the child-rearing process. Alloparenting frees the mother for further reproduction, and in fact humans in traditional societies have interbirth intervals close to half those of the other apes (around 29 months in humans versus around 60 months in chimpanzees), and early adoption of shared parenting must have been a key factor in human demographic success [55,56]. We are in effect the rabbits among the apes. By contrast, alloparenting is largely precluded in chimpanzees by the risk of infanticide [47]—chimpanzee mothers jealously guard their infants, holding onto them even when they have died. Consequently, chimpanzee twins rarely survive outside captivity [57].

Alloparenting requires a generalization of the maternal relationship, with the mother's interest in the infant's needs emulated by other adults or older offspring [56]. It also incidentally requires community-wide generalization of communicative signals—ritualized signals developed inside the maternal relationship will not be enough. The evolutionary mechanisms for such a generalization of aspects of theory of mind may have been various, but one possibility is a mechanism that has not been much discussed, namely what can be called 'cuteness selection' [58]. 'Cuteness selection' is the evolutionary mechanism coupling appealing features and caring responses. Konrad Lorenz first identified the human preference for the large eyes, globular skull, reduced jaw, chubbiness, high-pitched vocalizations and so forth both in offspring and our pets ([59]; see also [60,61]). The deep almost irresistible appeal of cute features to humans is commercially exploited in the pet trade, the comic industry whether Disney or manga [62], and in artistic kitsch [63].

This match between stimulus (cuteness) and response (caring) may have been originally driven by the same mechanisms identified by RA Fisher under the rubric of runaway sexual selection [64]—applied not only to mate selection but

also to infant survival success. Fisher's target was the handicap feature like the peacock's tail which caused Darwin such sleepless nights: what started off as an honest signal of fitness, once locked into an automated stimulus–response cycle, can accumulate to a degree where it is actually detrimental. Such an originally virtuous cycle may lie behind the irresistible attraction of 'cuteness' features: if parents invest more in 'cuter' offspring, their own preferences may be passed on as well as the stimulus that triggers the preference; similarly, if mates by this mechanism come to prefer 'cuter' mates, the process can accelerate. At the same time, alloparenting reduces the mother's investment, so that human mothers are more likely to abandon infants than chimpanzee mothers [56]. Thus 'cuteness' becomes an important counterweight to possible abandonment in times of stress, so reinforcing the fitness benefits of 'cute' features. Importantly, along with the external 'cuteness' stimuli go the caring response instincts, the empathy and the interest in emotional states it triggers. Alloparenting would greatly increase the fitness benefits of both the 'cuteness' stimulus and the caring response, and motivate the generalization of maternal instincts outside the maternal relationship. In this way, 'cuteness selection' offers a feedback process, similar to Fisher's runaway sexual selection, that could have generated the overall generalization of empathy and 'theory of mind' that characterizes our species.

If attending to an infant's needs and desires involves first-order theory of mind (a model of the other), trying to reassure, console, or express caring and affection can involve third-order theory of mind, namely a model of what the other will make of those signals. It is this third-order that is essential to the flexible use of language, the narrowing of reference and the use of novel expressions—i.e. the level of Gricean intentions. When I rub my chin vigorously intending you to think you may have egg on yours, you decode it by recognizing a non-instrumental action and wondering what led me to produce it. Since ape gestures (like human utterances) seem sometimes novel [52], and typically vague or ambiguous, they have to be resolved in context—suggesting that these systems already involve higher order intentionality [65]. It is this possibility for constructing signals that allows a language to grow, and indeed to function by pragmatically fleshing out the meaning. While higher order intentionality does not alone account for language, it is a precondition, and with the incremental growth of a conventional system of symbols higher level intentionality is itself reinforced [66]. Being able to express thoughts about others' states of mind will certainly enrich theories of mind, as studies of emerging sign languages have shown [67]. This 'maternalization' of the general population through 'cuteness selection' may not only have sown the seeds for language but may also help to explain the default trust and cooperation that is found inside (but not necessarily outside) human and other alloparenting groups [68]. In addition, 'cuteness selection' may offer a rival theory to the self-domestication theory recently invoked to explain cranial globularization and associated changes in the human lineage [69]. Researchers have pointed to many features of human evolution (including gene expression) that appear neotenic (paedomorphic), or more exactly heterochronic in origin, including extended childhood [70,71]. Traditionally, the cranial aspects of these processes have been attributed to reduction of the dentition consequent on dietary changes and to fire. Recently, though, it has been

suggested that the different cranial shapes of modern humans and Neanderthals can be attributed to ‘self domestication’, wherein a reduction of aggression alone carried with it a host of other features, as has been noted in domesticated animals [72]. There are problems with this theory though (see also [73]). Firstly, domestication is typically associated with drastic reduction of the brain (in pigs up to 35%; [74]), and although early modern humans show a small reduction from Neanderthal specimens, the reduction is in proportion to differences in body mass; generally, and over the much longer period, human evolution is of course associated with massive brain increases. Secondly, domestication is associated with an all-round loss of aggression (for that is the major target of domestication), but humans are one of the most aggressive of all mammalian species, it is simply that the aggression is typically outgroup directed, and is mostly premeditated and collective [75]. ‘Cuteness selection’ offers an account of the trend to human ‘neoteny’, anatomical gracilization and extended childhood that does not have these difficulties and would have worked on the much longer timescale of hominin evolution over the last couple of million years. That time scale matches recent endocranial studies showing growth in language-related brain areas separating earlier *Homo erectus* specimens pre-1.5 Ma from later ones [76], which is in line with the likely (but still controversial) possession of full language capacities by the common ancestor of Neanderthals and modern humans by ca. 700 Ka [40]. By contrast, the self-domestication thesis would link language to post-Neanderthal brain reduction, while the palaeontological and genetic evidence so far points to a pre-Neanderthal date for extensive spoken language [3,40].

One of the proximal effects of ‘cuteness’ is that it releases oxytocin in the recipients, a hormone that plays a role in many physiological processes, being upregulated in pregnancy and lactation. Its role in human and animal bonding has been much studied: it reduces fear, enhances trust and promotes prosociality [77]. It increases empathy and encourages gaze at interactants’ faces [78], and reduced levels are associated with autism and other interactional deficits [79]. Different levels of the hormone are associated with the greater sociality (empathy, tolerance, play) of bonobos compared to chimpanzees [80,81], and it and other endogenous opioids are upregulated in humans compared to apes [82]. Especially relevant for the current argument is that administration of oxytocin enhances communication and the effective targeting of messages [83]. It thus seems plausible that runaway cuteness selection could have driven the generalized ‘theory of mind’ essential to the way modern human language works.

The account outlined here and in [58] emerging from linguistics and anthropology converges with that given by Hrdy & Burkart [84] based on data from primatology. Both accounts see causal connections between alloparenting and theory of mind, but the Hrdy & Burkart account emphasizes the infant’s role in beguiling and manipulating reluctant caretakers, while the present theory puts the emphasis on adaptations in empathy and caring in the alloparents, triggered by cuteness features. Neither theory explains why other primate alloparenting species, largely among the callitrichids, show little evidence of theory of mind; both theories must assume that the large-brained human ancestors were fertile ground, or that (as in the present theory) a runaway process of cuteness selection happened not to be impeded by other physiological requirements like sexual dimorphism or a diet requiring massive dentition.

## 4. Human interaction and human social systems

Human social life is held together by intense social interaction. It has been estimated that humans spend up to 30% of waking hours in such interaction, compared to perhaps 20% among chimpanzees [85], although in both cases there are large differences across particular social groups. Fire must have greatly extended the social day and enhanced interactional opportunities [86]. A cross-cultural study of conversation suggests that individuals spend on average 4.5 h conversing, in which time they each may produce 16 000 words in 1500 turns at talking (extrapolating from [87]).

What drives the extraordinary human investment in interaction? Alloparenting, the outsourcing of infant care, depends on building bonds within social groups, and such bonds would also have been crucial in the earliest phases of hominin cooperative scavenging and hunting [88,89]. The multiple childhood attachments offered by alloparenting would have facilitated the acquisition of skills and information crucial to building cultural adaptations. There were probably multiple incentives of this sort for investing in interactional partners. In addition, Dunbar has suggested that language, because of its broadcast character, arose as a form of mass grooming, so knitting together large groups. However, this neglects both the fact that chimpanzees often groom in large chains [90], and that human conversational huddles are mostly in the twos or threes. However, pointing to the primary social functions of language is surely correct. Studies suggest that around two-thirds of what is discussed in informal conversation concerns social relationships [86,91]. The central role that social navigation plays in human social life is reflected in the way languages have elaborated conventions for signalling social relationships, for example, the elaborate honorifics of Japanese or Javanese, or in English the many indirect request forms (*Can I... , Could I possibly... , Would it by any chance be possible...etc.*) which achieve the same sort of effect. One of the central functions of language is thus the delicate juggling of highly sensitive social relationships. We have inherited from the common ancestor with our nearest primate relatives a fission–fusion society where relationships have to be constantly negotiated. In the pre-agrarian bulk of human history, this fluid social structure allowed conflict to be resolved simply by decamping, or else escalated by joining forces with kinsmen. In addition, humans in greatly elaborating divisions of labour and domains of expertise have engineered a system where social relationships are situationally dependent—I may be the nautical expert, but you the expert curer, orator, hunter or fisher: who leads depends on the activity. Conversation analysts have shown that these domains are actually jealously guarded and defended by the details of language use [92]. Even in the midst of the most rigid human social systems (rural caste India for example), social relationships are under constant juggling. The flexibility of human communication systems owe much to the fluidity of human social relationships and the social structures built out of them.

## 5. Conclusion

This paper has argued that human language rests on an infrastructure for communicative interaction. It also of course relies on many anatomical and neurocognitive adaptations for language, but the evolutionary launch pad for those features depends on the prior acquisition of the interactional

infrastructure that would have made language both possible and adaptive. Precursors for many aspects of this interactional infrastructure (turn-taking timing, contingent signals, multi-modality) can be seen in other primates, but despite precursors, the evolution of full-blown intention recognition and theory of mind seems more puzzling. Here we have argued that one possible account for how humans developed the extended ‘mind-reading’ that makes language feasible is based on the generalization of mother–infant sensitivities,

spurred by alloparenting and the possible runaway characteristics of ‘cuteness selection’.

**Data accessibility.** This article has no additional data.

**Authors’ contributions.** S.C.L.: conceptualization, writing—original draft, writing—review and editing.

**Conflict of interest declaration.** I have no competing interests.

**Funding.** I received no funding for this study.

## References

- Darwin C. 1871 *The descent of man*. London, UK: John Murray.
- Fitch WT. 2019. Darwin’s model of language evolution: an essay. See <https://homepage.univie.ac.at/tecumseh.fitch/darwins-model-of-language-evolution-an-essay.html>.
- Dediu D, Levinson SC. 2013 On the antiquity of language: the reinterpretation of neandertal linguistic capacities and its consequences. *Front. Lang. Sci.* **4**, 397.
- Berwick R, Chomsky N. 2016 *Why only US?* Cambridge, MA: MIT Press.
- Rilling JK, Glasser MF, Preuss TM, Ma X, Zhao T, Hu X, Behrens TE. 2008 The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* **11**, 426–428. (doi:10.1038/nn2072)
- Fisher S. 2019 Human genetics: the evolving story of FOXP2. *Curr. Biol.* **29**, R65–R67. (doi:10.1016/j.cub.2018.11.047)
- Evans N, Levinson SC. 2009 The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* **32**, 429–492. (doi:10.1017/S0140525X0999094X)
- Stivers T *et al.* 2009 Universals and cultural variation in turn-taking in conversation. *Proc. Natl Acad. Sci. USA* **106**, 10 587–10 592. (doi:10.1073/pnas.0903616106)
- Dingemanse M *et al.* 2015 Universal principles in the repair of communication problems. *PLoS ONE* **10**, e0136100. (doi:10.1371/journal.pone.0136100)
- Kendrick K *et al.* 2020 Sequence organization: a universal infrastructure for social action. *J. Pragmat.* **168**, 119–138. (doi:10.1016/j.pragma.2020.06.009)
- Müller M. 1861 *Lectures on the science of language*, vol. 1. London, UK: Longman.
- Burkart J, Adrieane J, Brügger R, Miss F, Wierucka K, van Schaik C. 2022 A convergent interaction engine: vocal communication among marmoset monkeys. *Phil. Trans. R. Soc. B* **377**, 20210098. (doi:10.1098/rstb.2021.0098)
- Fröhlich M, van Schaik CP. 2022 Social tolerance and interactional opportunities: environmental drivers of persistence and elaboration in orang-utan gestural communication. *Phil. Trans. R. Soc. B* **377**, 20210106. (doi:10.1098/rstb.2021.0106)
- Levinson SC. 2006 On the human ‘interaction engine’. In *Roots of human sociality: culture, cognition and interaction* (eds NJ Enfield, SC Levinson), pp. 39–69. Oxford, UK: Berg.
- Skirgard, H. *et al.* Submitted. Grambank data reveal global patterns in the structural diversity of the world’s languages.
- Bögels S, Levinson SC. 2017 The brain behind the response: Insights into turn-taking in conversation from neuroimaging. *Res. Lang. Soc. Interact.* **50**, 71–89. (doi:10.1080/08351813.2017.1262118)
- Levinson SC, Torreira F. 2015 Timing in turn-taking and its implications for processing models of language. *Front. Psychol.* **6**, 731. (doi:10.3389/fpsyg.2015.00731)
- Hilbrink E, Gattis M, Levinson SC. 2015 Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Front. Psychol.* **6**, 1492. (doi:10.3389/fpsyg.2015.01492)
- Casillas M, Frank MC. 2017 The development of children’s ability to track and predict turn structure in conversation. *J. Mem. Lang.* **92**, 234–253. (doi:10.1016/j.jml.2016.06.013)
- Salter G, Carpenter M. 2022 The developmental origins of infants’ earliest communicative gestures: from incipient to canonical forms. *Phil. Trans. R. Soc. B* **377**, 20210102. (doi:10.1098/rstb.2021.0102)
- Casillas M, Brown P, Levinson SC. 2021 Early language experience in a Papuan community. *J. Child Lang.* **48**, 792–814. (doi:10.1017/S0305000920000549)
- Lord C, Charman T, Cusack J, Guillaume D, Frazier T, Jones RM, Pickles A, Lounds TJ, Jeremy VV. 2020 Autism spectrum disorder. *Nat. Rev. Dis. Primers* **6**, 5. (doi:10.1038/s41572-019-0138-4)
- Levinson SC. 2016 Turn-taking in human communication, origins, and implications for language processing. *Trends Cogn. Sci.* **20**, 6–14. (doi:10.1016/j.tics.2015.10.010)
- Fröhlich M, Kuchenbuch P, Müller G, Fruth B, Furuichi T, Wittig R, Pika S. 2016 Unpeeling the layers of language: bonobos and chimpanzees engage in cooperative turn-taking sequences. *Sci. Rep.* **6**, 25887. (doi:10.1038/srep25887)
- Rossano F. 2019 The structure and timing of human vs. primate social interaction. In *Human language: from genes and brain to behavior* (ed. P Hagoort), pp. 201–220. Cambridge, MA: MIT Press.
- Levinson SC. 2019 Interactional foundations of language: the interaction engine hypothesis. In *Human language: from genes and brain to behavior* (ed. P Hagoort), pp. 189–200. Cambridge, MA: MIT Press.
- Levinson SC. 2019 Natural forms of purposeful interaction among humans: What makes interaction effective? In *Interactive task learning: humans, robots, and agents acquiring new tasks through natural interactions* (eds KA Gluck, JE Laird), pp. 111–126. Cambridge, MA: MIT Press.
- Holler J. 2022 Visual bodily signals as core devices for coordinating minds in interaction. *Phil. Trans. R. Soc. B* **377**, 20210094. (doi:10.1098/rstb.2021.0094)
- Holler J, Levinson SC. 2019 Multimodal language processing in human communication. *Trends Cogn. Sci.* **23**, 639–652. (doi:10.1016/j.tics.2019.05.006)
- Rodrigues ED, Fröhlich M. 2021 Operationalizing intentionality in primate communication: social and ecological considerations. *Int. J. Primatol.* (doi:10.1007/s10764-021-00248-w)
- Darwin C. 1872 *The expression of the emotions in man and animals*. London, UK: John Murray.
- Kobayashi H, Kohshima S. 2001 Unique morphology of the human eye and its adaptive meaning. *J. Hum. Evol.* **40**, 419–435. (doi:10.1006/jhev.2001.0468)
- De Vos C, Torreira F, Levinson SC. 2015 Turn-timing in signed conversations: coordinating stroke-to-stroke turn boundaries. *Front. Psychol.* **6**, 268. (doi:10.3389/fpsyg.2015.00268)
- Bögels S, Kendrick KH, Levinson SC. 2020 Conversational expectations get revised as response latencies unfold. *Lang. Cogn. Neurosci.* **35**, 766–779. (doi:10.1080/23273798.2019.1590609)
- MacLarnon A, Hewitt G. 2004 Increased breathing control: another factor in the evolution of human language. *Evol. Anthropol.* **13**, 181–197. (doi:10.1002/evan.20032)
- Merritt M. 1976 On questions following questions (on service encounters). *Lang. Soc.* **5**, 315–357. (doi:10.1017/S0047404500007168)
- Levinson SC. 2013 Recursion in pragmatics. *Language* **89**, 149–162. (doi:10.1353/lan.2013.0005)
- Pika S, Wilkinson R, Kendrick K, Vernes S. 2018 Taking turns: bridging the gap between human and animal communication. *Proc. R. Soc. B* **285**, 20180598. (doi:10.1098/rspb.2018.0598)
- Rossano F, Liebal K. 2014 ‘Requests’ and ‘offers’ in orangutans and human infants. In *Requesting in social interaction* (eds P Drew, E Couper-Kuhlen), pp. 335–364. Amsterdam, The Netherlands: John Benjamins Publishing Co.
- Dediu D, Levinson SC. 2018 Neandertal language revisited: not only us. *Curr. Opin. Behav. Sci.* **21**, 49–55. (doi:10.1016/j.cobeha.2018.01.001)

41. Levinson SC, Holler J. 2014 The origin of human multi-modal communication. *Phil. Trans. R. Soc. B* **369**, 2013030.
42. Thorup E, Nyström P, Gredebäck G, Bölte S, Falck-Ytter T. 2018 Reduced alternating gaze during social interaction in infancy is associated with elevated symptoms of autism in toddlerhood. *J. Abnorm. Child Psychol.* **46**, 1547–1561. (doi:10.1007/s10802-017-0388-0)
43. Mishra A, Ceballos V, Himmelwright K, McCabe S, Scott L. 2021 Gesture production in toddlers with autism spectrum disorder. *J. Autism Dev. Disord.* **51**, 1658–1667. (doi:10.1007/s10803-020-04647-5)
44. Bottema-Beutel K. 2017 Glimpses into the blind-spot: social interaction and autism. *J. Commun. Disord.* **58**, 24–34. (doi:10.1016/j.jcomdis.2017.06.008)
45. Rossano. 2013 Sequence organization and timing of bonobo mother-infant interactions. *Interact. Stud.* **14**, 160–189. (doi:10.1075/is.14.2.02ros)
46. Levinson S. 2000 *Presumptive meanings*. Cambridge, MA: MIT Press.
47. Myowa M, Butler D. 2017 The evolution of primate attachment. In *The cultural nature of attachment* (eds H Keller, K Bard), pp. 53–67. Cambridge, MA: MIT Press.
48. Liebal K, Vaish A, Haun D, Tomasello M. 2014 Does sympathy motivate prosocial behaviour in great apes? *PLoS ONE* **9**, e84299. (doi:10.1371/journal.pone.0084299)
49. Byrne RW, Cartmill E, Genty E, Graham KE, Hobaier C, Tanner J. 2017 Great ape gestures: Intentional communication with a rich set of innate signals. *Anim. Cogn.* **20**, 755–769. (doi:10.1007/s10071-017-1096-4)
50. Pika S, Fröhlich M. 2019 Gestural acquisition in great apes: the Social Negotiation Hypothesis. *Anim. Cogn.* **1**, 1–15. (10.1007/s10071-017-1159-6)
51. Liebal K, Schneider C, Ersson-Lembeck M. 2019 How primates acquire their gestures: evaluating current theories and evidence. *Anim. Cogn.* **22**, 473–486. (doi:10.1007/s10071-018-1187-x)
52. Fröhlich M, van Schaik CP. 2020 Must all signals be evolved? A proposal for a new classification of communicative acts. *Wiley Interdiscip. Rev.: Cogn. Sci.* **11**, e1527. (doi:10.1002/wcs.1527)
53. Dennett D. 1983 Intentional systems in cognitive ethology: the 'Panglossian paradigm' defended. *Behav. Brain Sci.* **6**, 343–390.
54. Bohn M, Liebal K, Ona L, Tessler M. 2022 Great ape communication as contextual inference: a cognitive modeling perspective. *Phil. Trans. R. Soc. B* **377**, 20210096. (doi:10.1098/rstb.2021.0096)
55. Hrdy S. 1999 *Mother nature*. New York, NY: Pantheon Books.
56. Hrdy S. 2009 *Mothers and others*. Cambridge, MA: Harvard University Press.
57. Kishimoto T, Ando J, Tataru S, Yamada N, Konishi K, Kimura N, Fukumori A, Tomonaga M. 2014 Alloparenting for chimpanzee twins. *Sci. Rep.* **4**, 6306. (doi:10.1038/srep06306)
58. Levinson SC. 2018 Spatial cognition, empathy and language evolution. *Stud. Pragm.* **20**, 16–21.
59. Lorenz K. 1943 Die angeborenen Formen moeglicher Erfahrung. *Z. fuer Tierpsychol.* **1943**, 235–409.
60. Glocker ML, Langleben DD, Ruparel K, Loughead JW, Gur RC, Sachser N. 2009 Baby schema in infant faces induces cuteness perception and motivation for caretaking in adults. *Ethol.: former. Z. fur Tierpsychol.* **115**, 257–263. (doi:10.1111/j.1439-0310.2008.01603.x)
61. Kringelbach ML, Stark EA, Alexander C, Bornstein MH, Stein A. 2016 On cuteness: unlocking the parental brain and beyond. *Trends Cogn. Sci.* **20**, 545–558. (doi:10.1016/j.tics.2016.05.003)
62. Gould SJ. 1980 'A biological homage to Mickey Mouse'. In *The panda's thumb* (ed. SJ Gould), pp. 95–107. New York, NY: Norton.
63. Morreel J. 1991 Cuteness. *Br. J. Aesthetics* **31**, 39–47. (doi:10.1093/bjaesthetics/31.1.39)
64. Fisher RA. 1930 *The genetical theory of natural selection*. Oxford, UK: Clarendon Press.
65. Graham KE, Furuichi T, Byrne RW. 2020 Context, not sequence order, affects the meaning of bonobo (*Pan paniscus*) gestures. *Gesture* **19**, 335–364. (doi:10.1075/gest.19028.gra)
66. Planer R, Sterelny K. 2021 *From signal to symbol: the evolution of language*. Cambridge, MA: MIT Press.
67. Pyers J, Senghas A. 2009 Language promotes false-belief understanding: evidence from Nicaraguan sign language. *Psychol. Sci.* **20**, 805–812. (doi:10.1111/j.1467-9280.2009.02377.x)
68. Burkart J *et al.* 2014 The evolutionary origin of human hyper-cooperation. *Nat. Commun.* **5**, 4747. (doi:10.1038/ncomms5747)
69. Neubauer S, Hublin J-J, Gunz P. 2018 The evolution of modern human brain shape. *Sci. Adv.* **4**, eao5961. (doi:10.1126/sciadv.aao5961)
70. Gould SJ. 1977 *Ontogeny and phylogeny*. Cambridge, MA: Belknap Press.
71. Somel M *et al.* 2009 Transcriptional neoteny in the human brain. *Proc. Natl Acad. Sci. USA* **106**, 5743–5748. (doi:10.1073/pnas.0900544106)
72. Wilkins A, Wrangham R, Fitch WT. 2014 The 'domestication syndrome' in mammals: a unified explanation based on neural crest cell behavior and genetics. *Genetics* **197**, 795–808. (doi:10.1534/genetics.114.165423)
73. Johnsson M, Henriksen R, Wright D. 2021 The neural crest cell hypothesis: no unified explanation for domestication. *Genetics* **219**, iyab097. (doi:10.1093/genetics/iyab097)
74. Kruska DC. 2005 On the evolutionary significance of encephalization in some eutherian mammals: effects of adaptive radiation, domestication, and feralization. *Brain Behav. Evol.* **65**, 73–108. (doi:10.1159/000082979)
75. Wrangham R. 2018 Two types of aggression in humans. *Proc. Natl Acad. Sci. USA* **115**, 245–253. (doi:10.1073/pnas.1713611115)
76. Ponce de León M *et al.* 2021 The primitive brain of early *Homo*. *Science* **372**, 165–171. (doi:10.1126/science.aaz0032)
77. Israel S *et al.* 2009 The oxytocin receptor (OXTR) contributes to prosocial fund allocations in the dictator game and the social value orientations task. *PLoS ONE* **4**, e5535. (doi:10.1371/journal.pone.0005535)
78. Luo L, Ma X, Zheng X, Zhao W, Xu L, Becker B, Kendrick KM. 2015 Neural systems and hormones mediating attraction to infant and child faces. *Front. Psychol.* **6**, 970. (doi:10.3389/fpsyg.2015.00970)
79. Camerino C. 2020 The new frontier in oxytocin physiology: the oxytonic contraction. *Int. J. Mol. Sci.* **21**, 5144. (doi:10.3390/ijms21145144)
80. Staes N, Stevens JM, Helsen P, Hillyer M, Korody M, Eens M. 2014 Oxytocin and vasopressin receptor gene variation as a proximate base for inter- and intraspecific behavioral differences in bonobos and chimpanzees. *PLoS ONE* **9**, e113364. (doi:10.1371/journal.pone.0113364)
81. Rilling JK, Scholz J, Preuss TM, Glasser MF, Errangi BK, Behrens TE. 2012 Differences between chimpanzees and bonobos in neural systems supporting social cognition. *Soc. Cogn. Affect. Neurosci.* **7**, 369–379. (doi:10.1093/scan/nsr017)
82. Rockman MV, Hahn MW, Soranzo N, Zimprich F, Goldstein DB, Wray GA. 2005 Ancient and recent positive selection transformed opioid cis-regulation in humans. *PLoS Biol.* **3**, e387. (doi:10.1371/journal.pbio.0030387)
83. de Boer M, Kokal I, Blokpoel M, Liu R, Stolk A, Roelofs K, van Rooij I, Toni I. 2017 Oxytocin modulates human communication by enhancing cognitive exploration. *Psychoneuroendocrinology* **86**, 64–72. (doi:10.1016/j.psyneuen.2017.09.010)
84. Hrdy S, Burkart J. 2020 The emergence of emotionally modern humans: implications for language and learning. *Phil. Trans. R. Soc. B* **375**, 20190499. (doi:10.1098/rstb.2019.0499)
85. Dunbar RIM. 1996 *Grooming, gossip and the evolution of language*. London, UK: Faber & Faber.
86. Weissner P. 2014 Embers of society: firelight talk among the Ju/'hoansi bushmen. *Proc. Natl Acad. Sci. USA* **111**, 14 027–14 035. (doi:10.1073/pnas.1404212111)
87. Mehl MR, Vazire S, Ramirez-Esparza N, Slatcher RB, Pennebaker JW. 2007 Are women really more talkative than men? *Science* **317**, 82. (doi:10.1126/science.1139940)
88. Tomasello M. 2014 *A natural history of human thinking*. Cambridge, MA: Harvard University Press.
89. Bickerton D. 2014 *More than nature needs*. Cambridge, MA: Harvard University Press.
90. Nakamura M. 2000 Is human conversation more efficient than chimpanzee grooming? Comparison of clique sizes. *Hum. Nat.* **11**, 281–297. (doi:10.1007/s12110-000-1014-2)
91. Dunbar RIM, Duncan NDC, Marriott A. 1997 Human conversational behavior. *Hum. Nat.* **8**, 231–246. (doi:10.1007/BF02912493)
92. Heritage J. 2013 Epistemics in conversation. In *Handbook of conversation analysis* (eds J Sidnell, T Stivers), pp. 370–394. Chichester, UK: Blackwell.