6 Conclusions

6.1 Reprise

Let us retrace the argument of this book. We began by pointing out the gap between the observable diversity of languages and the relatively fixed anatomy, physiology, and neurocognition that makes language possible. It is our interactional abilities that seem to play a crucial part in filling that gap, the clutch that engages the engine as it were. The crucial role that those abilities play is evident from the surprising amount of communication that can take place without an established language in common (as outlined in Section 2.2), from their central part in language acquisition by children, and the role those abilities play in 'filling out' the sketchy nature of linguistic expression. Those abilities involve both the mechanics of interaction – multimodality, turn-taking, repair, sequence organization – and the inferential and mental modelling involved in attributing purposes and actions to utterances. Many of the details of the organization of informal talk are uniform across cultures and languages, in strong contrast to the diversity of languages. Moreover, aspects of them appear early in infancy, suggesting the unfolding of a developmental programme. Further, aspects like the timing of turn-taking and the contingency between initiating and responsive communicative acts offer some observable continuities with non-human primates, once again in contrast to a focus on language structure. These facts surely point to the path that the evolution of language must have taken over the last million years or so.

The question then arises, how do we account for the very particular direction human communication has taken? I have offered two speculations. The first is that language may have borrowed its semantic architecture partly from our spatial cognition. All our surviving cousins in the great ape family use gesture as their flexible, face-to-face communicative system, suggesting that the last common ancestor and our own subsequent lineage were gestural communicators.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

Since gesture is a spatial form of communication mostly about space, it would have opened up spatial cognition as a source of structure. Spatial cognition also involves predicting what arrays will look like from the other direction – otherwise we would miss our landmarks and lose our way. In this way it also exercises our ability to take the other's point of view, the extension of which is crucial to understanding the minds and actions of others.

The second speculation is that our ability to model what is going on in other people's minds may have its origin in a generalization of the mother–infant relation into the general population at large. I have suggested that a runaway process of 'cuteness selection' may have been involved, whereby preferences for cute infants may have generalized to preferences for cute mates, so engendering the observable tendencies in human evolution towards greater gracility of build and greater tolerance of others with its attendant reduction of in-group aggression. The trigger here may have been the outsourcing of childcare, which by freeing mothers for further reproduction built the basis for the human demographic explosion. But the outcome has been a generalization of maternal empathy and mind reading to the population at large, so building the basis for an elaborate communication system founded on recognizing others' communicative intents. Such a system allows the short-circuiting of complex inference through the use of previously established signals, so allowing the slow accumulation of practices that we call linguistic conventions. Much of the complexity of the syntax of languages can be attributed to the demands of the interactive uses of language.

We then went on to ask what has motivated all this communication and interaction in the first place? The intensity and duration of human interaction is unmatched, and may have been driven by a number of complementary functions, from updating crucial information, maintaining large numbers of relationships in and beyond vulnerable foraging groups, winkling out 'free riders', and so forth. But particularly evident in the details of conversational practice is the delicate handling of social relationships, which are easily damaged by lack of empathy and respect. When corrections in the other's behaviour need to be made, they are often made with tact and artifice – as in the pointed but gentle jokes between friends. When calls on others' time and resources are made, they are usually made with considerable circumspection, so that fulfilling a request can be seen as a freely given gift. What this

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

suggests is that in the fission/fusion structure of social life we have inherited from the great ape family, care is taken to hold the group or the relationship together. In an Aboriginal community I am familiar with, or among the Rossel Islanders, the offended simply decamp, taking their dependents with them. It is this massaging of delicate, independent egos that may explain the great investment of time and energy put into social interaction with our fellows. In the relationships we cannot walk away from – kinship relations or relations of authority – even more care has to be taken.

The consequent rituals of interpersonal life offer insights into how we structure our social systems. The patterns of exchange of politeness or honorifics allow the analyst to map out the relationships between individuals and groups on the dimensions of closeness, both horizontal and vertical. Similarly, one can see how many of the institutions of society can be built by restricting different aspects of informal interaction. The constructional principles of social systems become manifest.

The overall claim of the book is, then, that in trying to understand how our species came to occupy our extraordinary niche in the biological world, much more attention should be given to the interactional base which gives us our 'carrying capacity' for language and culture. Taken as a whole, one can read the interactional spoor on the tracks that led humans to evolve their complex communication system, the miracle of language. The evolutionary story, as far as we can tell, goes something like this:

1. As we split from the closest apes about 6 million years ago, along with similar biology we inherited aspects of their social and communicational repertoire, namely a limited vocal repertoire of largely reflex signals, along with a more inventive gestural communication system with a precise temporal metric of turn-taking – signals of about two seconds long with 200 ms gaps between them.
2. By 4 million years ago our ancestors were bipedal foragers in the savannah, their upright posture making the entire ventral surface of the body part of their signal system and freeing the hands for tools and gesture. Exposed to predators and rival scavengers, they needed to act cooperatively in large fission/fusion groups, putting a premium on aggression-reduction and communication.
3. These factors favoured more complex communication, requiring advanced abilities to read intentions and understand the others'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

point of view. Evolution exploited two pre-adaptations. First, motivated by 'alloparenting', caring and empathy of mother–infant pairs was generalized through to the wider population by a run-away process of 'cuteness selection', leading to extended intention-attribution or 'theory of mind'. Second, the process of adopting the other's viewpoint, together with reliance on gesture, led to the exploitation of spatial cognition. These processes led to human forebears becoming social junkies, high on oxytocin, and utilizing spatial primitives to construct propositional concepts.

4. By 1.5 million years ago our ancestral hominins were cooperative hunters and foragers using complex tools, with a cultural system sustained by elaborated sign language, no doubt with some simple vocal accompaniment.
5. Increasing use of the vocal channel may have been occasioned by increasing preoccupation of the hands with other tasks, night-time communications encouraged by fire, the advantages of communication over distance, or simply by a maximization of the bandwidth made available by multimodal communication. By 1 million years ago ancestral humans were acquiring the breath control and vocal tract for complex speech, leading to relatively rapid cultural and technological innovation in larger groups. Yet the ancestral timing characteristics mentioned in point 1 persisted. Both biological and cultural evolution were now partly guided by group selection – improved communication and technology improved the life chances of the group.
6. By the time of the common ancestor of anatomically modern humans and Neanderthals (perhaps 600–700 kya), modern language capabilities were largely in place. Humans now had complex propositional language squeezed into the ancestral timing of turn-taking, with the result that the cognitive processing of language was now one of the main pressures on cognitive development.
7. Language afforded the construction of much more elaborate social systems, and the conceptual means to construct social institutions. The ancestral fission/fusion group dynamics were stabilized by the fine-grained adjustments in interpersonal relations, the juggling of relative rank and social closeness, made possible through linguistic interaction.
8. Language today bears the imprint of its origins as the social glue for holding together groups through sustained verbal interaction.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

6.2 How to Construct a Language: The Shape of Paleolithic Languages

So far, we have sketched the bare-bones progression of the language capacity, as shown by what we know about the likely development of the interaction engine, and the clues from the genetics of the physiology of speech and hearing. But concluding that the evidence points to full-blown vocal language capacities by about 700 kya, leaves us not that much the wiser about what ancient languages were actually like. Did the last common ancestor with Neanderthals have advanced syntax, a property that seems to demarcate man from beast? What could they and couldn't they talk about – what would they have used language for? What would the semantic system have been like? What speech acts would they have used?

Obviously much of this will be forever in the dark. Three-quarters of a million years ago is twenty times further back in the past than the last Neanderthals, who themselves were twice as far back as the painted caves of Lascaux, which is itself twice as far back in time than we can trace actual language histories. Given this depth of time, can we say anything about the languages spoken by the last common ancestor shared with us and Neanderthals? Surprisingly, I think we can. That is because the interaction engine clearly framed their communicative action like it does ours, and even to some extent that of our ape cousins. We can extrapolate that they too used short vocal bursts – one to two seconds long, with a norm of a 200 ms turnaround between utterances. We can be sure, too, that they used adjacency pairs – utterances expecting specific types of contingent responses (apes do the same with gestures). It is very likely that they elaborated these with pre-sequences and post-sequences, making up complex sequences of verbally expressed actions. Judging from the universality and utility of commands or requests, questions and assertions, greetings and partings, they likely had specialized devices for expressing them (even chimpanzees use imperative gestures). Given that early hominins had relied more heavily on gesture than on the vocal channel, we can be sure that gesture played at least as important a part in their multimodal communication as it does in ours. Other multimodal signals, especially gaze with our evolved pale sclera (the whites of the eyes) and facial expressions rendered especially visible by the relative loss of facial hair, are also surely traceable back that far (genes for skin

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

melanin suggest that we were well on our way to nakedness by over a million years ago).[1]

If the speculation about a spatial source of semantic structure is correct, then we can be pretty sure that the last common ancestor had a rich vocabulary for spatial description. Judging from experiments we have done comparing great apes to humans,[2] it is very likely that unlike our familiar left/right/front/back set of coordinates, our ancestors used geographical coordinates like north/south/east/west, as do most foraging communities today. Elaborate specifications of location and direction are then enabled. More generally, the structure of a verb to its arguments (expressed by the nouns that go with the verb), borrowed from verbs of motion ('going from x to y') or transportation ('carrying x to y'), is likely to have been generalized to a broader vocabulary. We have speculated that our theory of mind, and the attribution of meaning or intent to others, emerged from a generalization of empathy out of the maternal relationship. So, it is reasonable to assume that early language expressed propositional attitudes, like 'wanting p', 'thinking p', 'guessing that p is likely', 'saying p', and so forth. This would have made possible many of the key functions of language available today: gossip about shifting social relationships, assessing third parties' reliability, planning future expeditions, recollecting past events, giving technical tips and warnings. The consequent information flow within a community would have offered substantial survival skills to groups – informing about watering holes, recollecting hunting or gathering sites, assessing potentially dangerous rival groups or groups that look like promising allies. We should not forget that although human populations were low through prehistory, they were fractionated into small ethnic groups and groups of quite different origin (like the Eurasian Neanderthals, other archaic humans, and our own African lineage), making encounters dangerous.[3]

Now consider that the evidence shows that the main motivation for developing complex communication in the first place was the need to juggle intricate social relationships. Humans, like chimpanzees and bonobos, came from a line of apes with fission/fusion societies. Like

[1] Rogers, Iltis, & Wooding 2004. [2] Haun *et al*. 2006.

[3] Ancient population sizes can be inferred from genes, e.g. it is estimated that at 120 kya, human populations in sub-Saharan Africa might have been between 120,000–325,000, suggesting perhaps 200–300 distinct languages and cultures at that time (Sjödin *et al*. 2012).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

bonobos, human foraging groups tend to have relatively egalitarian social organizations, where many relationships of friendship, alliance, mating, and exchange can be forged by individuals. But we have seen that these are built around the axes of social distance (close versus distant) and rank (high versus low). Language offers the means for fine-tuning those relations, therefore, we can be reasonably sure that early language had the means to communicate friendship, respectful equality, and respected rank. This respect may have been in the form of the honorifics that many societies use, or through the elaboration of indirection, or even both.

Indirection, we noted, often has its origins in *pars pro toto*, as where one action like *Do you by any chance have matches?* can act to truncate a sequence, so making the request (*May I borrow them?)* redundant. This has also been observed as the origin of ape gestures (a process dubbed ritualization), as when the infant holds up its arms to be carried, so requesting a carrying sequence of actions. This is a powerful process for developing complex meanings out of simpler expressions. Conventions of meaning, we earlier noted, arise when one novel signal has been successfully decoded on a first occasion of use, leading to ease of interpretation on subsequent uses – the pattern then has to pass through the community, and a convention is born.

What about syntax, the carapace of language, constructing larger units out of parts, the armature that supports the words in a sentence, specifying their relationship to the verb? Chomsky has argued that syntax has given us the recursive capacity to express unlimited thoughts, and thus endowed humans with their creativity. But we have seen that recursion, for example in the ability to embed one chunk within another indefinitely, is in fact a remarkable structure of sequential organization in interaction, and so not restricted to syntax (which may have borrowed the trick rather than have donated it). We have also seen that the repair system in language optimizes the ease of response by reproducing the chunk that was understood and highlighting the part that wasn't (as in *Amanda went to York to do what?*). This utilizes and even forces a chunking of the structure – one reason why a holographic system of communication (using an unanalysed chunk to mean a whole proposition, as in early child language) wouldn't work very well. No highly developed inferential communication system like a language will work well without the capacity for efficient repair – our understandings would just drift slowly apart – which is why repair

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

has a universal organization and a high frequency of use, while providing a prime need for chunking of the linguistic signal.

If one considers carefully the interactive uses of language, it becomes clear that many of the functions and properties of constructions reflect the uses to which they are put. We noted that the kind of movement operation that converts *I remember the wood before it was burnt down* --> *The wood, I remember it before it was burnt down* serves an interactional purpose, seeking a nod of recognition for the referent of *the wood*. Nearly all the complex operations in syntax have to do with getting information into an appropriate position for interactive understanding, yielding counterparts like *Anne didn't do it* and *It wasn't Anne that did it* (where the latter appears to deny a presumption in an earlier turn that she did).

If we put all of this together, the idea that the language of the last common ancestor with Neanderthals was a primitive inventory of simple signals seems very unlikely. It may instead have already been a sophisticated, supple system with complex constructions devised to do the jobs that lay behind its evolution in the first place: juggling complex human relationships and passing on crucial information that sustained culture, technology, and subsistence, all within the demanding context of the turn-taking system.

6.3 The Case for the Interaction Engine

There remains a loose end: I have described the interaction engine as built especially on four main parameters, multimodality, timing, contingency, and mind reading. A question is whether these really cohere, which we now turn to.

In this book I have argued for the existence of a special human predisposition for a particular kind of social interaction, which rather than being made possible by language, on the contrary, itself makes language possible – possible to acquire by infants, and possible to pass on over generations with increasing richness. It is interaction that provides the niche for language use, and language bears the imprint of that context, with its fundamental unit the clause adapted to the turn of the turn-taking system. The speed of language processing, the double tasking involved in conversation, will all have left their imprint on the processing machinery for language. Audiologists know full well that the optimal distance for language exchange facing each other is

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

under 2 metres – further than that and unvoiced consonants become less clear, and significant loss of signal occurs at 4 metres. Apes have air sacs that can be used to boom long distance. There is no trace of these in the human fossil record. Instead, we have evolved for the close huddle.

But the arguments for the interaction engine are weaker than they are for many other evolved traits because it is fairly clear that the elements have been slowly accumulating over the millions of years since our split from the apes. Our early bipedalism freed the hands both for tool use and gesture, the white sclera made possible effective pointing and direction of attention, the up-regulation of oxytocin triggered by our fellows made interaction more rewarding, the generalization of some maternal sympathetic instincts to all conspecifics spurred the growth of theory of mind, the growing importance of vocalization endowed us with a highly mobile and resonant vocal tract, and so forth. These are the prehistoric sediments that laid the foundation for our interactional capacities. The curious nature of the current system provides clues to its origin – for example, the compression of extreme complexity into short turns with quick interchanges suggests that language evolved within a pre-existing temporal niche, originally adapted to the exchange of much simpler and more limited signals. Now children struggle to master this demanding system well into middle childhood.

Despite the fact that these capacities have accumulated over deep evolutionary time, there is some evidence that they have grown together intimately enough to act in some ways as a package. The most striking evidence comes from autism, brilliantly and sympathetically studied by the neuroscientist Uta Frith, who drew attention both to the associated deficits in theory of mind, and to its physiological bases.[4] A standard informal list of symptoms typical of autism spectrum disorder (ASD) looks like this (from the website of the US National Institute of Mental Health):[5]

- Making little or inconsistent eye contact
- Tending not to look at or listen to people
- Rarely sharing enjoyment of objects or activities by pointing or showing things to others

[4] See the concise review by Frith 2008.
[5] www.nimh.nih.gov/health/publications/autism-spectrum-disorder

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

- Failing to, or being slow to, respond to someone calling their name or to other verbal attempts to gain attention
- Having difficulties with the back and forth of conversation
- Often talking at length about a favourite subject without noticing that others are not interested or without giving others a chance to respond
- Having facial expressions, movements, and gestures that do not match what is being said
- Having an unusual tone of voice that may sound sing-song or flat and robot-like
- Having trouble understanding another person's point of view or being unable to predict or understand other people's actions

But these characteristic symptoms are all within the set of properties we have examined as characteristic of the interaction engine – coherent and synchronized multimodal signals with mutual gaze, shared attention and theory of mind, turn-taking, and timing. There are striking studies that show how, for example, infants with autism look in the wrong place during interaction, missing crucial multimodal signals.[6] They typically show no social smiles, do not gesture, do not easily recognize others' intentions, and find it difficult to take the other's perspective as required by deictic words like *this* or *there*. They also often show marked deviations from expected contingency patterns.[7]

Now, autism is a spectrum condition, meaning that it varies in severity, and also in the prominence of certain symptoms, making crisp generalizations difficult. It is highly studied, and despite its spectrum character there is little doubt that it constitutes a genuine syndrome. It is a highly heritable trait, as shown by studies of identical twins. Given its variable expression, one would expect no single genetic origin, and that proves to be the case, since many individual cases involve new mutations that may affect multiple different genes. Nevertheless, there is a very strong association between autism and oxytocin, the hormone we have earlier encountered as associated with human and animal bonding, sexual pleasure, and establishment of trust.[8] Recent work shows that autism is strongly associated with mutations of the oxytocin reception gene OXIR.[9] There are also more

[6] Jones & Klin 2013. [7] Bottema-Beutel 2017. [8] Kosfeld *et al.* 2005.
[9] LoParo & Waldman 2015.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

general, population-wide associations of oxytocin levels with sociability and communicative competence. There are even studies that show that oxytocin administration improves the theory of mind operations involved in targeting communication to a specific audience.[10] Although it may be simplistic to think of human society as bonded by chemical amity in the way that ants are, there are strong hormonal influences on the foundations of social interaction.

There is an interesting contrast between autism and Down's syndrome. Although high-functioning autistic individuals may be of high intelligence and verbal ability, these are areas in which Down's individuals are typically impaired. Nevertheless, Down's individuals are interactively competent, prosocial, and show few or none of the specific problems with interaction that are diagnostic of autism like avoidance of mutual gaze, problems of intent recognition, or problems with turn-taking or timing.[11] The contrast between autism and Down's syndrome is what psychologists call a double dissociation, with each set of behavioural traits associated exclusively with each condition.

There are other syndromes that may relate to interactional impairment. DSM-5, a manual of mental disorders, has recently added social pragmatic communication disorder (SCD) to its list of known distinctive clinical profiles.[12] Individuals diagnosed with SCD show persistent problems with communication both linguistic and multimodal (for example, giving inappropriate responses) that cannot be explained by low IQ or ASD. ADHD (attention deficit hyperactivity disorder) also correlates with interactional problems and serious social difficulties, partly associated with aggression.[13] The reduction of aggression in humans is clearly linked to cooperative tendencies and increasing skeletal gracility. If one makes a distinction between 'cold' calculating aggression versus 'hot' reactive aggression – which is also reflected in neural pathways – unlike chimpanzees, humans have reduced the tendency for 'hot' aggression, while amplifying 'cold' aggression.[14] Schizophrenia, associated with auditory hallucinations or hearing voices, is another syndrome that relates in an interesting way to the interaction engine. There is little doubt that our inner lives are partly peopled with imagined dialogues. Louis Gould in the 1950s

[10] de Boer *et al*. 2017. [11] Jenkins & Ramruttun 1998.
[12] American Psychiatric Association 2013. See also Bishop & Norbury 2002.
[13] Carpenter Rich *et al*. 2009. [14] Wrangham 2018.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

found by using electromyography that schizophrenics when hearing voices are actually sub-vocally producing speech, and it was suggested that the phenomenon may arise because the patient's own voice is misattributed or not recognized.[15] This does suggest that a lot of our mental life is formed of inner conversations.

6.4 Built for Interaction?

ASD raises questions about the extent to which humans are creatures highly evolved for communicative interaction. What case is there for this, if we cast our eyes over the entirety of human physiology and psychology? Evolutionary processes have clearly made a large investment in our language capacities. A host of minor and major adaptations of brain, vocal tract, audition, and the genes that build them have taken place: relative to the other apes, we have a greatly enlarged Broca's area (BA 44, 45), the connective fibres of these frontal areas of the brain to the temporal lobe and auditory cortex have been enlarged, a direct connection from motor cortex to the larynx and to the breathing system give increased voluntary control of vocalization, and our hearing is fine-tuned to the complex consonantal distinctions made in human languages.[16] We are at the start of a revolution in our understanding of the genome, and the genes involved in developing language abilities are only beginning to be unravelled: early findings of a regulatory network involving FOXP2 and CNTNAP2 (another gene implicated in language) were made possible by the discovery of a single mutation and its effects on the growth of the neural connections behind language production.[17] All this infrastructure supports specialized neurocognition, for example specialisms of the auditory cortex to the processing of speech sounds, their encoding as motor programs, the retrieval of lexical items.

Language, then, provides compelling evidence for the specializations of human anatomy and cognition for communicative interaction. Beyond the strictly verbal processes going on in the individual, studies are beginning to show the synchronization processes involved in interaction. It has long been known that if in interaction one person puts their hands behind their head, the other is likely to, or if one

[15] Bick & Kinsbourne 1987. [16] Fitch 2018. [17] Fisher 2017.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

adopts an accent, the other is likely to accommodate it by minimizing the accent divergence. Now we also know that brain synchronization or entrainment also takes place, beyond any entrainment simply caused by the speech signal itself.[18] Work on the neurocognition of turn-taking shows that there is very early speech act identification (the precondition for quick responses) and even earlier tracking of possible cues to turn completion.[19] The neural signal also shows a high sensitivity to lengthened gaps of a few hundred milliseconds in conversation, where a gap after an initial speech act or first part of an adjacency pair suggests a negative response.[20] None of this is likely to have parallels in the other great ape species since they lack the fine control of vocalization and the ability to vocally imitate.

The human hand is a marvel of bio-engineering, with its twenty-seven bones and thirty-six muscles, offering around 60 degrees of freedom in its movement, which must constitute a problem in its neural control.[21] The development of the opposable thumb is unique among the primates. The presumption has always been that the evolution of the hand, and in particular its opposable thumb, was driven by early human tool use. However, the complexity and flexibility of the hand seems under-determined by the making of stone tools. The interesting speculation arises whether some of the flexibility and fine control could have been driven by gesture. We have seen that there is a long-standing theory that human vocal language was preceded by a signed gesture system, and while the typical gestures we make when speaking may not require exquisite control, sign languages do. All sign languages make use of highly detailed hand shapes, often involving, say, the independent movement of two or three digits, and they require very rapid transitions from state to state on both hands.

Our upright stance, in addition to freeing our hands, makes available a broad range of postures for signalling confidence, diffidence, trust, and caution. In addition, the upright torso can indicate muscular tension or relaxation, and, as every underwear advert makes clear, a level of physical fitness. But the deployment of the body in conversation can play an important role in how the conversation develops. For

[18] Pérez, Carreiras, & Duñabeitia 2017.
[19] Bögels, Magyari, & Levinson 2015. [20] Bögels, Kendrick, & Levinson 2015.
[21] Santello, Baud-Bovy, & Jörntell 2013.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

example, a twist of the trunk into an unstable position can, by suggesting movement, indicate a momentary digression, before the main business will be resumed.[22] Bodily orientation, and the relative alignment of head, gaze, and trunk, can act to segment talk, and to pull in bystanders or cut them out.[23]

There are numerous other possible adaptations of human physiology to communicative interaction. We have already reviewed the white sclera of the human eye that makes it possible to easily track gaze either to other protagonists or to referents. We have also seen that neonate infants are already sensitive to mutual gaze. For humans, the face is an important repository of information, about identity and more fleetingly about emotion and intention. The relative lack of hair on the human face makes very fine details of facial musculature visible – there's even a theory that the loss of hair might have gone along with three-cone colour vision, allowing us to obtain further colour information about emotional states.[24] Although the muscles of the human face are very closely similar to those of chimpanzees, the chimpanzee face, due to the way the muscles connect to the overlying tissue, does not seem to move as freely and this accords with reports of more restricted facial expressions than humans, most expressions being concentrated around the mouth.[25] Combined with the human relative lack of facial hair, the prominence of eyebrows, and the attention accorded to human eyes, this suggests an evolved platform for facial signalling.

Our communication system, like many animal social systems, relies on rapid and accurate individual recognition. In conversation, we keep track of to whom we have told what, and adjust our whole demeanour according to the specific social relationship. In the human case, the identification relies critically on facial recognition. It used to be thought that other primates use similar cues, but close examination shows this is not so. Old and New World monkeys are poor at recognizing individuals from facial cues alone, and when they succeed they use means other than the human-specific face-selective neural pathways.[26] Humans are also good at recognizing from faces whether someone is likely related by kinship to another known person, which may have played a special role in the early human organization of

[22] Schegloff 1998. [23] Kendon 1990. [24] Changizi, Zhang, & Shimojo 2006.
[25] Burrows *et al*. 2006. [26] Rossion & Taubert 2019.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

large groups.[27] Our voice recognition excels – we can distinguish a familiar voice from the initial sample of one word, a "Hello" delivered on the telephone (this is how familiars typically identify themselves by phone). Other primates are unlikely to have such a large repertoire of social contacts, but interestingly voice recognition and identification seems quite conserved across species (for example, processed largely in the right temporal lobe).[28]

There are many psychological adaptations to our intensely interactional social life, reflected in the many psychiatric conditions occasioned by maladaptation to it. Various kinds of depression, anxiety, bipolar disorder, and so on, are associated with different interactional styles and commitments, and the causation may go in both directions.[29] Interestingly, the 2020–2022 pandemic has demonstrated the effort that goes into impression management in video phone situations, known colloquially as 'Zoom fatigue'.[30] In addition, it has heightened awareness of the multiple conditions caused by social isolation – such as reduced immunity, increased risk of heart problems, and mental illness, effects very evident in those incarcerated in solitary confinement.[31] Humans are so built for social interaction that its deprivation actually causes disease.[32]

6.5 The Interactional Imagination

If our interactional abilities are so deeply baked in our psyche, one might expect many ways in which interactional assumptions might unwittingly impose themselves on our imagination. Probably a great deal of our daydreaming involves mental rehearsals of important interactions, or regretful replays of interactions that went wrong.[33] Witness the famous last lines of James Joyce's *Ulysses*, where Molly's reverie goes:

I was a Flower of the mountain yes when I put the rose in my hair like the Andalusian girls used or shall I wear a red yes and how he kissed me under the Moorish Wall and I thought well as well him as another and then I asked

[27] Alvergne *et al*. 2009. [28] Wilson & Petkov 2011. [29] Cigarini *et al*. 2018.
[30] That fatigue can also be traced to slight delays in turn-taking occasioned by the medium, see Boland *et al*. 2022.
[31] Strong *et al*. 2020. [32] Pietrabissa & Simpson 2020.
[33] For a review of studies of self talk, see Brinthaupt 2019.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

him with my eyes to ask again yes and then he asked me would I yes to say yes my mountain flower and first I put my arms around him yes and drew him down to me so he could feel my breasts all perfume yes and his heart was going like mad and yes I said yes I will Yes.

Socrates even went so far as to consider the idea that thought is basically inner speech, 'the talk which the soul has with itself about any subjects which it considers'.[34] If so, the thoughts one entertains might be much coloured by the language one happens to speak, a doctrine known as the Sapir-Whorf hypothesis, and one much resisted by psychologists, despite some strong evidence for effects in that direction. To explore internal speech, psychologists gave subjects beepers to record what exactly they were thinking when the beep randomly occurred. It turned out that around a quarter of the time, people were involved in inner speech or dialogue (although there was a great deal of individual variation), and they could be trained to record the actual words. So, there is some objective evidence for what is an ineluctably subjective experience, the prevalence of internal speech and dialogue.[35] This accords, of course, with our conscious awareness of rehearsing our responses for an upcoming important meeting or some unavoidable confrontation.

But there may be less obvious ways in which our psyches are pervaded through and through with interactional assumptions. When Victorian travellers and colonists ventured into foreign lands, they much remarked upon the superstitions of the 'natives' who seemed to live in a world peopled by godlets, evil demons, and malevolent spirits – they ignored of course their own heartfelt religiosity. Magic and witchcraft became staples of twentieth-century anthropology, and indeed the facts are interesting. I have myself spent years working in a Melanesian society where no death is considered accidental – every fatality, save the youngest infant, has been caused by a sorcerer, who might be masquerading as the victim's best friend. The suspicions and recriminations blight what would otherwise be a simple, but genial life in a land of tropical plenty. Not only are there sorcerers but there are also gods inhabiting many jungle locations whom, if disturbed, can cause death or injury. Indeed, the natural world is full of spirits and guardian gods – many trees and river

[34] Plato 1921, *Theaetetus*, 189e–190a. [35] Hurlburt, Heavey, & Kelsey 2013.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

pools are believed to be inhabited by spirits, snakes may be gods in disguise, and fireflies are the souls of the dead. The world is thus governed by unseen powers that we may brush against; best to treat all these invisible actors with respect, and seek professional help from diviners who can read the signs. In this society, everybody believes in sorcery, including the best-educated. Yet that Melanesian world is not as remote from our own as at first it seems. Although the famous Salem witch trials were held in the seventeenth century, the last such trial occurred as recently as 1878, and the last British conviction under the Witchcraft Act occurred in 1944. Superstitions still encroach on our reason – who has not glanced over their shoulder on a dark, stormy night? Moreover, organized religions thrive throughout the modern world, in which devotees commune with an unseen almighty interactant: all prayer, spoken or silent, is after all a conversation with the deity.

We have, in short, a tendency to endow the world with the intentional agency we are familiar with from all our social interactions. Scientists are not immune from contagion from such magical thinking – Newton's occult beliefs led John Maynard Keynes to say 'Newton was not the first of the age of reason, he was the last of the magicians.'[36] In general, biologists operate as if there was a grand designer (while repudiating the suggestion of course), presuming function in every organ, and we decipher nature much in the same way that we decipher an ancient script, by presuming a message we were intended to be able to read. Political systems, too, operate as if whole polities were unified actors, and not the shambolic coalitions they usually are. Paranoid fantasies about imagined infiltrators, modern-day sorcery like QAnon or mysterious actors in the 'deep state' cloud our rational understanding of political process. We are underlyingly animists.

This kind of magical thinking may pervade our reasoning in a more subtle way. Conversation only works because we employ a special kind of reasoning to understand utterances. If I ask 'Do you have a cup of coffee?' and you reply 'There's a Starbucks around the corner', I know immediately the answer to my question although you didn't supply it – you take it that I want a coffee, but you can't supply it and so supply instead a possible alternative way to satisfy the desire. Unlike normal logical reasoning, conversational reasoning

[36] Keynes [1946] 1978.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

is ampliative (you get more information out than you put in), it is abductive (you infer a premise that would explain the observed utterance), and it is defeasible or context-dependent. These properties follow from the fact that conversational reasoning involves inferring the intention or goal of an utterance, and that in a rich context we feel subjectively certain of the interpretation most of the time. Now those who study the many ways in which human everyday reasoning falls short of logical standards note that we fail on just these sorts of grounds: we add to the premises, we leap to conclusions, we are influenced by the context, we assume the logical puzzle has been put to us for a purpose and we make definite inferences where none are warranted. If I say 'If you turn on the switch the motor starts', you assume that I meant to convey that when you first turn on the switch then as a result the motor will start, and otherwise it won't – but these assumptions about the order of events, their causal relation, and the inference that if you don't turn on the switch the motor won't start, are presumptions based on the nature of cooperative conversation. If you make these assumptions outside a conversational setting, you will be making just the kind of mistaken inferences that Kahneman and Tversky famously enumerated (and which earned Kahneman a Nobel prize in economics).[37] It seems pretty clear that the source of many of these inferential mistakes is that we carry over the reasoning essential to conversation – where we assume a cooperative interlocutor intends us to make these inferences – to other kinds of reasoning.[38]

So, in all these different ways our thought processes may be deeply imbued by the kind of thinking that makes human social interaction work. In the nineteenth century, the anthropologist Edward Burnett Tylor drew attention to the widespread beliefs in animism, which he held to be the primitive roots of religion. The great Swiss psychologist Piaget showed that every child is an animist, at first believing that everything is alive, and only gradually restricting living things to plants and animals. In the modern world, we may no longer make offerings to trees and lakes, but animism has a new lease of life in our fascination with intelligent machines.

[37] See e.g. Tversky & Kahneman 1974.
[38] See Levinson 1995, Sperber, Cara, & Girotto 1995. See also Horn 1973 on early confusions in logic.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007

6.6 Coda: The Further Need for a Science of Interaction

Who has not been bewitched by a smile, fallen in love with a voice, or been entranced by sparkling eyes? Equally, all of us have felt the frustrations of not being able to connect, sensed underlying suspicion, or open hostility. These are the central experiences of our consciousness – the connections to other people and how they are expressed in the conduct of daily interaction. They are the stuff of our dreams and nightmares, and occupy our waking thoughts in moments of reflection. Novels, films, and works of art explore these feelings and intuitions.

This book has been about this central core of our experience, but less about its content than about the unconscious frameworks that guide and construct it. As pointed out, it is curious that there is little science of human interaction, certainly little in proportion to its importance to us. Quite a lot of what we know has been assembled by social scientists using qualitative methods that may or may not generalize, and which stay close to the surface, even though it is now possible to extend investigations into, for example, the underlying neuroscience.[39] Here I have assembled some of the fragmentary scientific evidence about these unconscious frameworks, their evolutionary origin and the mental mechanisms that make them work. This is less a survey of everything we know than a selection of some areas that are better explored, and some areas that are little explored but striking in their obvious importance. Some of the science is quite technical or recherché, but I have tried to make it accessible to the general reader. It transpires that we know remarkably little about how human interaction works, rather less than we know about how it works in other species, including some birds, partly because experiment and data collection is hedged in by ethical issues. But in addition, interaction seems relatively effortless most of the time, it seems transparent to us, and besides, our natural interests lie in the content not the mechanism. It is the medium in which we swim, and so its workings are almost invisible to us. In fact, the mechanisms are incredibly complex, highly evolved, and are perhaps the most crucial hallmark of our species. Understanding more about how interaction works may prove important in understanding not only modern afflictions from ASD to the success of popularism, but also how mankind evolved in the deep past. If this book has awakened the interests of a few stout souls, it will have served its purpose.

[39] Bögels & Levinson 2017.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 23:24:07, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009570343.007