4 *The Origin of the Interaction Engine and Its Role in Language Evolution*

4.1 Precursors to Language

There is an apparently unbridgeable gulf between the inarticulate beasts and humankind. There are simply no good animal models for language, no obvious intermediate steps, no clues to the bridge that must have got us to a special place in the universe, as the communicator extraordinaire. This chapter is about how, nevertheless, we may be able to discern some of the crucial intermediary steps.[1]

Recollect the characterization of the interaction engine developed in the prior chapters, comprising especially the following components: multimodality, contingency between actions, the ascription of intentions, and precise timing, all under the umbrella assumption of cooperative communication. The intense face-to-face interaction, the interest in other minds, the complex sequential structure of conversation and its fine timing, all look as human-specific as language itself. Some of the features we have looked at seem more ethological, part of the human behavioural repertoire. Others seem more cognitive, including the cooperative tendencies, intention reading and of course language itself. In this chapter we pursue the phylogenetic origin of a few of these traits, including turn-taking and multimodality on the one hand, and intention ascription, empathy, and cooperation on the other. Finally, we will turn to language and offer a few speculations about the origin of the propositional core of language, which frames its expressive capacity.

In trying to understand where our communication system comes from, it is essential to compare our communication system to those of the other primates. If one thinks in the traditional way of human

[1] Recent decades have seen a torrent of work on language evolution (particularly readable general introductions are Hurford 2014, Planer & Sterelny 2021, Johansson 2021, Mithin 2024). The treatment in this chapter focuses just on the contribution of a full set of interactional abilities to language origins.

communication as based on language, then there indeed seems to be a Rubicon between humans and other animals – no interesting intermediate steps suggesting how our linguistic abilities might have been slowly accrued in evolution. Perhaps these intermediates would have been evidenced by all those extinct hominins that stand between us and *Homo habilis*, the first great tool user whose origins go back over two million years. But the bridge seems to have been lost.

However, if one thinks of language as resting on an antecedent faculty which still enables it today, and which every child uses to bootstrap itself up into its native language, then suddenly the intermediate ladders become evident. That is one of the virtues of turning the beam of attention from language onto the antecedent powers that make it possible. This is the strategy we pursue here. The chapter begins by examining the behavioural side of the interaction engine, focusing on turn-taking and showing how this focus allows us to say a surprising amount about the communicational continuities between us and the other primates, and even about the likely abilities of fossil hominins. Then, as a bridge to the more cognitive aspects of the interaction engine, we take the persistence of gesture in human communication as a clue to the role it may have played in language evolution. This leads us to spatial cognition, which is what drives most gesturing, and we show that spatial concepts may have provided a backbone for language semantics. Then we turn to that fundamentally mental (and uniquely human) element in the interaction engine, theory of mind, and propose a novel origin for this in the way we had to broaden our empathetic response to allow the outsourcing of childcare – the secret to our demographic success. If our interactional abilities have played a crucial role in the evolution of language, one might expect that to be clearly evident in the structure of languages, but that is not so obvious. So, at the end of the chapter, we examine the extent to which we can discern the impact of the interaction engine on the structure of modern languages.

4.2 The Phylogeny of Turn-Taking

Let us start by considering the salient character of human communicative ethology, the pulsed alternation of signals that we have called turn-taking. It is possible to trace the distribution of turn-taking across most of the primate order, at least as far as the very partial existing

literature provides the data. There are some 450 species of primate, a family tree with a time depth of up to 80 million years, ranging from the primitive Prosimians to the Platyrrhines (New World monkeys) and the Catarrhines (Old World monkeys and apes), which is the branch including humans (Figure 4.1). There is significant research on the vocal behaviour of many of these species, often under the rubric of 'duetting'.[2] Unfortunately for communication scholars, much of this primate work focuses less on the details of the internal structure of call sequences and more on the functions of the calls. A further problem is that the 'ethogram' tradition in primate work tends to follow a single animal and record its behaviour, rather than focusing on the interaction between animals. Nevertheless, there are some excellent studies of turn-taking behaviour among members of the different branches of the primate tree.

Some kind of turn-taking is evident even in the oldest Prosimian branch of the primate order. For example, published acoustical traces show clear turn-taking among individuals of the species *Lepilemur edwardsi*, a diminutive lemur of Madagascar.[3] Among New World primates, marmoset species have been closely studied, in part because they make good laboratory animals. For example, the Brazilian common marmoset (*Calithrix jacchus*) exhibits precision turn-taking acquired during the first year of life.[4] Starting by producing many calls in overlap, marmoset infants learn by about eight to nine months both to take turns properly and to produce the correct second response to an adult first part – adults may withhold communication to sanction improper use. This closely mirrors the human pattern, where turn-taking becomes regularized during the first year, and where there is also a close interaction between instinct and social learning. Turn-taking timing in marmosets is very different from human adult or child timing, with gaps of the order of five to six seconds compared to human modes of two-tenths of a second (or half a second for children), but there is contingent call matching.[5] This communicative pattern is different from the 'duetting' behaviour of many species because it is not restricted to territorial or mating behaviour, but may occur at any

[2] On the limitations of the existing literature, and the varied uses of the terms 'duetting', 'chorusing', and 'antiphonal' vocalizations, see Pika *et al.* 2018.
[3] Mendez-Cardenas & Zimmermann 2009. [4] Chow, Mitchell, & Miller 2015.
[5] Takahashi, Narayanan, & Ghazanfar 2013.

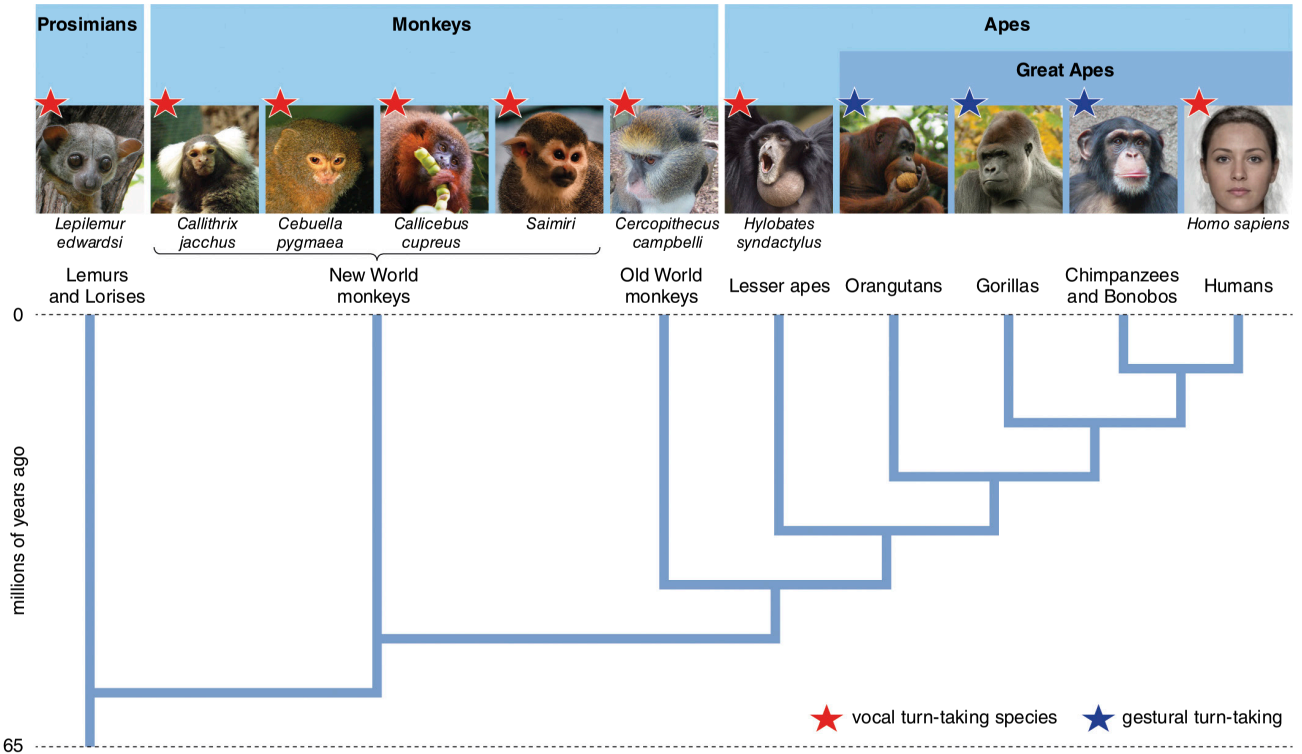

Figure 4.1 Turn-taking across the primate order (after Levinson 2016). Vocal turn-takers can be found right across the major groups of the primate order, but the great apes, not counting humans, are a curious exception: they are gestural turn-takers. (Images: average human courtesy Lisa DeBruine & Benedict Jones; remainder reproduced under Creative Commons license, Credits left to right: Frank Vassen, Raymond Spekking, Malene Thyssen, David Weitzberg, Steve Wilson, Badgernet, Suneko, Eleifert, Roger Luijten, Thomas Lersch).

time between conspecifics within range. Like humans, this marmoset species is a cooperative breeder, meaning that individuals other than the parents may be involved in infant care, and this appears to be a motivation for this kind of catholic communication with other individuals. Many other New World monkeys have been reported to engage in vocal turn-taking, including the pygmy marmoset *Cebuella pygmaea*,[6] the coppery titi *Callicebus cupreus*,[7] and squirrel monkeys of the *Saimiri* genus.[8]

Turning to Old World species closer to our human line of descent, there are detailed reports of the communication behaviour of Campbell's monkeys (*Cercopithecus campbelli*), which clearly have one of the most complex vocalization patterns among the Old World monkeys. Their calls are composed of elements that in a specific sequence may convey, for example, alarm on the sighting of a leopard versus an eagle.[9] Females exchange calls in a turn-taking A-B pattern with less than a one-second gap between turns. But young individuals have to learn the pattern, initially producing more inappropriate call patterns. Playback experiments using appropriate versus inappropriate calls showed that adults compared to their offspring paid much greater attention to well-formed exchanges. This displays again the pattern of a part instinctual, part learned exchange of signals. Japanese macaques (*Macaca fuscata*) are another species with attested turn-taking, exchanging three main call types with a fast pattern around 250 ms, close to the human norm.[10] Like the marmosets, they also leave a longer gap if there is no response before trying to initiate a sequence again. Other Old World monkeys, including geladas, have been reported to engage in turn-taking vocalizations.[11] Coming ever nearer to our own descent line, among the lesser apes, gibbons display elaborate, highly ritualized exchanges, with both chorus elements and antiphonal elements involving rapid turn-taking.[12]

Looking for evolutionary connections to our species, obviously the great apes are of special interest. But here we meet a discontinuity with

[6] Snowdon & Cleveland 1984. [7] Müller & Anzenberger 2002.
[8] Symmes & Biben 1988.
[9] Ouattara, Lemasson, & Zuberbühler 2009. For turn-taking in this species see Lemasson *et al.* 2011.
[10] Katsu *et al.* 2019.
[11] Richman 1976 in the interpretation by Pika *et al.* 2018.
[12] Geissmann 1999. See also Geissmann & Orgeldinger 2000 and Haimoff 1981.

some of the other turn-taking species already mentioned. The great apes are reported to have very limited and relatively simple repertoires of vocalizations, most of which are involuntary or are prompted by specific stimuli and tend to be produced in chorus or in overlap.[13] Little evidence for spontaneous systematic vocal exchanges has been found among chimpanzees, apart from long-distance calls for coordinating movement: only around 10 per cent of calls are responded to within five seconds.[14] Recently, though, a number of exceptions to this earlier generalization about great ape vocalizations have been described. For example, gorillas exchange soft contact calls or 'grunts', mostly only between age mates, in a turn-taking manner (other calls are more likely to overlap). These grunts occur with gaps of around half a second, a bit slower than typical human responses but within the same range.[15] Playback experiments violating the 'one at a time' rule for 'grunts' proved less interesting to gorillas than ones that observed gaps between turns.[16] Another recent finding is that bonobos exchange about half a dozen different call types, often in overlap, but that one kind of call (so-called peeps or pee yelps) were exchanged predominantly with about 250 ms gaps.[17] These recent findings are based on captive individuals where close-up microphones were able to record vocal interactions, so it is possible that there is more coordinated vocal exchange in the wild than has been captured by current methods.

However, while the great apes do not seem to display vocal virtuosity or much vocal turn-taking, they do have extensive and more flexible gesture repertoires. Even here though, overall rates of response are only of the order of 15–20 per cent.[18] But a different picture emerges particularly if one concentrates on mother–infant interaction. For example, in bonobos, gestures by either mother or infant to initiate the infant mounting on the mother for the purposes of joint travel show remarkable affinity to a human adjacency pair of request-and-compliance.[19] Further studies of joint-travel initiating gestures in both bonobos and chimpanzees in the wild show deep systematics.[20] The timings are rapid, often overlapping, but with human-like 200 ms gaps in the majority of cases. The gestures

[13] Call & Tomasello 2007. [14] Arcadi 2000.
[15] Lemasson, Pereira, & Levréro 2018. [16] Pougnault *et al.* 2020.
[17] Levréro *et al.* 2019.
[18] Call & Tomasello 2007, Liebal, Müller, & Pika 2007, Pika & Mitani 2006.
[19] Rossano 2013. [20] Fröhlich *et al.* 2016, Rossano 2019.

themselves have various origins, but are sometimes ritualized between mother–infant pairs from the initial movements involved in mounting, and so are specific to dyads, but others have commonalities across pairs.[21] There are some differences between bonobos and chimpanzees here, with greater emphasis in bonobos on visual rather than tactile gestures and thus on gaze in interaction, and a more cooperative, faster, and more anticipatory style in interaction. Although the most human-like interchanges have been reported from mother–infant pairs, there is also more general systematic gesture use, for example of food requests from females to dominant males in orangutans, again with rapid adjacency-pair structuring.[22] In general there is sufficient evidence that the kind of turn-taking found in vocalizations across the primate order is found instead more in the gestural domain among the great apes.

Putting the primate facts together we have the kind of distribution of turn-taking behaviour right across the primate order shown in Figure 4.1, with rapid turn-taking in the great apes primarily but not exclusively gestural in character. At present we do not know how general this characteristic is across all 450 species, but it may be concentrated among species with pair-bonding or high degrees of cooperative behaviour, including cooperative parenting. If it turns out to be general, it is likely a conserved evolutionary trait, but in any case, the close affinities in timing and flexibility between human vocal turn-taking and ape gestural exchange make a prima facie case for evolutionary precursors to this aspect of human behaviour.

4.3 Turn-Taking and the Evolution of Language

From the patterns of turn-taking among the great apes, it seems that, in our ancestral state at the time of the last common ancestor of humans with chimpanzees around 6 million years ago, we were mainly gestural rather than primarily vocal turn-takers. Ape gestures sometimes co-occur with vocalization – the vocalization can help to draw attention to the gesture. This multimodal background provides the context in which a gradual increase in dependence on the vocal channel likely developed. Because modern human physiology shows evolved

[21] Liebal, Schneider, & Errson-Lembeck 2019. [22] Rossano & Liebal 2014.

specialisms for vocalization we are able to say something about the time course of this increasing reliance on the vocal channel.

The first important datum comes from a 1.6 million-year-old (1.6 my) fossil hominin, containing a rarely preserved vertebral column. Known as the Nariokotome boy (or KNM-WT 15000) he was about eleven years old when he died, and would have grown into a tall, thin man. He is assigned to the species *Homo ergaster*, an African form of *Homo erectus*, the first hominin species to range over vast areas of Eurasia. These are the hominins that made the hand-axes, the pear-shaped distinctive tools of the Lower Palaeolithic, made with a clear symmetrical mental model in mind and worked with a skill that takes months of instruction for modern students to emulate. It is highly unlikely that such tools could have been made without instruction about material sources and techniques, implying an already advanced communication system capable of supporting complex cultures able to adapt to many different ecologies in both Africa and Eurasia. But this species, judging from the Nariokotome boy's vertebral column, was very unlikely to have been an advanced vocal communicator. That's because efficient verbal language requires advanced breath control. While many primates (including chimpanzees) vocalize on both the in-breath and the out-breath, humans use a special interrupt of the autonomic breathing system to make it possible to breathe just before speaking, so having the air pressure to power the vocal tract, and to do so with superfine control so that every stress and emphasis is reflected in released air pressure.[23] The interrupt is a nervous pathway that connects the motor cortex directly with the intercostal muscles used to inflate the lungs, and the requisite nerves pass through the thoracic vertebrae in an enlarged spinal canal, lacking in chimpanzees. That broad channel was also missing from the Nariokotome boy. It is reasonable then to assume that he was, like the apes, primarily using a gesture system with rapid turn-taking for close intentional communication, rather than finely controlled vocalization.[24]

This conclusion has been challenged on the basis that this is too much weight for a single fossil to bear, and there is a remote possibility that the Nariokotome boy suffered from spinal stenosis, a pathological narrowing of the thoracic vertebrae. This criticism is given a

[23] McKay *et al.* 2003, Torreira, Bögels, & Levinson 2015.
[24] MacLarnon & Hewitt 2004.

certain weight by the discovery of a few thoracic vertebrae from the European species of *Homo erectus* which do not appear narrowed in the same way.[25] However, the Eurasian *Homo erectus* with the larger vertebral canal was almost certainly not ancestral to our line, so may be less pertinent to our story. Even if the Nariokotome spine is pathological (which on balance seems unlikely), the only effect would be to push back the inception of human fine breath control earlier, somewhere between our split with the chimpanzee branch six million years ago and, say, 2 million years ago (2 mya). Another line of evidence comes from recent work on endocasts, the traces of brain structure visible on the interior of the skull, which show that the African *Homo ergaster* lineage underwent a brain reorganization around 1.5 mya, with the expansion of the prefrontal cortex (and the language-critical Broca's area) pushing back the precentral inferior sulcus. These changes are a likely signature of gradual language development.[26]

If, despite the interpretive uncertainties, we take the narrow spinal canal of the Nariokotome boy to indicate a lack of fine breath control, then that gives us a date before which highly developed vocal language probably did not exist. The *Homo ergaster* species he represents is assumed to be directly in our line of descent. The next important datum comes from extensive fossils belonging to Neanderthals from many Eurasian sites and from proto-Neanderthals from Atapuerca in Spain. The Spanish fossils from the Sima de Los Huesos cave are remarkable for their number and completeness, and for the recovery of ancient DNA from 430,000 year-old fossils.[27] The DNA shows that these individuals are likely proto-Neanderthals with Denisovan admixture (Denisovans are a sister branch of ancient hominins – see Figure 4.2) who diverged from the modern human lineage about 700,000 years ago (700 kya). All the evidence points to these individuals and the Neanderthals that followed them being fully articulate hominins. For a start, Neanderthals had the right genes as far as we can tell – critically, the same variant of FOXP2 as we have, a gene known to play an essential role in language capacity.[28] Second, despite earlier doubts, we now know they probably had the modern vocal tract that enables language.[29] Third, they had the same hearing sensitivity,

[25] Meyer & Haeusler 2015. [26] Ponce de León *et al*. 2021.
[27] Meyer *et al*. 2016. [28] Fisher 2019. [29] Barney *et al*. 2012.

concentrated in the bandwidth central to speech, as modern humans.[30] Fourth, they had the special enervation of the thoracic vertebrae implicated in precise breath control for speech – the property missing from Nariokotome boy. Fifth, recent discoveries seem to confirm that they were also symbolic creatures, using large amounts of coloured ochres, burying their dead with grave goods, and producing cave paintings and hand stencils.[31] Extraordinary structures found deep inside Bruniquel Cave, circular walls made of four courses of stalagmite and dated to 175 kya, speak of some kind of ritual use: way beyond daylight, and showing no signs of habitation.[32] Although there were no doubt subtle cognitive differences from modern humans, which may become clear as we learn more about ancient DNA, there can be little doubt that Neanderthals were an articulate variety of human.

Circumstantial archaeological evidence points in the same direction. Neanderthals utilized advanced technologies using flakes from prepared cores for tool manufacture, crafted aerodynamic wooden javelins, and composite tools with microliths fastened with adhesives made from tree bark. Fire and clothing made from skins allowed them to thrive in ice-age conditions – evidence of tanning, tools for softening skins, and even thread have survived. They brought down mammoths and other large and dangerous game animals which would have required planning and group effort. Game included rhinoceros, aurochs, and cave bear – all fierce animals over half a ton in weight and over 2 metres high. Some groups even specialized in hunting these huge animals.[33] New evidence shows that they adapted much more flexibly to different ecologies and climatic conditions than used to be thought, using varied toolkits and different foraging strategies.[34]

It is vanishingly unlikely Neanderthals were endowed with all these intellectual and cultural properties without those capacities having been enabled and passed on by language use over scores of thousands of years. It is also unlikely that they would have been able to transmit the advanced technology and hunting skills without the benefit of language. Since Neanderthals and modern humans shared a main common ancestor perhaps as far back as 700 kya, with multiple subsequent

[30] Conde-Valverde *et al.* 2021.
[31] Hoffmann *et al.* 2018. Mithin 2024 is more skeptical of Neanderthal symbolism.
[32] Jaubert *et al.* 2016. [33] Wragg Sykes 2020: chapter 8.
[34] Wragg Sykes 2020: chapter 8.

interbreeding events, vocal language must go back at least that far, perhaps to some earlier Atapuerca fossils attributed to *Homo antecessor* or to the subsequent *Homo heidelbergensis*, both species intermediate between *Homo erectus* and the common ancestor of modern humans and Neanderthals. Figure 4.2 summarizes these inferences.[35]

Given this great antiquity of language, we can assume that the Neanderthal mammoth hunters of Eurasia, like their cousins who were the ancestors of modern humans in Africa, were fully articulate humans with languages perhaps not dissimilar to those of modern hunter-gatherers. Given the vast geographies and timescales involved, we can also be sure there were many languages. As we learn more about the contribution of genes to specific brain areas and the vocal tract, we may be able to home in on some of the properties – there are hints for example that their languages may have been tonal like Chinese. It may even be possible by looking at characteristics of Eurasian languages not found in African languages to pick up ancient echoes of Neanderthal tongues.[36]

If the early members of the human lineage started off as primarily gestural turn-takers like the apes, what happened to this system? The answer of course is that it is still with us, in the shape of the gestures that haunt our words, clarifying what we mean and specifying shape and space. Hereby hangs a second origin story (see Section 4.5).

Although the evidence for Neanderthal language capacities continues to stack up, the account I have told here remains controversial.[37] The desire for human exceptionalism runs deep – the desire to draw a clean line between man and beast. But that is not in line with the general thrust of evolutionary theory, nor indeed with the general thrust of history, which is littered with now-abandoned tenets of cultural, religious, and racial exceptionalism. So those who wish to deny the existence of Neanderthal language will continue to point to the many minor details of genetic and anatomical differences between Neanderthals and modern humans and consider them tell-tale signs of the inarticulateness of our nearest relatives. At the time of writing, for example, an analysis of the genes associated with the oscillations

[35] See Dediu & Levinson 2013, 2018 for much further detail.
[36] Roberts, Dediu, & Levinson 2014.
[37] For the radical alternative, see Berwick, Hauser, & Tattersall 2013, Berwick & Chomsky 2016.

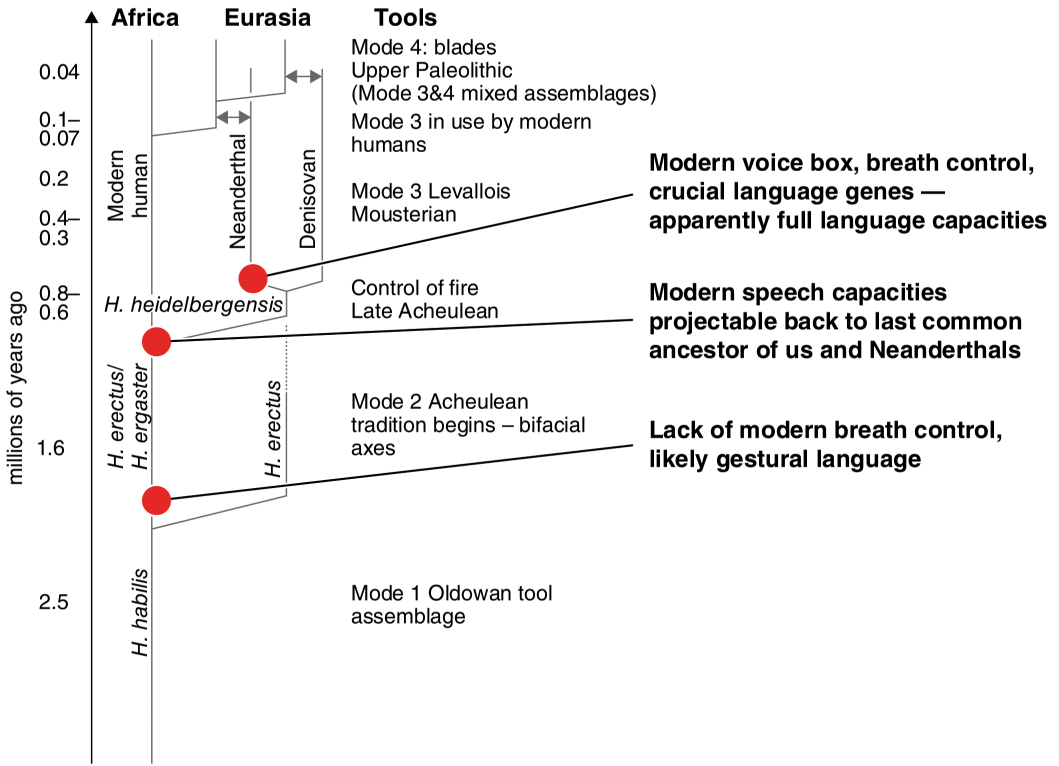

Figure 4.2 Hominin phylogenetic tree with inferred language capacities (after Dediu & Levinson 2013). The tree shows the relations of the species whose genes have contributed to modern humans; time is on the vertical dimension in millions of years with the deep past at the bottom as in an excavation. Modes 1–4 refer to increasing complexity of stone tool types. Language capacities inferred from ancient DNA and fossils.

of the brain suggests small differences between the lineages, taken to be signals of language capacity.[38] But those oscillations are primarily driven by the auditory signal itself, and the latest information from the study of fossil middle ears is that despite anatomical differences the auditory sensitivities of modern humans and Neanderthals were near identical, in contrast even to some much earlier members of the Neanderthal lineage.[39] This precise match in utilized bandwidth ought to be a compelling argument: across mammalian species there is a close correspondence between vocalization output and auditory sensitivity.[40] Although as research proceeds we can expect to find further subtle differences in genome and anatomy between Neanderthals and modern humans, there seems little doubt that both ancient lineages used language, quite likely in very similar ways to our own informal conversation.

A further problem for the exceptionalists is the recent unearthing from ancient DNA of the genetic entanglement between modern humans, Neanderthals, and indeed other ancient hominins. Interbreeding took place repeatedly, as indicated in Figure 4.3, and with reciprocal exchange of genes both ways, with fertile offspring that eventually contributed to our own genome. The evidence points to strong gene flow associated with frequent interbreeding with both sexes of each lineage involved. The degree of intimacy is hinted at by the fact that modern populations outside Africa have inherited Neanderthal oral microbiome (ancient kissing?) and persistent venereal disease.[41] It is also attested by the shared lithic technology – there is a distinctive cultural assemblage known as the Châtelperronian, which is closely similar to the typical Upper Paleolithic blade technology associated with the first modern humans in Western Europe, but it is actually found with Neanderthal genetic material.[42] This hints at many cultural transfers, probably both ways, since Neanderthals had expertise for living in extreme northern environments that modern humans had to learn to inhabit. In short, our genes, our cultures, and probably our languages were entangled over more than 100,000 years of repeated interaction.

[38] Murphy & Benítez-Burraco 2018. [39] Conde-Valverde *et al.* 2021.
[40] Charlton, Owen, & Swaisgood 2019.
[41] Pimenoff, Mendes de Oliveira, & Bravo 2017, Weyrich *et al.* 2017.
[42] Welker *et al.* 2016.

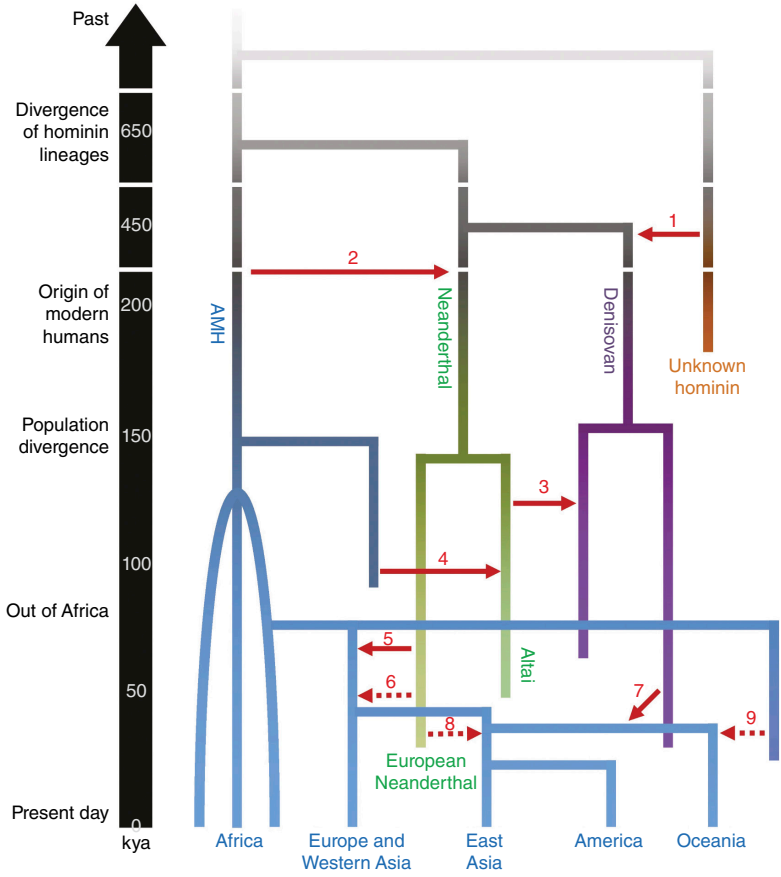

Figure 4.3 Known interbreeding events between modern humans, Neanderthals, and other hominins (from Dediu & Levinson 2018). Time is here represented going downwards with anatomically modern humans (AMH) at the bottom. Interbreeding events are represented by arrows linking branches of the tree at different time depths (Altai indicates a distinctive Siberian branch of the Neanderthal lineage).

4.4 The Ontogeny of Turn-Taking and Gesture

The cradle of civilization is, of course, quite literally the cradle itself. Recapitulation theory is the name given to Ernst Haeckel's 1870s theory that 'Ontogeny recapitulates phylogeny', the idea that embryonic development passes through stages that reflect the earlier history of the organism. The theory is discredited: it ain't so simple, even though

modern 'Evo-Devo' (evolutionary developmental biology) utilizes some of the same ideas about heterochrony – changes in the timing of development that can have a dramatic effect on physiology, and which have been exploited by evolutionary processes to generate new species (see Section 4.6). Nevertheless, children's development of features of the interaction engine can give us clues to their origin, by virtue of their timing – if they are very early and relatively invariant across individuals they are more likely to have an endogenous source and be part of a programme of development, a natural unfolding of abilities of the kind we see in the development of motor skills such as walking. So here we review some recent findings.

It has long been noted that infants engage with their caregivers in 'proto-conversation', a structured exchange of signals (smiles, laughter, coos, etc.) long before they know any words.[43] Our own research suggests that a system like this is in place at three months, but by nine months (still before language is produced) it is highly developed. Figure 4.4 shows a graph showing how overlap of turns between caregiver and infant recedes as the child develops over three years and three months, and at the same time gaps in response by the child get slowly shorter, so at three and a half years the child is responding on average within 500 ms. What is being exchanged early on (at least on the baby's part) is of course not language but inarticulate vocalizations, which as you can see in the graph are in fact produced rather quickly in the second three months of life (the earliest period at which it was easy to measure them).

It has been shown that caregivers' responses divide over whether they are responding to infant fussing and crying as opposed to prelinguistic vocalizations which are syllable-like and occur with various emotional tones. Caregivers respond in overlap to cries, but with a short gap (around 400 ms) to proto-linguistic kinds of sounds.[44] We noted earlier (in Section 3.2) that there is a class of human vocalizations that do not obey turn-taking constraints – these include emotional cries, laughter, in-breaths, exclamations (like *wow*!), and may belong to an earlier evolutionary layer as it were, similar to the involuntary cries of apes. It is interesting that caregivers impose this dichotomy on infant vocalizations. Some researchers have found a 'four-month

[43] Bruner 1983. [44] Yoo, Bowman, & Oller 2018.

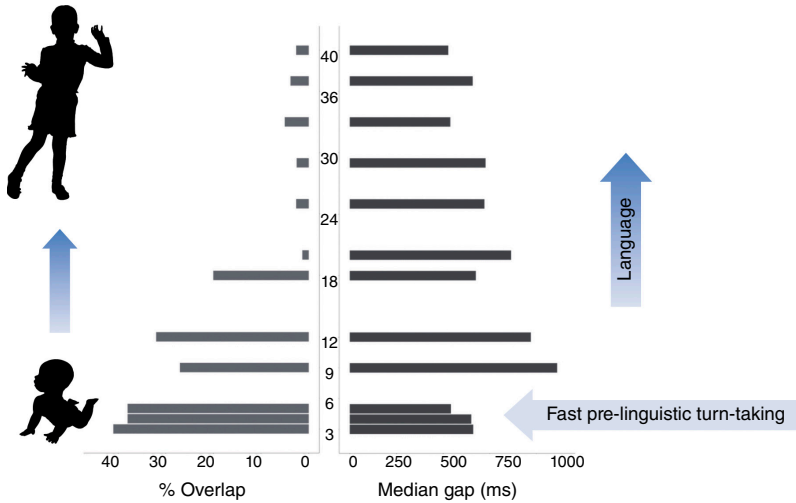

Figure 4.4 Turn-taking from infancy to childhood. Infant age in months is shown vertically, with the amount of overlap with caregivers' turns on the left, and length of gaps to the right. Pre-linguistic vocalizations early on show quick turn-taking, and the speed of response seems to slow as infants struggle with encoding language (data from Casillas & Frank 2017, Hilbrink, Gattis, & Levinson 2015).

breakthrough', as it has been called, when the infant's vocalizations suddenly become predominantly alternating with the caregiver's.[45]

The later sequence of development is interesting – at nine months the responses of the infant are slower (Figure 4.4). This is the period when infants begin to show the ability to engage in joint attention over an object, an essential stepping stone to learning words. Subsequently, as the infant progressively engages with language, turn-taking speed does not increase significantly. Toddlers can get quite frustrated by their inability to jump into an adult conversation – they are simply too slow. In fact, it takes years for children to acquire adult speed at turn-taking. A recent study, measuring answers to questions, found that while the average adult response was within 250 ms, the average 4–5 year-old's responses were around 481 ms, and 6–8 year-olds were averaging 487 ms, still far from adult speed.[46]

[45] Ginsburg & Kilbourne 1988. [46] Stivers, Sidnell, & Bergen 2018.

What these observations suggest is that, early on, infants quickly acquire (or have natively available) a relatively rapid, half-second response timing. But as language develops with ever-increasing complexity, their response timing slows, and it doesn't reach adult rapidity till late middle childhood. These findings make sense in the light of the extraordinary processing demands that turn-taking makes on our cognitive system – as detailed in Section 3.2, adults are processing the incoming speech so fast that they can predict how it will end, and can simultaneously begin preparing their responses. This is an impressive adult trick, only possible because of their speed of language processing – an acquired skill.

Let us return to the evolutionary questions. This very frequent early turn-taking does not seem to have the same intensity among the non-human apes. Although we tend to think of this as something the mother is imposing upon the infant, training it as it were for social life and language, there is an interesting alternative perspective from primatology, namely that the infant is constantly testing the mother, and has reason to do so.[47] This is because along with other species who practice 'alloparenting' (that is, the sharing of child care with non-parents), human mothers are far more likely than great ape mothers to abandon or neglect their infants.[48] Alloparenting lowers the parental investment in the child, and in times of shortage or conflict it may be strategic to cut the losses. The cuteness of babies and their natal fatness (not found in other apes), may be essential insurances against neglect (see Section 4.7).

But the particular evolutionary interest here is that the very early turn-taking exhibited by the infant may have an instinctive basis rooted in phylogeny. One possibility that the child development data suggests is that early hominins, like the apes, had a simple but effective communication system, combining vocalizations with gesture, part of the human phylogenetic background. As we've seen, ape gestures are exchanged at about the same pace that human adults exchange utterances, with quick gestures of around a second or two duration and a response timing around 250 ms. This is not cognitively taxing if there are just a few stabilized gestures to produce and comprehend. If this

[47] Hrdy 2009: 119.
[48] Hrdy 1999: chapters 12 and 14 discuss infanticide and differential neglect in simple societies.

package of one- to two-second chunks with 250 ms between alternations was part of our primate heritage, it would explain the relative rapidity of early infant turn-taking – it would be part of primate ethology. But in the human case, the child thereafter has to learn to squeeze ever-increasing linguistic complexity into the same one- or two-second chunk, and at the same time attempt to maintain a rapid alternation. Human development seems consonant so far with the phylogenetic story we have told.

The turn-taking evidence is in this way suggestive of a deep phylogenetic continuity from apes to humans. If so, one might expect human infants to gesture before learning much language. In some respects this expectation is met. There has been sustained study of index-finger pointing, and a great deal of evidence that it becomes universally available across disparate cultures from about ten months, once the nature of mutual attention over an object (triadic attention as it has been called) is well established.[49] Children of this age or soon after are able to use pointing creatively to help people find things, indicating early cooperative instincts. So deictic or pointing gestures are used to identify objects before children know their names. But the kind of gestures that adults do the whole time while they speak, so-called iconic or depictive gestures in particular, seem curiously absent (or at least non-synchronous with speech) in children until well into the second year.[50] Children seem to learn the words – mostly verbs – that go with the gestures as much as six months before they learn to make the corresponding gestures. This delayed inception of iconic or demonstrating gestures seems attenuated in the context of sound-symbolic words (like *squishy* or *thump*), suggesting that such words may have played an important role in early language.[51]

Now apes do not naturally point, but they do make suggestive gestures that seem iconic or depictive in character, so on the face of it there seems a real discontinuity here. This may perhaps be a problem for the gesture-first theory of language evolution entertained in the next section. And the fact that iconic gestures come in long after the corresponding verbs is also a problem for the idea that vocalization and gesture are part of a seamless multimodal package, and all that

[49] Liszkowski *et al.* 2012.
[50] McNeill 1992: chapter 11, Ozçalışkan, Gentner, & Goldin-Meadow 2014.
[51] Kita *et al.* 2010.

has happened in the transition from ape to human is that the burden of communication has been shifted from the gestural mode towards the oral. It is possible that the conceptual and motoric skills involved in making iconic gestures simply take longer to mature, but nevertheless are an integral part of a natural and inevitable maturation. It is after all a fact that in all languages, as far as we know, both pointing and iconic gestures routinely accompany speech unless there are specific cultural taboos against using them, suggesting a strong native basis.[52] But the pattern of development of gestures in human infants compared to apes does appear distinctive for reasons we partly do not understand.

4.5 Gesture as the Trojan Horse that Gave Language Its Propositional Structure

Every mobile animal with a home base needs to be able to find its way back.[53] Some of the most extraordinary cognitive feats by animals are navigational: the arctic terns circumnavigate the globe annually from pole to pole converging on an area north of Antarctica in the southern summer. The lesser golden plover migrates from Siberia to Tasmania and back again. As winter threatens, monarch butterflies retreat from the Great Lakes to specific fir trees in mountains west of Mexico City. Bluefin tuna circulate from arctic waters to Brazil and back via the Caribbean. Brazilian green turtles return from South America to a tiny dot in the Atlantic, Ascension Island, to breed.[54] These navigational feats are aided by exotic senses like magnetoreception, polarized light detection, or special thermoreceptive organs.[55]

By contrast, human navigational abilities are natively poor. They are poor even by contrast to our primate cousins the chimpanzees, whose territorial range is much more restricted than that of human hunter-gatherers.[56] Newspapers abound with stories of walkers lost for days, and even experienced woodsmen and naturalists get lost just a few miles from their starting point. We did not inherit that veritable bonanza of special sensory apparatus available to the birds, the fishes, and the beasts. But we have made up for it by developing technologies

[52] See, for example, Kita & Essegbey 2001.
[53] A fuller version of the argument in this section can be found in Levinson 2023.
[54] Waterman 1989. [55] Hughes 1999. [56] Green *et al*. 2020.

of navigation, and elaborate cultural solutions for spatial orientation and description. Take for example the Guugu Yimithirr speakers of northern Queensland, who have developed a very fine sense of direction by virtue of the natural training their language and gesture system gives them. The language has no words for 'left' or 'right' directions, but instead uses a cardinal direction term like our 'north', 'south', 'east', 'west' (although oriented slightly differently). Because there is no other way to say where things are, they need to be able to say things like 'Watch out, there's a snake by your northern foot'. Once when travelling off-road with Guugu Yimithirr speakers I drove into a bog, because it took me too long to process the warning given in cardinal direction terms! The verbal system is supplemented by a gesture system which gives directions accurate to within a few degrees, and which can also indicate distance by the height of the arm. Up to one in ten words in Guugu Yimithirr are cardinal direction terms, mostly supplemented or even supplanted by a gesture. These people can point with unerring accuracy to places close and far, and do not easily get disoriented even under experimental conditions.[57]

This is a cultural solution with lifelong training compensating for a missing innate ability. We have of course developed over thousands of years a plethora of cultural prostheses to compensate for our feeble native abilities, from charts and maps to theodolites and compasses, radar beacons, and GPS devices. Even the humble path or road serves a navigational function for most of us. Our nearest relatives, the relatively sedentary chimpanzees, need to have an unfailing sense of direction in a dense forest canopy if they are to retrieve tools and return to known food sources.[58] It is worth asking why we are natively such poor navigators even though we wander over such vast territories.

The mental maps of rats and other mammals are stored in the hippocampus (a limbic structure duplicated on both sides of the brain), using a system of specialized cells, the discovery of which earned John O'Keefe, May-Britt Moser, and Edvard Moser a Nobel prize in 2014. There are directional cells, boundary cells, and in addition grid cells, which record mental maps at different granularities. Humans have the same system, and experiments with London cab drivers showed that their hippocampi grew as they learned the city's warrens and mazes.[59]

[57] Haviland 1993, Levinson 1997, 2003a. [58] Normand & Boesch 2009.
[59] Maguire *et al.* 2000.

But humans have repurposed the hippocampus to do a lot of further jobs: they have retained the spatial functions in the right hippocampus, but use the left hippocampus largely for verbal and episodic memory.

O'Keefe and associates have suggested that the intricate coding of vectors in the hippocampus may be a source for linguistic structure.[60] This suggestion jibes with an old speculation, that goes back to Greek and Roman grammarians but flourished especially in the seventeenth century, that the core of language structure is based on a spatial analogy.[61] It is fairly obvious that temporal ideas are expressed spatially: nearly all English spatial prepositions are employed in time expressions (*on Wednesday*, *at noon*, *from morning to night*, *in a week* …), but they are further extended to many more abstract domains (as in *on deliberation*, *at odds with*, *from despair to elation*, *in denial*). The spatial vertical dimension lies behind many expressions of change, as in *fall sick*, *rise triumphant*, *inflation up*, *prices down*, and the vertical structures valuations as in *top quality*, *lowest calibre*, *above all*, *beneath contempt*. Spatial motion is extended to changes of state (*go to sleep*, *come to believe*, *pass from a solid to a gas through liquid form*, *went from poverty to riches*) and plays a central role in the grammar of aspect, the encoding of the internal temporal qualities of events (*he was going to tell*, *he stopped lying*, *Sue went on criticizing*, *he would come to find out*). These sorts of patterns hold not only for English, but many other languages around the world. The Classical languages with their long history proved fertile hunting grounds for nineteenth-century scholars who discovered how spatial cases and concepts are an apparently inexhaustible source of new grammatical structures. The study of this 'grammaticalization' as it is called is still a major strand of linguistic theory.

The strong version of the theory was revived in the 1970s under the rubric of 'localism',[62] with the idea that spatial expressions provide the template for grammatical notions like case relations or 'thematic roles' (that is, the roles noun phrases play with respect to their governing verbs) and thus bestow the crucial bonds between a verb and its arguments. Thematic roles like agent and patient are arguably the very core of grammar – in English an agent typically surfaces as the

[60] O'Keefe & Nadel 1978, Nadel 1991, O'Keefe 1996.
[61] Fortis 2020, Wüllner 1831.
[62] Gruber 1965, Anderson 1971, see also Lyons 1977:718–724.

subject of a sentence and the patient as the object of a transitive verb. Spatial concepts like 'go from X to Y' become generalized to possession like 'the gift went from X to Y' and on to state change like 'The light went from red to green'.[63] In this way, a spatial notion as in *to Rome* becomes a grammatical dative as in *give to Bill*, and a marking for destination like *the train leaving for London* becomes a grammatical benefactive as in *she suffered for her children*. Localism holds that all the basic grammatical relations have underlying spatial concepts.

Furthermore, space can also be thought to be the donor of many fundamental semantic concepts. The kind of semantic primitives required for language include things, places, paths (notions like *to*, *from*, *via*), events (motions), and states (things in places) – just the sorts of entities encoded by specialized cells in the hippocampus. Spatial language itself, although very various across diverse languages, draws on universal frameworks of spatial frames of reference – the different ways of anchoring things in space, via reference to an ego's point of view, an object's surfaces, or an environmental anchor.[64] The extraordinary thing about the Aboriginal language Guugu Yimithirr is that it utilizes the environmental or geographical frame to the exclusion of the others, so inculcating an unerring mental compass. The degree to which this kind of training can lead to an alternate cognitive system is shown by an experiment carried out on children from another hunter-gatherer society, the Hai||om of Namibia, who when taught to dance by demonstration (instructed 'Do it like this!'), learnt the dance moves in terms of cardinal directions, not in terms of bodily left and right. So, when facing south, and shown to start off with their right (and western) foot, they would mimic the motion; but when turned around 180 degrees, they would now lead with their left (and western foot) not the right foot![65] It is this cognitive flexibility that shows that humans do not have an innately fixed spatial orientation system, unlike the birds and the beasts.[66]

It used to be thought that the hippocampus plays only a minor role in language. A major reason was that a famous patient HM who had both hippocampi impaired by an operation to cure epilepsy was able

[63] Jackendoff 1972, Talmy 1972.
[64] Levinson 2003a, Levinson & Wilkins 2006. [65] Haun & Rapold 2009.
[66] An experiment contrasting human and ape subjects suggests that apes may think primarily environmentally like the Guugu Yimithirr or the Hai||om. See Haun *et al*. 2006.

to use language at least superficially normally. His ability to acquire new memories, however, was eradicated. Close analysis of his language shows that in fact it was not normal, especially in its tracking of referents.[67] Moreover recent studies show that the hippocampus is crucially involved in keeping track of who did what in discourse,[68] and, most remarkably, that when learning a second language the hippocampus grows just as it does when learning new spatial routes.[69] It is now thought that the hippocampus plays a crucial predictive role in language production and comprehension.[70] Intracranial electrodes inserted in the hippocampus during surgical operations show that the theta rhythm, a distinctive neural oscillation involved in spatial reasoning, also plays a crucial role in tracking verbal expectations and predictions – the sort of thing essential to being able to produce one's conversational turn on time.[71] This overlap between the spatial and language functions of this oscillatory frequency is telling – this is the typical timing of the syllable in language.

The possibility then arises that language has invaded the human hippocampus, partially capturing the left-hand one which is most closely integrated with language. Inarticulate mammals will not suffer from such a 'hit' to their spatial capacity, but humans may have a weakened sense of direction and a feebler spatial memory as a result. The interesting possibility then is that because language has cannibalized the human left hippocampus, language has inherited the spatial frameworks intrinsic to this neural tissue, thus accounting for the 'localist' observations about the spatial foundations for semantics and grammar. This kind of co-option of pre-existing brain tissue for new functions is probably what has made language possible. There is an interesting parallel in the exploitation of the left occipitotemporal sulcus (also called the visual word form area) for reading – a process that has been dubbed the 'cultural recycling of cortical maps'.[72] In that particular case, the part of the visual system evolved for recognizing small linear structures next to the face recognition area, has been repurposed for letter recognition, with a consequent apparent loss of face-recognition acuity in literate people. Because this area of the brain

[67] MacKay, Stewart, & Burke 1998, MacKay 2011.
[68] Duff & Brown-Schmidt 2012. [69] Mårtensson *et al.* 2012.
[70] van de Ven, Waldorp, & Christoffels 2020. [71] Piai *et al.* 2016.
[72] Dehaene & Cohen 2007.

is natively adapted for recognizing detailed linear structures, nearly all the writing systems of the world have this spidery character. There is thus a reciprocal effect of 'cultural recycling of cortical maps' – the brain shapes the cultural exploitation, but the cultural exploitation also reshapes the brain. There are in fact profound effects of literacy on the brain; the differences are substantial enough to form an anatomical signature of literacy.[73] The degree of flexibility or brain plasticity involved here can be gauged by the extraordinary finding that in the blind the visual areas have been re-assigned to process language.[74] Given these parallels, the idea that the left hippocampus has been recruited for language purposes, and in so doing, has left an indelibly spatial mark on language, seems plausible. Such a recycling of brain tissue for new purposes would have weakened our directional sense.

But why would the hippocampus have been recruited in this way? Here we come back to gesture. Gesture is a spatial modality, and in fact about three-quarters of gestures convey spatial information.[75] When people are describing places, objects, or directions, gestures almost invariably accompany words. Although a great deal of current talk has spatial content, in the eras before agriculture, when humans were all foragers and hunter-gatherers, spatial communication must have been of pre-eminent importance and even greater frequency. In the Australian language Guugu Yimithirr, spoken by a group that were traditionally hunter-gatherers, up to one in ten words is a directional – one of the cardinal direction terms. By locking all spatial coordinates into a fixed north-south-east-west system, a great specificity and precision can be conveyed by gesture, as illustrated in Figure 4.5. Here a Guugu Yimithirr speaker narrates how a fugitive hid inside a hollow tree, later leaping out westwards to catch the speaker as a boy. The whole narrative is coherently locked to landscape details, and can be followed on a map.[76]

Now recollect that flexible interactional communication in our nearest ape relatives, the chimpanzees and bonobos, is largely in gestural mode, and the presumption is therefore that early hominins in our line were

[73] Reading exploits brain plasticity to build greatly enhanced connections (white matter tracts) between the hemispheres (the corpus callosum), and enlarged areas of grey matter in crucial locations (Castro-Caldas *et al.* 1998, Carreiras *et al.* 2009, Dehaene 2009, Huber *et al.* 2018).
[74] Bedny, Richardson, & Saxe 2015, Bedny & MacSweeney 2019.
[75] Cooperrider, Gentner, & Goldin-Meadow 2016. [76] Levinson 2023.

Figure 4.5 The power of gesture: a Guugu Yimithirr speaker's gestures set up the story of how a man ambushed the speaker by jumping westwards out of a hollow log. In this community in northern Queensland, gestures faithfully reproduce cardinal directions of motions and alignments (Levinson 1997, 2003a).

primarily gestural communicators too, an inference strengthened by the finding that the African variety of *Homo erectus* may not have had full vocal control (Section 4.3). These early hominins were successful big game hunters, and cooperative decisions about where to hunt and directions for where to help retrieve game must have been crucial – spatial information that lends itself to gesture, itself inscribed in space.

We have seen that pointing constitutes a milestone in the development of communication in childhood, appearing universally around the first birthday, before the first words.[77] Soon after, infants use gesture in a cooperative way, for example to point to mislaid objects, or even to places recently occupied by an absent person in order to indicate that person.[78] Pointing is an incredibly powerful tool, used by the Guugu Yimithirr, for example, to indicate locations near or far to within a few degrees of arc.[79] It plays a crucial role in contact situations where there is no common language (as reviewed in Chapter 2) and in the birth of new languages, as when new sign languages evolve out of 'home sign' ad hoc manual signing systems. In the relatively new Balinese sign language Kata Kolok, for example, one sign in six in interaction is a pointing gesture.[80] Interestingly, great apes in the wild have never been observed to point, probably because they lack the cooperative instincts which make joint attention and collaboration possible.[81]

There is an additional important reason why language may have gravitated towards spatial cognition. We have seen that the design of utterances takes into account what the recipient is likely to make of them – that is, the essence of open-ended communication consists in being able to take the other's point of view. This is also of course the foundation of cooperative behaviour. But taking the other's point of view also has a quite literal interpretation, what things look like from the other's perspective. The development of this spatial perspective-taking was first explored by the great Swiss developmental psychologist Jean Piaget: he and his long-time collaborator Bärbel Inhelder presented children with a three-dimensional model of three mountains and placed a doll on the other side – when did the child, they wondered, come to be able to imagine what the scene looks like from the other side?[82] They thought on the basis of their experiments not before age six, but recent experiments with simpler scenes show children can imagine the

[77] Liszkowski *et al.* 2012. [78] Liszkowski *et al.* 2012. [79] Levinson 2003a.
[80] De Vos 2012. [81] Tomasello 2006, 2022. [82] Piaget & Inhelder 1969.

opposite view at age four or below.[83] This visual perspective-taking may play a crucial part in the growth of the child's ability to model his or her interlocutor's mental states. It is interesting to note that disruptions to 'theory of mind', as in autism, are also associated with weakened spatial abilities[84] and abnormal gestures.[85] It is also interesting that the hippocampus is involved not only in mental maps of our local terrain but also of our social life, representing close versus distant kin on one spatial dimension, and bosses and underlings on another vertical one (these social maps will play a role in Section 5.6).[86]

We have followed a trail of clues that assemble into a coherent picture of how spatial concepts may have played a crucial role in the evolution of language. Let us now spell out the argument that emerges from these observations:

1. Human native spatial abilities are poor, but we make up for it with linguistic and cultural prostheses;
2. The explanation may be that language has cannibalized the hippocampus, the seat of the mammalian mental GPS;
3. Consequently, language may have borrowed conceptual primitives from spatial cognition, these being differentially combinable in different languages;
4. The hippocampus may have been colonized because:
   (a) space was prime subject matter for communication among early hominins,
   (b) gesture uses space to represent space, and was a likely precursor to language,
   (c) perspective-taking is essential for flexible communication.
5. Spatial cognition may thus have been a pre-adaptation for linguistic concepts, providing us with some of the conceptual framework that makes it possible for us to express propositional thought in vocal form.

If the special role of gesture may explain the central role of spatial cognition in the organization of language, it will not explain why apes – who make extensive use of gesture – have not gone down the same route. Indeed, apes do not point. To understand this, we need

[83] Hughes 1975. [84] Lind, Bowler, & Raber 2014. [85] Hughes *et al*. 2019.
[86] Montagrin, Saiote, & Schiller 2018, Schafer & Schiller 2018, Tavares *et al*. 2015.

to understand how humans came to have their abiding interest in what other individuals think and particularly think of their fellows, to which we now turn.

4.6 The Route to Empathy and Theory of Mind

Recollect that the interaction engine has as a critical component the ability to think 'Why is she making that non-instrumental action or gesture now?', in other words, to think about possible communicative motives, and to distinguish them from non-communicative intents.[87] Without this, raising an index finger or making an unfamiliar vocalization cannot be recognized as communicative signals, and there would be no way for the infant to crawl its way into the communicative world of language users. The door to that world is the door into other minds.

Infants are surprised by experiments in which objects seem to escape basic laws of physics – they seem to be innately endowed with some kind of naive theory of physical phenomena, or at least they have the means to rapidly concoct it.[88] In the same sort of way they have been supposed to natively have available a theory, or the means of constructing such a theory, of other minds. The 'theory of mind' would provide for an intuitive psychology of other agents, including the attribution to other persons of (possibly fallible) beliefs and intentions driving their behaviour. Nevertheless, the infant progresses slowly, starting from birth with awareness of mutual gaze,[89] through recognition of specific others, to the full-blown realization of the potential differences in others' points of view and knowledge between the age of three and four.[90] Although it is now known that some aspects of this progression are available to apes (for example, understanding what other apes see and desire),[91] they never seem to achieve the kind of understanding of others' thoughts and intentions exhibited by a three- or four-year-old child.[92]

[87] A fuller exposition of the argument in this section, with additional references and detail, can be found in Levinson 2022b.
[88] Baillargeon 1994. [89] Farroni *et al.* 2002. [90] Wimmer & Perner 1983.
[91] Moore 2017 argues that this already provides a crucial foundation for communication, on which human infants build.
[92] Call & Tomasello 2008. This modifies the results of Premack & Woodruff 1978.

There has been a huge amount of research into how infants and children acquire theory of mind.[93] Infants seem to progress first from an understanding of others' desires, followed by an understanding of others' beliefs, then to realizing there can be limitations to others' beliefs, next to entertaining that others' may hold false beliefs, and finally to considering others' possible deceptions. The developmental timetable though can differ substantially across cultures, showing that culture and learning are involved. Most dramatically, deaf children born to hearing parents and limited to the restricted 'home sign' systems mentioned in Chapter 2, are delayed by years. While normally linguistic children (including deaf signers) achieve false-belief understanding by four or five years, deaf home signers get to the same stage at age eleven or later. Some home signers, who have invented a pidgin sign system of their own, even fail false belief tasks in adulthood.[94] This shows that theory of mind is, unlike naive physics, not something that simply unfolds during development – it requires critical input from others. But the fact that acculturated apes, even those reared by humans, never fully attain theory of mind makes clear that there is also some innate basis to it.[95]

The question then is where our theory of mind originates – and more generally what is the origin of our capacity for taking the other's point of view, and the empathy that goes with it. Empathy, then, has both a cognitive or perspectival side and an affective or emotional one. In all societies, empathy is recognized as a core value, for in all societies people (and likely many species of animal) at least grieve for their dead. Christianity turned this pathos into a central doctrine – in Catholic Christendom images of the stricken bleeding Christ awakened faith in worshippers. The weeping Mary under the cross, the arrows piercing St Sebastian, St Agatha having her breasts cut off, images of such scenes were made to evoke pity and reimagine the suffering – as Tertullian said 'the blood of the martyrs is the seed of the Church'. The invocation of empathy plays a major role in our cultural life (Figure 4.6).

In 2015, I returned to Rossel Island, a remote island in Papua New Guinea, where I had been doing fieldwork off and on for twenty

[93] For a useful review, from which the following details are taken, see Wellman 2018.
[94] Pyers & Senghas 2009. [95] But see Heyes & Frith 2014.

Figure 4.6 Empathy in social life: (a) the Madonna weeps at the crucifixion; (b) President Obama weeps over the Sandy Hook massacre; (c) a beggar invokes sympathy to beg for alms. Image credits (a) Mater Dolorosa, ascribed to Pedro Roldánc c. 1670, Bode Museum, Berlin (Photo: S. C. Levinson); (b) Getty Images (Joe Radele); (c) Myriams fotos, Pixabay

years, to find that my chief assistant and host Yidika had died. He had been invaluable to me because he alone, of all the islanders, had learnt to help me transcribe the complex local language Yélî Dnye. He had also facilitated my research in many ways, and so wherever I went, he would come too, walking around the island or travelling by canoe. As was customary, I went and wept by the grave, and indeed I felt genuine grief at the loss of a friend and companion. This was widely reported, and thereafter wherever I went around the island people would come and weep copious tears for me, in pity for my loss (there is a special verb in Yélî Dnye, *ch:anê* 'evoke pity, feel sorry for someone'). The degree of empathy seemed extraordinary to me at the time, and indeed it got in the way of work – in our own society, it would be proper to say 'I am sorry to hear of your loss' or the like, but not to burst into tears for a stranger's loss. This set me thinking about the special role of empathy in human social life: is this the lifeblood that powers the cooperation that is so distinctive of humans? Is empathy, and its cognitive counterpart, theory of mind, the human equivalent of the pheromones that glue together the societies of social insects like ants and bees? In a small, traditional, kinship-based society like Rossel Island, it seems natural that altruism and prosocial behaviour would be driven by an empathetic understanding that one should alleviate the suffering of one's own kin. Likewise, in such societies, where there is evidence that the empathetic response has been exploited or trust betrayed, accusations of sorcery or the beginnings of feud arise.

In humans, empathy really works, in the sense that sharing anguish actually diminishes it physiologically, while measurably increasing the stress level in the empathizers.[96] The mechanisms have also been examined: feeling empathy for others releases oxytocin, a hormone associated with emotional attachment, and this leads to more prosocial behaviour.[97] The greater prosociality of bonobos compared to chimpanzees has been linked to hormonal differences of this kind.[98] In the brain, the very same circuits involved in first-person pain are activated when observing others' pain.[99] Empathy also works materially: beggars get their alms, we give to charities, governments give aid to poor nations.

Do apes have empathy for their fellows? Although Darwin doubted that apes were aware of their own mortality, he studied their expressions of grief. Since then there has been much observation and research. Overall the story is mixed. On the one hand, there seems little doubt that apes grieve for their dead.[100] On the other hand, experimental investigations do not seem to show that any ape species, with the partial exception of orangutans, is especially motivated by sympathy to perform prosocial acts.[101] Nor is there evidence that non-human primates show empathy in cases other than distress.[102] In contrast, humans enjoy vicariously the successes of others, hence our enjoyment of football matches, the Oscars, or the Olympic Games.

Where then does human empathy originate, what mechanisms gave rise to it? Perhaps if we understood this we would understand the roots of human cooperation, which remains an evolutionary mystery. Evolutionary theorists recognize that community-wide altruism and cooperation are hard to explain: the normal mechanisms of evolution promote individual self-interest (recollect the theory of the 'selfish gene'). One mechanism that gets us part of the way is kin selection: in a kinship-based society, selfless behaviour in favour of individuals who share many of the same genes is a way of ensuring preservation of much of one's own genetic material – a mechanism that underlies the workers of insect societies. Some think that only 'group selection' in competitive circumstances offers an explanation, where a group with good internal

[96] Peräkylä, & Sorjonen 2012, Peräkylä *et al*. 2015. [97] Barraza & Zak 2009.
[98] Staes *et al*. 2014. [99] Bernhardt & Singer 2012.
[100] Gonçalves & Carvalho 2019. [101] Liebal *et al*. 2014.
[102] Myowa & Butler 2017.

cooperation can outperform one that lacks it, even though this mechanism has a dubious history in evolutionary theory.[103]

I will argue here that part of the answer lies in how humans have generalized parental instincts beyond the parental bond. Parental investment in offspring is of course a central topic in primatology, and a great deal is known about how varied parenting is among the different species. Chimpanzees, our closest cousins, very rarely share parenting, largely because mothers fear infanticide.[104] In contrast, human infants are typically raised by a number of adults and older siblings, made possible by the fact that human infants are weaned much earlier than chimpanzee infants (twenty-nine months on average in non-industrial societies, compared to four to six years for chimpanzees).[105] Indeed, because mothers are freed of total responsibility for infant care early, humans breed at twice the rate of any other ape. Clearly, this reliance on childcare assistance other than from parents has the consequence that the kind of private gestural communication system of specific chimpanzee mother–infant dyads would be a hindrance in a human society. Outsourcing childcare presupposes a community-wide communication system.

Those of us in the West live in societies where the nuclear family is often the norm, but in most of the world, people live in extended families or close to grandparents and other kin and rely heavily on help with infant care. When working on Rossel Island I was frequently amazed to encounter familiar infants miles from their natal village in the arms of a young niece or even being suckled by an aunt, having been lent out for entertainment or because the mother needed respite. Hunter-gatherer peoples make extensive use of these substitute parents.[106] The predominance of the nuclear family in industrialized societies blinds us to the crucial role that elder siblings, aunties, nieces, and above all grandmothers have played in the rearing of young throughout the bulk of human history. There is even a theory that the menopause evolved to allow grandmothers to indirectly contribute to the reproductive success and fitness of their descendants.[107]

Now, every mammal by virtue of its offspring's dependence on nursing has a close relationship to those offspring, and has a natural

[103] Boyd & Richerson 2009. [104] Myowa & Butler 2017.
[105] Hawkes *et al*. 2017. [106] Hrdy 2009: chapter 4.
[107] Hawkes, O'Connell, & Blurton-Jones 1997.

interest in the wellbeing and mental states of their dependents. Mothers then need to attend to infant distress, and figure out what causes it. Humans in particular have an exceptionally close mother–infant relation. Although newborn infant chimpanzees are as helpless as human ones, the type and intensity of interaction between mother and infant contrasts with the human relationship: there is less maternal looking and shorter mutual gaze in chimpanzee dyads.[108] In humans the prolonged mutual gaze, the engagement in imitative games, and the exaggerated vocal and visual displays typical of 'motherese' are a form of species-specific behaviour, and perhaps related to the fact that the infant is more often laid down, with touch replaced with distal engagement.

In this maternal relationship to the infant, empathy is guaranteed by maternal selection, evolution's winnowing out of unsuccessful mothers. It is not hard to show that mothers have a greater empathetic response than non-mothers to others in distress.[109] The evolutionary puzzle of human empathy is how and why this kind of empathy is shared outside this maternal relationship, indeed widely across conspecifics. We feel sorry for the beggar in the street, the homeless person on the street corner, the child who fell over, and we may intervene at our own cost, even risking our lives to rescue a stranger's drowning child. We feel sorry for them because we can imagine ourselves in their stead, performing the spatial and mental transpositions that Piaget studied in children. Why we would try to help a stranger even to the extent of risking our own lives is clearly an evolutionary puzzle.

So, here's a story about how we might have got there – how we might have evolved empathy and theory of mind, how we might have generalized maternal caring instincts to the population at large. Konrad Lorenz, observing his daughter with a puppy, noted that cuteness (his *herzig*) is a 'releaser' of caring and empathetic instincts.[110] 'Cuteness' has physical attributes that are hard to resist – large eyes, short snout, bulging cranium, chubby cheeks, short limbs, and clumsy movements (Figure 4.7a). Stephen J. Gould pointed out in a playful essay that our favouring of cuteness even worked to transform the images of Mickey Mouse over fifty years from a lanky teenager to an attractive juvenile. There is a curious Japanese cult of the cute, known locally as *kawaii*, expressed in dress styles, dolls, and cartoons, where the figures exhibit

[108] Bard *et al.* 2005. [109] Plank *et al.* 2021. [110] Lorenz 1943, Gould 1980.

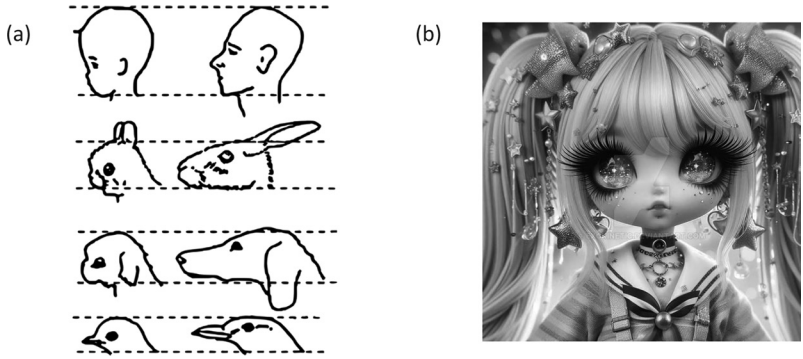

Figure 4.7 (a) Infantile features that elicit cuteness reactions in humans, including reduction of nose, globular head, and relatively large eyes (from Lorenz 1943). **(b) Doll illustration in Japanese *kawaii* or cuteness aesthetic** (*Kawaii* cute harajuku doll by Exokinetic on DeviantArt, Creative Commons license).

the large eyes, reduced jaws, and globular heads of Lorenz's generalizations (Figure 4.7b).

Lorenz's notions of instinctive reactions triggered by 'releasers' are now viewed as oversimplifications of more complex processes. For one thing, we now know much more about the underlying mechanisms – in mice, cuteness releases oxytocin, and oxytocin in turn triggers maternal behavior and response to vocal signals.[111] But an interesting link was already made by Lorenz, from cuteness releasers to neoteny, the retention of childlike features that characterizes the human species: 'The characteristic which is so vital for the human peculiarity of the true man – that of always remaining in a state of development – is quite certainly a gift which we owe to the neotenous nature of mankind.'[112]

Neoteny, or the retention of childlike physiology and behaviour, is something that seems to characterize humans, with their large heads, small jaws, reduced dentition, and the like.[113] The general slowness of human development with long gestation, prolonged childhood, late sexual maturity, and lengthened life expectancy, all seem in line with this account.

But neoteny is another concept that has met with some modern suspicion, since it is a blanket concept that can hide diverse processes. Neotenous features are better seen as a superficial byproduct of

[111] Marlin *et al*. 2015. [112] Quoted by Gould 1980:107. [113] Gould 1977.

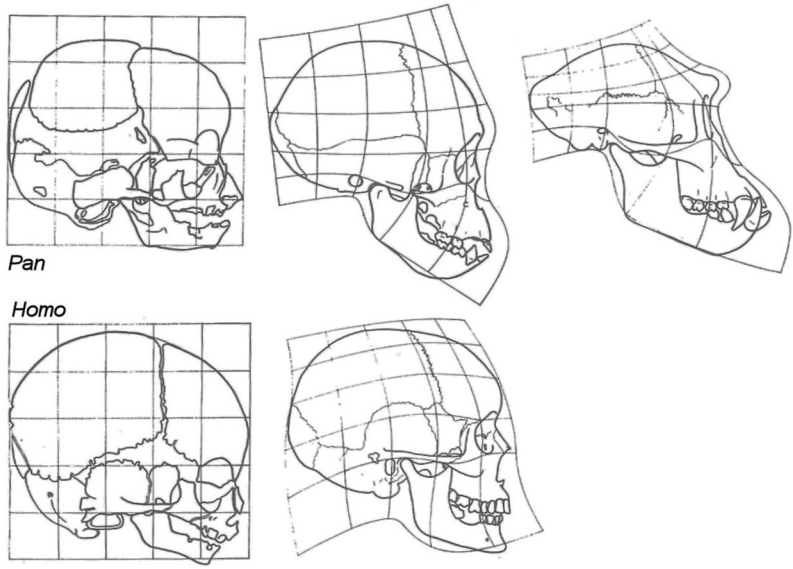

Figure 4.8 Development of the skull in chimpanzee (top) versus human (bottom). There is a striking resemblance between the shape of the skull of the young chimpanzee and the adult human, suggestive of human neoteny (image derived from Starck & Kummer 1962, partly after D'Arcy Thomson 1942).

heterochrony, a central evolutionary mechanism for creating adaptive variants simply by slowing down or speeding up or otherwise playing with the temporal development of particular organs and processes.[114] On this view, the apparently neotenous features in humans may in fact be the outcome of a complex interplay of developmental accelerations and retardations, with mismatches between, for instance, childlike dentition and adult-like long legs.[115] Take, for example, the developing skull shapes of the chimpanzee versus that of humans, as in Figure 4.8, where the adolescent chimpanzee skull looks most similar to the human adult, while the adult chimp has developed a powerful jaw with a proportionally smaller cranium. In principle this could simply be a byproduct of heterochrony, with humans frozen, as it were, at an adolescent stage.[116] But it could also be the effect of selection for reduced dentition, for example because of reduced aggression or the

[114] Gould 1977. [115] See, for example, Bogin 1997.
[116] For a sceptical view, see Shea 1989.

tenderizing of food by cooking; or because (as is currently speculated) selection of genes for cortical development effectively globularized the brain case in anatomically modern humans.[117]

Despite these caveats, Gould's conclusion that 'a general temporal retardation has clearly characterized human evolution'[118] still seems valid. A recent development has been the ability to measure neoteny in the development of the brain. For example, an analysis of some 300 genes expressed in the brain shows that twice as many of these are delayed in expression in humans compared to chimpanzees – a delay that presumably confers on us the brain plasticity instrumental to our long learning period.[119]

Now it is easy to see how some of these neotenic effects could be caused by selection for 'cuteness'. Cuteness selection would work like this: if 'cuteness' releases empathy and nurturance from caregivers, then retaining cuteness for longer will, other things being equal, raise the chances of survival. In humans, there is no rush for puberty, which may cut us off from many types of support. Cuteness will have its effect on the viewer automatically, willy-nilly, and is thus not an emotion that we associate with great art, but rather with kitsch.[120] Its automaticity is essential for the survival of young mammals. If adults are shown infant faces, they react with distinctive brain responses in the orbitofrontal cortex and they flush slightly, showing a rise in facial temperature.[121]

How could selection for cuteness get going? Suppose individual A prefers infantile-looking mate B, who produces as offspring an individual C with infantile characteristics (due of course to B's genes for infantile appearance). But C also has A's genes preferring infantile-looking mates, so will choose an infantile-looking mate D, and their joint offspring will both look infantile and prefer infantile-looking mates. And so on. This is actually the recipe for what the great mathematical geneticist R. A. Fisher called 'runaway sexual selection',[122] which is one way the peacock could have got its tail, or male gorillas their grey backs, or humans the loss of much of their body hair. The critical feature is the genetic predispositions that lock the signal to the response. If originally the peacock's longer tail was a true indicator of strength under a handicap, but now the tail is an automatic attractor

[117] Neubauer, Hublin, & Gunz 2018. [118] Gould 1977:365.
[119] Sommel *et al.* 2009. [120] Morreal 1991. [121] Kringelbach *et al.* 2016.
[122] Fisher 1930.

of mates, a positive feedback loop can result that promotes longer and longer tails even when these are detrimental to actual fitness. This runaway mechanism has usually been evoked to explain sexual dimorphism but, as Fisher acknowledged, it could be responsible for many traits unlinked to secondary sexual characteristics.

There might be various reasons why individual A might look for infantile characteristics in a mate. If A is male, then he could be targeting a maximally fertile female (human female fertility declines in the twenties, and humans have a uniquely extended postmenopausal life expectancy). If B is female, she could be targeting a less muscular male whose lower testosterone predicts lower aggression and a more consensual partnership. This is probably the mechanism which has reduced aggression among bonobos compared to chimpanzees. If we change the formula slightly and have A as a parent preferring a cute-looking infant over a less cute one, so investing more in cuter offspring, we will obtain very much the same effect, namely a trend towards more gracile, infantile-looking individuals. In historical populations short of resources, such differential investment in offspring can amount to infanticide or at least diminished chances to reproduce, so giving cuteness a fitness advantage.[123]

Fisher's point was that, once the preference for the *signal* of greater fitness – here cuteness – was fixed, rather than the preference for fitness itself, such a process could lead to an ever stronger and stronger signal, until limiting factors curbed the runaway process. This offered an account of the peacock's tail, which had so troubled Darwin – as he confided in his correspondence, 'The sight of a feather in a peacock's tail, whenever I gaze at it, makes me sick.'[124] Cuteness is indeed just such an irresistible signal – while doing fieldwork among Mayan Indians, our son could not resist playing with the puppies even though he suffered terribly from the resultant fleas.

There are then various ways in which this neotenic preference could confer a fitness advantage and subsequently spread through a population. A runaway process like this presupposes that some preference for cuteness already exists. On Lorenz's account, the preference might be expected to be general across mammals, as a mechanism binding mothers and infants together. In the runaway cuteness selection model just outlined, the attraction felt for infants could have been exploited by sexual selection, now switched to mate attractiveness. There is indeed

[123] See, for example, Voland 1984. [124] In a letter to Asa Gray, 3 April 1860.

some cross-cultural evidence for the perception of gracile or infantile features as attractive, especially to males.[125] A potential problem with the sexual selection account is that females in human traditional societies almost always successfully reproduce, so an account focused directly on the attractiveness of offspring may be more successful. Here, another critical feature of humans is relevant: unlike chimpanzees, humans utilize extensive alloparenting (caregiving other than by the mother) – with the early weaning mentioned previously, grandparents, older siblings and fathers can relieve the mother, thus allowing her rapid successful further reproduction. Alloparenting will only work if persons other than the mother find the infant attractive and compelling. This provides the impetus for the generalization of susceptibility to cuteness. Interestingly, alloparenting lowers the mother's investment in her offspring, making abandonment in times of difficulty (famine, loss of old partner) more thinkable. The primatologist Sarah Blaffer Hrdy shows that alloparenting species are indeed much more likely to abandon offspring. In this context of maternal ambivalence, the cuteness of offspring may be essential for survival.[126]

The relevance of this discussion for the evolution of communication is this. Cuteness features, which can include high-pitched vocalizations, novel gestures, and other motoric elements, invoke caring behaviours. Caring behaviours include understanding sources of comfort and discomfort for the infant, amusing and distracting infants, and foreseeing possible accidents. These involve 'theory of mind', attributing to the infant needs, wants, sources of emotional distress or satisfaction, and foreseeing future possible actions of the infant. Turn-taking provides a means of providing care: if the infant enjoys jiggling, or cooing, alternating behaviours are likely to arise. In this way, the reciprocal development of cuteness releasers and caring behaviours could provide a basis for the development of the interaction engine.

We noted earlier that the great apes are gesture turn-takers on a rapid basis similar to human vocal turn-takers. But this behaviour has been reported especially for mother–infant dyads, although it occurs also in other asymmetric relations, for example when adult orangutans

[125] Jones 1995. An alternative theory is that 'neoteny' results simply from the relaxation of the selection for robustness (see Brace's comments on Jones's article, 736–737).

[126] Hrdy 2009, Hrdy & Burkard 2020.

beg and give food to each other.[127] So the suggestion here is that what humans have uniquely done is generalize the empathetic properties of mother–infant interaction to the society at large.

Incidentally, a rival hypothesis to this goes under the rubric of human 'self-domestication'. Darwin considered the possibility that humans had, so to speak, tamed themselves, but could find no case where conscious breeding of humans convincingly took place (although, it now transpires, some slave owners in pre-Civil-War southern states in the USA did do so). But the concept is loosely applied to mean that women selected mates for non-aggressive tendencies, and in so doing brought into the species a slew of other traits associated with reduced aggression and with domesticated species – feminization, reduced dentition, retention of juvenile traits, greater sensitivity to other species – since these seem to automatically go along with domestication.[128] This rival account is targeted primarily at aggression-reduction, while the cuteness selection explanation has lengthened childhood, gracile build, empathy, and cooperation as the direct targets of selection. In many respects the two explanations are on the same territory, but cuteness selection has some advantages. One particular typical trait of domestication is reduced brain size (35 per cent smaller in domestic pigs), and this is absolutely not a characteristic of human evolution. In fact, the self-domestication theory has been especially invoked to explain the post-Neanderthal globularization of the brain case, in the context of a slight reduction of brain size in modern humans – but this difference in size disappears if proper allowance is made for the greater bodyweight of Neanderthals.[129] A second problem is that humans are in fact one of the most aggressive species on the planet; it is intra-group aggression that is reduced. More specifically, if one makes the distinction between 'hot' reactive aggression and 'cold' proactive aggression, it is only the hot aggression that is reduced in humans compared to chimpanzees.[130] The primatologist Sarah Blaffer Hrdy asks us to imagine travelling with a plane-load of chimps: 'Any one of us would be lucky to disembark with all ten fingers and toes still attached, with

[127] Rossano & Liebal 2014.
[128] Wilkins, Wrangham, & Fitch 2014 offer a possible mechanism. See also Hare, Wobber, & Wrangham 2012.
[129] Hare 2017 fails to take body-size changes into account.
[130] Wrangham 2018, 2019.

the baby still breathing and unmaimed. Bloody earlobes and other appendages would litter the aisles.'[131] Richard Wrangham argues the best way to explain this is that superior human communication made it possible to form coalitions against bullies and despots, so obtaining the levelling that characterizes hunter-gatherer social systems. That better communication is precisely what a more developed theory of mind delivers.

One of the effects of cuteness mentioned above is that it releases oxytocin in the recipients, a hormone that plays a critical role in many physiological processes. Oxytocin is upregulated during pregnancy and lactation. The role of oxytocin in animal and human bonding has been much studied: it reduces fear, enhances trust, and promotes prosocial behaviour.[132] Repeatedly it has been shown to be involved in increased empathy and gaze at interlocutors' faces.[133] Oxytocin, when administered to healthy subjects, increases their ability to tailor messages for particular recipients, so enhancing communicative effectiveness.[134] In genetic conditions linked to social and communicative problems, like Prader Willi syndrome and Autism Spectrum Disorder, oxytocin levels are reduced.[135] Different oxytocin levels are associated with the greater social bonding and cooperation in bonobos compared to chimpanzees.[136] Oxytocin and other endogenous opioids seem to be upregulated in humans compared to apes.[137] In short, oxytocin is the proximal enhancer of social cooperation and effective communication. It is just one of many hormones that respond to and regulate our social interaction, and make possible the high levels of cooperation in human social life – if ant social life is regulated by pheromones, we too operate what might be called a system of 'chemical amity'.

To summarize this section, we have suggested that empathy and its cognitive counterpart theory of mind play a prominent role in human social life, and were crucial to the evolution of a communication system based on recovering the sender's frame of mind. The hypothesis is that this sensitivity to others' mental states may have evolved by generalizing maternal caring instincts, through a process of runaway 'cuteness selection', so that the general population became more sensitive to these mental states (a 'maternalization' as it were of all adults). That

[131] Hrdy 2009:3. [132] Israel *et al.* 2009. [133] Jiao *et al.* 2020.
[134] De Boer *et al.* 2017. [135] Camerino 2020. [136] Staes *et al.* 2014.
[137] Rockman *et al.* 2005.

process was bound up with changes in the appearance and behaviour most likely to evoke those sensitivities (an 'infantilization' of adults as well as children). These processes would have made alloparenting a general viable strategy – non-mothers having the empathetic response to the cuteness features of someone else's infant – so allowing humans to reproduce at double the rate of the other great apes. Hence, the argument goes, we evolved both the reduced intra-group aggression and the general tendency to neoteny that characterize human evolution. These processes would have enhanced the conditions for cooperation and joint action, and made possible the inferential basis for the communication systems we call languages. These enhancements would themselves have offered further distinct fitness advantages through group activities, group competition, and survival, so cementing cooperation as a viable default mechanism.

A society of trust and empathy is a potential group of 'suckers', always subject to exploiters or free riders. Generalized trust would always have to be balanced by suspicion, punishment, and repression. Babel, the fractionation of languages, has always offered some hard-to-fake guarantee of in-group membership, while gossip and reported reputation may have helped to guard against internal exploitation. So, for example, on Rossel Island in Papua New Guinea, the prescribed amity of kinship is hedged by the constant possibility of suspected sorcery and feud. It is an interesting thought that perhaps we owe our fatal failure to evolve stable beneficial political systems to the side-effects of human generalized trust with its inevitable counterpoint, the suspicion that within our midst there lurk fellow citizens abusing and exploiting that trust.

4.7 The Possible Origins of Grammar in Interaction

This chapter began by examining aspects of the interaction engine that have clear precursors elsewhere in the primate order, and especially among the other great apes. But language presupposes a huge amount of structure beyond the turn-taking and action sequences we can see in chimpanzees or orangutans. First, there is semantic structure, and we've sketched reasons for thinking some of this may have originated in a gestural protolanguage which naturally drew on spatial concepts, which in turn seem to act as a template for grammatical relations in modern languages. So, a highly evolved gestural language of the kind that early *Homo erectus* may have had could have seen the development of

the basic propositional structure that allows the description of diverse states of affairs, helping to structure the gestural code. Another divide between us and the other great apes is theory of mind, the ability to model other individuals' thought processes to a high degree. The origin of that, we've speculated, could lie in outsourcing childcare, which requires a generalization of quasi-maternal instincts across all possible carers. It also motivates a community-wide, flexible communication system, which then makes available the many cooperative patterns of activity that constitute a culture. But what about *grammar*, that highly articulated skeletal structure of a sentence that allows movement or substitutions in some directions and not others in a way unique to every language? Some essential elements – for example a fixed word order – are already visible in the gestural communications of deaf 'home signers'.[138] But what is the source of all that amazing grammatical complexity that makes learning a foreign language such a formidable task for the adult?

Grammar has often been thought of as *sui generis*, something that evolved mysteriously and perhaps by chance out of properties of the human mind – the position that the linguist Noam Chomsky has held.[139] On Chomsky's view, the critical element in this mental revolution is recursion, but we have already seen how this may actually have its roots in interactional sequence structure (Section 3.5). The more traditional alternative view is that complex syntax comes by hard graft, being learnt late by children and often arising out of bookish learning.[140] But if, as this book maintains, the main and original functions of language lie in social interaction, one might expect the structures of languages to wear that interactional origin on their sleeves. There are, however, reasons why that may be less than obvious: first, as we've argued, human communication skills have been sedimented over aeons,[141] many of these layers have a partially independent character, and the grammatical layer is one of these. Second, our insight has been blunted by centuries of grammatical scholarship that has ignored the contexts of use.

[138] Goldin-Meadow *et al*. 2008 claim 'home sign' systems and pantomime tends to have an SOV (subject-object-verb) order, while established sign languages tend to be SOV or SVO (Napoli & Sutton-Spence 2014).
[139] Berwick & Chomsky 2016.
[140] Karlsson 2007 shows how grammatical complexity, specifically depth of embedding, is partially a function of literacy.
[141] See Levinson & Holler 2014.

Here we can do no more than try and make the case that linguistic structure is indeed adapted to its primary use in conversation.[142] A prominent property of that niche is the turn-taking we have reviewed. The turn-taking system works by allocating minimal units on a first-come, first-served basis. If one looks at the units, they turn out to be typically particles like *Yes*, *No*, *Huh?*, nouns, or question words like *Why?*, *When?*, or noun phrases like *Sue, the delivery man*, or the minimal sentence or clause, forms like *He's gone?*; *She ate it*; *She gave it away to a student*. Of these units, theories of grammars highlight the basic clause, because it both expresses a proposition and articulates the basic machinery that binds verbs to their arguments or nouns. It is the core of bookish grammars. It also plays a central role in conversation, since the first parts of adjacency pairs are (leaving aside ritual things like greetings) normally in this form – questions, requests, offers, and the like.[143] The responses, in contrast, can often be truncated. The basic English clause with five words or so fits nicely into the average turn length, which is around two seconds. If the clause is a pan-linguistic structural unit motivated by its role as the first part of adjacency pairs, the noun phrase may be motivated by its frequent role as a second part (as in 'Who came to the door?', 'John's mum').

But conversation analysts have pointed out that the turn is a porous unit, jointly constructed with the addressee. Take the following example, where Pam starts off agreeing with the prior speaker with the (particle-prefaced) basic clause *You've got to tell Mike that*, understandably complete in context and delivered with final intonation. But getting no response, she continues with a subordinate clause *because they want that on film*. A complex sentence is interactionally constructed within a turn which has been extended (notice that Carney comes in a bit late in overlap, marked with the brace).

<13> (simplified from Schegloff 1996:59)

Pam: (in breath) Oh yeah you've gotta tell Mike that. Uh- cuz they want that on film.
[
Carney: Oh: no: here we go ag(h)(h)ain …

[142] For excellent book-length treatments see Ochs, Schegloff, & Thompson 1996, Couper-Kuhlen & Selting 2018.
[143] Thompson & Couper-Kuhlen 2005.

There is an argument that a great deal of complex syntax is custom made for and by interaction. For example, grammarians call a sentence structure like the following which introduces a referent right at the beginning a 'left-dislocation': *The last paragraph, I seem to remember it being different*. But this actually arose in the following context:

<14> (from Geluykens 1992:24)
A: The last paragraph
B: yes
[
A: em, I seem to remember it being different from what's printed…

The point is that the new referent *The last paragraph* has been fronted to see if the referent can be recognized, and on assent by B the rest of the sentence is delivered. This turns out to be the primary niche for this construction, its likely origin, and the whole construction is thus typically jointly made.[144] A similar story holds for another construction in the grammar books, 'right-dislocation', as in *He's an odd man, that professor*. These are often occasioned by a lack of recognition of the referent, signified by a short pause, so a fuller or additional description is supplied:

<15> (from Stivers, Enfield, & Levinson 2007:91, simplified)
L: Your friend 'n mine was there
(0.2)
L: Mister R
J: Oh he's …

But this repair structure is now conventionalized, so it can be used in contexts where it can add no information, as in *It's a bit of a weighty subject, that*.

With a long-established language it is often hard to discern how conversational context may have contributed to the origin of a construction because written records are usually not conversational in character. But with a language even now in formation it is easier to spot the conversational origins. A nice case of this is the origin of relative clauses in the New Guinea pidgin language Tok Pisin. The construction bears the imprint of its origin with the relative clause marker *ia* derived from English 'here', which became a deictic or

[144] Geluykens 1992.

demonstrative meaning 'this one here'. So, the following structure seems to have arisen from an interactive exchange something like 'I saw this dog, you know the one with black ears' – 'Yes' – 'I was afraid of that one':

<16> (from Sankoff & Brown 1976:655)
Me save wanpela dok ia [ya bilong en blekpela] na me save fret long en
'I saw this dog [which has black ears] and I was afraid of it'

Relative clauses are the main source of centre-embeddings in language, and centre-embeddings are the best evidence for complex recursion in grammar. But as we have seen, far deeper recursion is actually found in interaction structure than in grammar, and it is interesting to see that this itself may be the source of centre-embeddings in syntax.

Complex bipartite constructions often invite collaboration in utterance construction, as in the following where a preliminary *when*-clause allows the main clause to be guessed and pre-empted by the interlocutor:

<17> Lerner 1996:241
Marty: Now most machines don't record that slow. So I'd wanna- when I make a tape,
John: be able to speed it up
Marty: Yeah

We have seen that in order to participate in rapid turn-taking it is necessary to be able to predict how the incoming turn will end – Example 17, for instance, demonstrates that this kind of prediction can indeed be done by participants. What one might expect is that languages would try to concentrate crucial indicators of the social action being done at the front end of turns, because that is the information that the responder needs in order to plan their response. We see this for example in English question marking, where the *Wh*-words are fronted to the beginning of a sentence, and Yes-No questions have an initial inversion making them early detectable. Similarly, English imperatives leave off the subject so the bare verb marking the imperative can appear right at the front (as in *Leave that right there!*). But not all languages are so obliging – many languages leave the *Wh*-words in the normal sentence position (so instead of English *Where did he say he was going to?* one is likely to get something like 'He said he is going to where?'). But in these sorts of cases there are likely to be early clues

to the function of the utterance. For example, in the many languages which do not mark Yes-No questions grammatically, or do so at the very end, there is likely to be a marked pitch offset at the very beginning of the turn.[145]

Questions in fact are a kind of construction which has played quite a central role in arguments about the nature of grammar, especially the Wh- or content question. Take a sentence like *You used to play what with her?* It only occurs naturally as a repair initiator, as for example in:

<18> (From *The Catcher in the Rye* by J. D. Salinger, 1951[146])
Holden: 'I used to play checkers with her all the time.'
Stradlater: 'You used to play what with her all the time?'
Holden: 'Checkers.'

Similarly for questions with multiple *Wh*-words – like *To whom did you give what?* – which grammarians have puzzled over because of constraints on which Wh-word can end up in front: these only have uses in repair sequences.

The mechanisms of repair quite systematically constrain the nature of language syntax. We've seen that participants, when they can't understand what has been said, try to minimize the effort required by the speaker of the problematic utterance to repair it. To do so they must locate the problem, and so for example repeat all but the missing or incomprehensible bit:

<19> (from Kendrick 2015:170)
Kel: 'But like the only picture other people f- (0.2) can see is like the one of me on the bridge with my hair like ((whoosh sound))'
(0.9)
Hea: 'What one on the br[idge.'
Kel: [In Newcastle

This makes use of the chunking provided by grammatical units, but it also motivates them: an essential ingredient for inviting repair is being able to substitute a *Wh*-phrase for a larger unit. In addition, people repair their own utterances when they have misfired – and this also requires tracking units, because speakers must make clear how

[145] Sicoli *et al*. 2015.
[146] I owe this example to the blog www.thoughtco.com/echo-question-language

much of what they said is being jettisoned. So one gets the following kind of self-repair (where the abandoned element is marked with a dash, signifying a cut-off, a glottal stop): *And from green left to pink – er from blue left to pink* where the whole constituent or unit is replaced.[147] Repairs may be repeated, as in Example 20, and it then becomes imperative that the hearer can track back and mentally discard the rejected parts, which requires the speaker to go back to the beginning of a chunk.

<20> (from Fox, Hayashi, & Jasperson 1996:206)
K: 'Okay, let's see if– before I go and look at the solution if I can–'
C: 'Mhm'
K: 'follo- if I can break it out here'

The general point is that the repair system forces a flexible chunking or constituent structure, which is exactly what the grammar has evolved to provide.[148] Repair may then have played a role in the development of complex syntax, motivating the ability to move, replace, and draw attention to specific chunks of message.

One area where interaction organization imposes strongly on grammars is the linguistic format of speech acts – the social actions performed by language, like questioning, requesting, threatening, promising, and the like. It is clear that the grammatical devices of question formation – in English the use of a fronted *Wh*-word, or the inversion of the subject and auxiliary in yes-no questions like *Is he coming?* – constrain the uses to which such utterances are put, although questioning is in fact only one of them. This is a topic already taken up in Chapter 3 and we return to it again under the rubric of politeness in Chapter 5. Some formats are designed for very specific interactional uses. Take for example third-person imperatives that many languages have, translating as something like 'Let him come here': this presupposes two speech events, one in which I tell you 'Let him come here' and another in which you tell him 'You are to go there'.

In Section 4.5, it was pointed out that spatial cognition requires many of the semantic concepts that play a role in propositional structure, namely our ability to represent the world: spatial thinking involves

[147] Levelt 1989:478ff suggests that the replacements follow the rules of coordination. See also Couper-Kuhlen & Selting 2018:130ff.
[148] Schegloff 1979.

entities moving in directions from locations to other locations, and so on.[149] So much of the propositional structure of language may have been borrowed from the antecedent conceptual resources of spatial wayfinding or gestural depiction. But the instantiation of those concepts in grammar, it is here being argued, may also owe a lot to the way in which conversational routines encourage the use of standardized expressions or markers. Initially these may be spread over two parties, and then get truncated into a single turn or exchange which retains some of the earlier structure. A simple example of this are some of the so-called indirect requests of English, with formats like *Can I have*, *Do you have*, *I wonder if*. They occur in extended sequences such as:

<21> (from Merritt 1976:325)
Customer: 'Do you have Marlboros?'
Seller: 'Yeah. Hard or soft?'
Customer: 'Soft please'
Seller: 'Okay' ((provides))

But they also occur in the truncated form, where the question is patently a request:

<22> (from Levinson 1983:361)
Customer: 'Have you got Embassy Gold please?'
Seller: 'Yes dear' ((provides))

In this way the *Have you got* becomes a standard request form in service encounters regardless of doubts about the availability of the goods.

Languages have an incredible delicacy of expression. Consider all the different ways of asking a question:

<23> a. Someone called last night, did they?
b. Someone called last night, didn't they?
c. Didn't someone call last night?
d. Someone called last night?
e. Did someone call last night?
f. Did anyone call last night?
g. No-one called last night, did they?
h. No-one called last night I suppose?
i. No-one called last night, right?

[149] Jackendoff 1983.

These are roughly organized from top to bottom in the order of expectancy: a. expects a positive answer 'Yes', with declining expectations through e.; the tide then turns towards greater expectation of a 'No' from f. through i. There are further gradations made available through intonation.[150] Other languages may do it differently, with a beautiful palette of particles with different forces. Why do grammars provide such a smorgasbord of choice between fine discriminations of expectancy? It turns out that expectancy really matters in conversation. Conversationalists try to avoid asking questions that may not have a known answer; and they try to minimize the 'epistemic gap' between speaker and addressee by carefully estimating the probabilities of the response type and choosing a matching form. Conversation analysts have noted this desired alignment of expectations across participants under the rubrics of 'preference organization' and 'epistemics' and it seems to hold across languages. The motivations seem to be, firstly, to minimize disruption of the ongoing central topics of conversation, secondly to emphasize the shared mental world of speaker and addressee (and thus agreement about what is known), and thirdly to minimize any challenge to the other's competency or 'face' – so in Example 23, it might be rude to presume that the addressee has likely failed to inform the speaker of a caller trying to reach them. The disruption occasioned by getting the estimation wrong can be seen in the following exchange:

<24> JS:II:48 (Pomerantz 1984:77)
01 A: 'D'they have a good cook there?'
02 (1.7)
03 A: 'Nothing special?'
04 B: 'No. Every- everybody takes their turns.'

Here A's question goes unanswered for nearly two seconds – long enough for A to suspect a misfire and offer an answer of her own, B's silence being interpreted apparently as a reluctance to contradict the positive expectation of the question.

But whatever the precise motivations, the evidence is that getting the polarity of the question right makes a big difference: roughly three-quarters of all answers to polar questions are affirmative in a sample of ten mostly unrelated languages, and affirmations are on average up

[150] See Quirk *et al.* 1972:807–824.

to 500 ms faster than negative answers.[151] So, conversation will only proceed smoothly if these probability estimations are mostly correct. This provides a powerful motivation for languages to develop these finely graded estimations of a likely response.

There are vast numbers of constructions in any language's grammar (recent grammars of English have nearly 2,000 pages). They have all originated from shared patterns of use, mostly patterns forged in conversational interaction under the further constraints and biases of our cognition (they have to be learnable, even 'catchy' if they are to spread throughout a speech community). Grammars are, as it were, quite largely repositories of frozen conversational strategies. Being able to adjust the prominence of information, or supplement it on the fly, produces a constant trickle of innovations. The argument, then, is that grammars are adapted to their conversational niche, which has partly forced upon them the properties they have.

To summarize, the gestural origins of language may already have donated simple structure (partly derived from spatial cognition) to language. The source of the rest of the highly complex grammatical machinery is heavily contested by linguists, some thinking that there is a specific innate endowment that constrains possible grammars, others thinking that the complex web of grammar is spun by cultural evolution under constraints that come from general (including spatial) cognition. What one can clearly see though is that much of the complexity is motivated by interactional needs. It is possible then that the grammars of languages may be built to a large extent by the sedimentation of conversational practices over deep time, in a process linguists call 'grammaticalization'. The very basic units, like the clause and the noun phrase, perform fundamental functions as major types of conversational turns, while more complex constructions arise partly through the systematic truncation of interactional routines. The repair system forces the segmentation of a turn, so that the parts can be recycled in repair initiation (as in 'You met who at the station?'). Since the turn-taking system limits speakers to one unit or clause at a time, and at the same time requires prediction of how a turn will end, this motivates the linguistic elaboration of dependencies – for example, given an *if*, the recipient is warned that a second *then* clause can be expected before the turn is finished. Since the prime and original job of language

[151] Stivers *et al*. 2009.

is to fill the turns that we exchange in informal talk, it is not surprising that languages have evolved over time to fit this function. It is only surprising that, with the exception of the school called 'interactional linguistics',[152] linguists have tended to neglect this basic fact.

4.8 Summary: The Role of Interactional Abilities in the Evolution of Language

This chapter has argued that interactional patterns offer a clear bridge between human and other primate communication systems, a bridge that is missing in the gap between the expressive capacities of language and primate call and gesture systems. We went on to see that this bridge gives us insights into the evolution of language, for example in the persistence of turn-taking timing over great ape species, including ourselves. This is clearest in the gesture capacity of the other great apes, and this led us to focus on the possible role of gesture in language evolution. Following that trail suggests that gesture, through its spatial basis, may have donated a great deal of conceptual structure to language, inherited by spoken languages to this day along with the associated gestures. One element of our interactional system though provides a gulf between man and beast, namely our interest in other minds and our ability to 'read' the intentions and emotions of others. We've speculated that one plausible avenue leading to our so-called theory of mind is through our outsourcing of childcare, which in giving us a reproductive advantage was a key to our demographic success. The transfer of precious offspring to others requires trust that the carers will adopt a maternal perspective, caring for the wants and needs of the infant. In contrast, no chimp can trust another with its infant. The consequent generalization of empathy – including the ability to take the other's point of view – through human populations opened up the possibility of Gricean communication, namely the development of signals that indicate communicative intents.

This chapter has highlighted the contributions of the interaction engine to language evolution, and even, as sketched in Section 4.7, to the grammars of current languages. There is no doubt that the evolution of language has been a key factor in the development of almost everything that marks us out as a species, but despite a great deal

[152] See, for example, Couper-Kuhlen & Selting 2018.

of recent work its development remains largely shrouded in mystery. Here we have picked out three lines of investigation that seem promising: very specific properties of interactional behaviour like turn-taking that seem highly conserved across primate species; the possible midwife function of gesture, another feature conserved across the great apes; and the crucial role that theory of mind uniquely plays in human communication. The first of these had been ignored until recently, but is now a lively focus of research especially by primatologists. Most authors on language evolution acknowledge the special role of theory of mind (including the linguist Jim Hurford,[153] or the philosophers Kim Sterelny and Ronald Planer[154]), and many think it is the key factor (following the psychologist Michael Tomasello).[155] Others though downplay it, arguing that it is the powerful symbolic code of language that is itself largely responsible for theory of mind. So, they focus instead on the symbolic nature of language or the structured nature of the signal (like the archaeologist Steven Mithin,[156] or the linguists Derek Bickerton[157] and Noam Chomsky[158]).

But the role of gesture in language evolution is particularly controversial, because it faces an obvious difficulty. Theories that suggest, as in this book, that a gestural language may have preceded spoken language, must then explain why we ever abandoned it as the central medium. The challenge is that the sign languages of the deaf demonstrate that sign languages can communicate just as effectively as spoken languages, and that once evolution finds an adequate solution, there will be no reason to abandon it: the small steps taken by the 'blind watchmaker' offer no mechanism to get from one fitness peak to another discontinuous one that is not appreciably higher.[159] How then to square the salient facts that all the other great apes have a gestural means of communication for their flexible social interaction, and that we retain an almost obligate use of gesture when speaking? One solution is to appreciate that all along primate communication has been multimodal, using both the gestural and vocal channels; what then has happened in human communication is that the burden of communication has been shifted increasingly from the gestural into the verbal

[153] Hurford 2014. [154] Planer & Sterelny 2021. [155] Tomasello 2008.
[156] Mithin 2024. [157] Bickerton 2014. [158] Berwick & Chomsky 2016.
[159] An argument cogently made by the sign language expert Emmorey 2005, among others. See also Fitch 2010.

channel. For human communication not to have got stuck in the gestural channel, the centre of gravity must have been shifted into the verbal channel before the gestural proto-language achieved anything like the expressive power of current sign languages, so perhaps over three-quarters of a million years ago.

What this book adds to the lively ongoing debate about the origins of language is an emphasis on the role that the interaction engine, a bundle of special abilities and behavioural propensities, likely played in the early steps out of great ape communication systems into the complexities of language. That a system forged so deep in antiquity could continue to have such a profound effect on how we converse and through that, on the structure of our languages, should be intriguing. But its effects are also felt in the conduct of our social life, to which we turn in Chapter 5.