# 14 Interactional Foundations of Language: The Interaction Engine Hypothesis

STEPHEN C. LEVINSON

## *1. Communication without Language*

This chapter argues that humans are endowed with a special interactive ability. It is this ability that underlies most of our species' great achievements: our language, our ability to cooperate, and hence our capacity as a culture-bearing species. In particular, I argue that it is not language that makes human communication possible, but rather a special underlying communicative ability that makes language possible. Despite the fact that this view runs against the orthodoxy, there are plenty of reasons to believe in its veracity.

For a start, there are humans born who are cut off from the normal means to learn language—they are profoundly deaf, and unless they have the luck to be born in a culture where there is a developed sign language, or where cochlear implants are mandatory, they are forced to invent from scratch some kind of sign language to communicate with those who look after them (Goldin-Meadow, 2003). Such "homesign" systems have been described from around the world (Van Deusen-Phillips, Goldin-Meadow, & Miller, 2001), and have certain features in common: lexical signs and simple sentence structures tend to emerge, despite great difference in the kind of feedback such children get. Such homesign systems have formed the basis out of which more developed sign languages have sometimes emerged (Senghas, Kita, & Özyürek, 2004).

That we have the means to communicate effectively without language is also clear from encounters with other cultures—every early explorer (and even adventurous tourist) has had some experience of this. In 1930, the Leahy brothers penetrated the central highlands of New Guinea, at that time thought to be uninhabited—their first encounters with the thriving indigenous population and their gestural communications were captured on film (see *First Contact* [Connolly & Anderson, 1987], a film incorporating the early footage) and demonstrate the effectiveness of gesture and pantomime under potentially hostile conditions.

Additional evidence comes from interactions with prelinguistic children; for example, one-year-olds can communicate about absent entities or indicate a searched-for item, by pointing to the item or the place where it was (Liszkowski, 2013). Reciprocal interactions develop with infants from at least three months after birth, and by nine months, there is little doubt that a dawning understanding of communicative intentions is developing. This is discussed at length in section 4.

Various language impairments are also revealing. Severely aphasic patients can communicate remarkably well in their home environments with limited resources—with little more than assent and dissent, complex communication can be sustained (Goodwin, 1995). Children with severe speech impediments, such as some Down's syndrome individuals, may be fluid communicators, while others with intact language, such as high-functioning autistic children, may have intact language but lack interactive intuitions and motivations and so be unable to use it effectively (a double dissociation we return to in the conclusions). Understanding precisely what such autistic individuals may lack in the way of interactional abilities may ultimately offer the key to the topic of this chapter (Frith, 2003).

Finally there is experimental evidence about our abilities to communicate without language, or even without any communicative conventions. Galantucci (2005) showed that when players of a coordination game, separated in space, are deprived of an easy means of linguistic or symbolic communication, they can invent an arbitrary symbol system and coordinate on its use. De Ruiter et al. (2010) used a similar method to show that coordination and communication was possible even when participants were deprived of the chance to build conventional symbols by rapid changes in the stimuli. The team went on to show that the sender and receiver during such exchanges activated overlapping areas of the brain, especially the right posterior superior temporal sulcus, a region previously associated with the attribution of intention (Noordzij et al., 2009).

The idea that communication does not rely on conventional symbols, but rather that such symbols rely on an underlying communicative ability, can be traced to the philosopher H. P. Grice (1957, 1989). His analysis suggested that all meaning is ultimately related to intention recognition:

S $means_{NN}$ z by U iff:
1. S intends U to cause z in H
2. S intends 1 to be achieved by H recognizing 1.

Here $meaning_{NN}$ stands for *nonnatural or intentional meaning*, sometimes talked about in terms of M-intentions (meaning intentions). So if I want to silently communicate to you, across the table in the seminar room, that you have a cappuccino moustache, I could vigorously wipe my upper lip: my intention is to cause a belief in you (that you have some foodstuff on your upper lip) by means of your recognizing that intention, and so coming to believe that you have some foodstuff on your upper lip! Why else would I be rubbing my lip so vigorously while gazing at you? Grice, and subsequent commentators, have argued that this special reflexive intention recognition could underlie complex symbol systems: once you have solved one of these joint intentional puzzles using some action, the next time we use the same action you'll leap to the same conclusion. Schelling (1960) showed that tacit communication, even without any overt signal, allows people to coordinate actions if they share the goal—each will think what the other might be thinking (these are games of pure coordination); furthermore, repetitive games give rise to salient solutions using earlier precedent. Lewis (1969) built on this and argued that all conventions have the character of arbitrary solutions on which participants jointly coordinate and that language is a huge edifice of such conventions. These authors thus provide us with an ultimate reduction of meaning to intention recognition, together with ways in which such complex inferences can be short-circuited by building up conventions in a community (we return to the psychological reality that may underlie this philosophical reduction in section 4).

We thus have both the demonstration of the ability to communicate complex propositions without language (homesign, first contact situations, and experimental semiotics and pragmatics) and some theoretical reconstruction of how it is possible. It is tempting to think that this communicative ability is in abeyance once we have complex linguistic communication at our disposal, but nothing could be further from the truth. Even the simplest sentence requires inferential resolution: If I say *The cat is on the mat*, already the definite articles constitute a coordination problem—which cat, which mat—and what sense of *on* is intended (is the cat a picture woven into the mat?)? But in interactive language use, the inferential requirements are compounded: conversation consists of exchanges of *speech acts*, sequences of actions such as greetings, offers, acceptances, requests, compliances, assessments or appreciations, and the like, typically coming in pairs of initiating action and response. But speech acts are often not directly signaled by the form of the utterance—there's no one-to-one mapping between form and function (Levinson, 2013): if I say *What are you doing tonight?* it might be a prelude to an invitation, proposal, or request rather than a question about your domestic routines—language use would be impossible without the deployment of the same inferential abilities that we see involved in radical tacit communication (Levinson, 2000; Sperber & Wilson, 1986).

## 2. *The Puzzle of Language Diversity*

There are some 7,000 languages today, grouped into some 430 families or highest level clades that can be reconstructed (Hammarström, Forkel, Haspelmath, & Bank, 2016). Languages of different families tend to show little similarity in their sound systems (with inventories of phonemes from 11 to 140), their word formation (little morphology or so elaborate that it more or less exhausts the syntax), their syntax (all manner of word orders, incomparable constructions, and such) or their semantics (absence of conditionals, proper color words, varied segmentation of the body into body parts, and such). Beneath the diversity, some structural parallels can be noted, but the lowest common denominators are few and far between (Evans & Levinson, 2009). In a study attempting to map the way languages have colonized the possible "design space," we found that as far as the mapping of semantic and functional distinctions into morphosyntax is concerned, unrelated languages are widely dispersed in their solutions (Levinson, Hammarström, & Roberts, 2018).

How then to account for the fact that we are the only language-bearing mammal? If language structure was built into the genes, this ought to be clearly revealed by crosslinguistic comparison. If instead there is just some special abstract facility for building complex syntax, as Berwick and Chomsky (2016) assumed, then that itself won't account for our abilities to learn such diverse languages. The more special, restricted, and uniform any such endowment is, the less it can account for language diversity. An alternative line of thought would emphasize the constraints from general cognition (perception and memory, for example; see Christiansen & Chater, 2008; Christiansen & Chater, 2016) that will limit the possible target languages, but such constraints are unlikely to differ greatly from those in our nearest primate cousins, and thus will probably not account for our unique language abilities. For that, what we crucially need are learning mechanisms. One such mechanism is our capacity for vocal learning, that is, the ability to hear a sound and mimic it in production. To date, there is not much evidence of this ability elsewhere in the primate order (as opposed, for example, to the passerines). But vocal learning only primes the pump for learning sound systems, crucial though that is.

It is here that the need becomes evident for some mechanism that can motivate and enable the child to learn the structure of the native language it must acquire. We suggest here that the child is endowed both with strong instincts to communicate and with instinctual understanding of the interactional system that humans use to communicate (Lee, Mikesell, Joaquin, Mates, & Schumann, 2009; Levinson, 2006). Using this understanding, the infant can bootstrap itself into the world of intersubjectivity (see section 4). On this account, languages are free to vary under drift and cultural selection because human language capacities lie very largely in an invariant, underlying, communicative infrastructure rather than in a universal "blueprint" for language structure.

## 3. *Universals of Interactive Language Use*

If there is an instinctual basis for patterns of interactive language use, we would expect strongly universal patterns of use to be evident. Here we make the case that these can be clearly discerned.

The conversational deployment of language is its core use—face-to-face interaction is the central niche in which it is learned, in which it evolved, and the home for the bulk of all language use. We each output on average a couple of hours a day, over 15,000 words, but distributed into small bursts of communication (on average around two seconds long), which alternate with the output of others. Any Martian ethologist would find this pattern of distributed recurrent short bursts of signaling one of the most notable features of human communication, observing that the groups involved can be of different sizes, that participants tend to face or orient to one another, and that the current speaker alone tends to produce concurrent hand movements, facial gestures, and such. Research on language use in conversation has mostly been undertaken by the nonexperimental sciences, partly because the phenomena are hard to capture in a laboratory setting, and partly because the study of language has often been construed more abstractly, outside its original and core functional niche. Extensive corpus work has been done, however, especially by the conversational analysts (see, e.g., Schegloff, 2007), but also by corpus phoneticians (Heldner & Edlund, 2010; Levinson & Torreira, 2015) and students of gesture and situated communication (Seyfeddinipur & Gullberg, 2014).

Although there is substantial cultural patterning of conversational behavior, its fundamental organization has recently been shown to be strongly universal. The alternating turn-taking, for example, has distinctive properties recurrent across all languages that have been examined: the timing is precise, with gaps of ~200 ms, overlapping speaking tends to be minimal and brief (~5% of the speech stream), and the exchange of roles seems organized by the same principles outlined by Sacks, Schegloff, and Jefferson (1974). A speaker gets a minimal clauselike unit as an initial turn, which can be extended by tacit or explicit agreement to another such unit, but after each such unit, speaker transition may occur (unless a speaker is specifically selected, the first starter becomes the next speaker). A study of 10 languages from five continents showed only minor differences in timing (Stivers et al., 2009).

The turn-taking system with its rapid transitions has interesting implications for language processing: the typical modal 200-ms gap is much tighter than the latency of speech planning and encoding (600 ms for a single word, 1000 ms or more for a clause), implying that next speakers must plan their utterances well before the incoming speech has ceased. This would require predicting both the content and the timing of the incoming turn, so that the relevant response can be planned to come in on time. We believe, on the basis of a series of experiments, that the recipient of an incoming turn tries early to predict the speech act (whether the utterance is a question or offer, for example) and begins planning the appropriate response as early as possible, withholding the response (if necessary) until turn-end signals are detected in the incoming turn, as sketched in figure 14.1 (see Bögels, Magyari, & Levinson, 2015; Levinson, 2016; Levinson & Torreira, 2015). This would account for the relatively stable response time, with a mode of 200 ms, close to the limit of human responses to simple "go" signals. Bearing in mind that response times for all tasks increase logarithmically as the number of choices increase (Hick's Law), and that a speaker selects from a vocabulary of at least 20,000 words, it is clear that this is a remarkable feat. The system of short response times of around 200 ms thus puts very considerable pressure on the processing systems for language.

It is worth asking why this system is the way it is. The possibility arises that this system arose prior to the

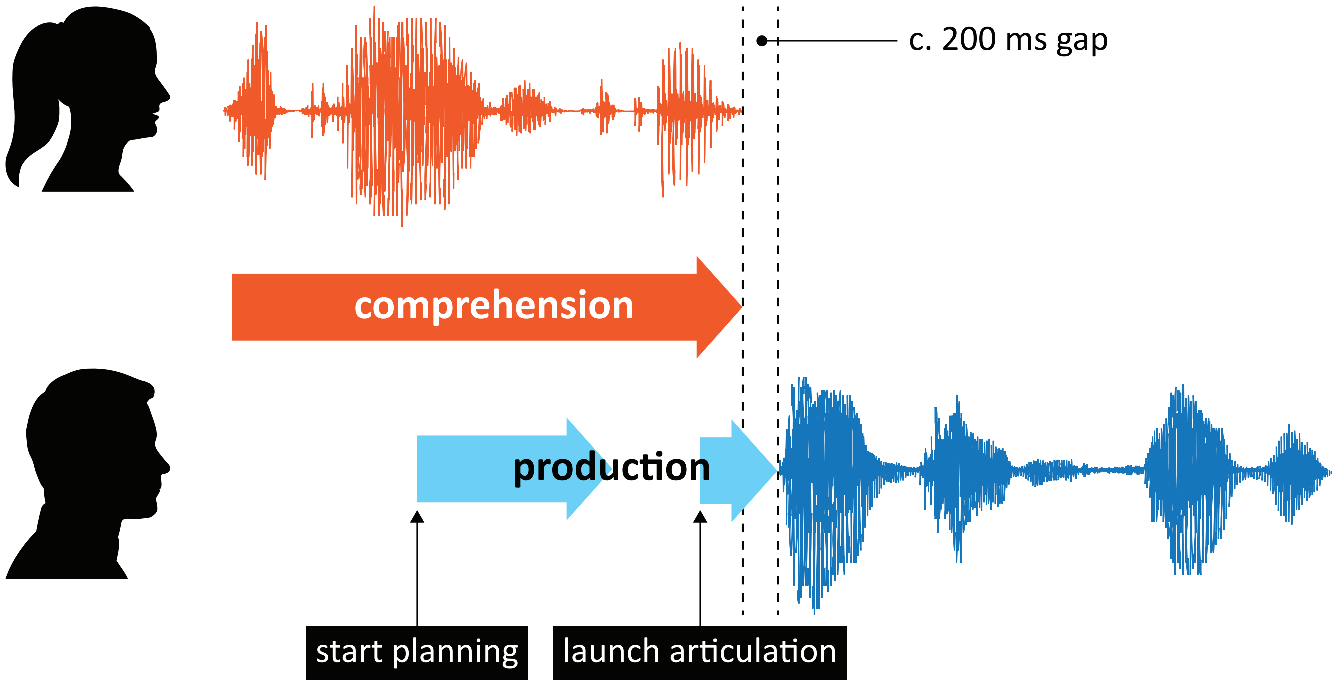


FIGURE 14.1 Overlapping comprehension and production in conversation (after Levinson, 2016).

evolution of complex language, and that the linguistic system has itself had to adjust to it. Infants today have to learn to squeeze ever increasingly complex encodings into these same short two-second bursts with quick response times, and struggle with it throughout the childhood years (see section 4). Perhaps the same slow development took place in phylogenetic time over the million years leading up to complex language (Dediu & Levinson, 2012).

In any case, the turn-taking system is a highly specific part of the ethology of human communication, which requires some account. One may ask what hurries participants on? One factor is that given the norm of a quick response time, delayed responses come to have a specific semiotics, signaling difficulties in responding in the desirable direction. For example, after a request, a delay is a likely harbinger of a refusal (Kendrick & Torreira, 2015; Pomerantz & Heritage, 2013). To avoid such suggestions interlocutors have to keep the pace up.

Turn-taking is just one layer of universal organization characteristic of interactive language use. Another such system is *repair*—a system for correction, clarification, and the maintenance of intersubjectivity. Schegloff, Jefferson, and Sacks (1977) outlined an ordered system of priorities, where provision is first provided for self-repair mid-turn, then at turn-transition, and then by other-initiation in next turn. Considerable crosslinguistic work has been done on the system of other-initiated repair (Enfield et al., 2013), and across languages it occurs at a rate of 1.4 times per minute (Dingemanse et al., 2015). The rate gives one some idea of its functional importance: without this means of correcting intersubjective understandings, our communications would rapidly derail.

In a study of a dozen languages from five continents, it was found that there are remarkable parallels in the shape and format of these initiators of repair, with most languages having a schwa-based particle (cf. English *huh?*) where the whole prior turn is within the scope of doubt (e.g., it was not heard or understood; Dingemanse, Torreira, & Enfield, 2013) and then a series of ways of targeting precisely items of lesser scope (Dingemanse et al., 2015). Even the bodily behavior of repair initiators is similar across cultures, often involving a freezing of kinesic movement (Floyd, Manrique, Rossi, & Torreira, 2016). Without a system of this kind, our understandings in interaction would soon go their separate ways.

Another area where crosslinguistic work shows remarkable parallels is in the higher order organization of turns into sequences. As we noted, conversations consist of exchanges of *actions* or speech acts, where these are very variably coded in linguistic form. A basic sequence is composed of an initiating action, followed by its response, such as questions and their answers, offers and acceptances, requests and compliance. Such "adjacency pairs" form a core on which elaborate extensions of structure can be built (Schegloff, 2007). There can be presequences such as the following preoffer sequence (p. 30):

A: Whatcha doin'? ← *Pre-offer* } *Pre-sequence*
B: Not much. ← *Go ahead*

A: Y'wanna drink? ← *Offer/Proposal* } *Base adjacency pair*
B: Okay. ← *Acceptance*

There can be insert sequences such as the following other-initiated repair (OIR) sequence (p. 102):

A: Have you ever tried a clinic? *Question (base first pair-part)*
B: What? } *Insert sequence (OIR)*
A: Have you ever tried a clinic?
B: ((sigh)) No, I don't want to go to a clinic *Answer (base second pair-part)*

And there can be post-adjacency-pair expansions, as in the following example after a repair insert (from p. 152 with many simplifications):

A: You sound HA:PPY hh ← *base first pair-part*
B: u- I sound ha:p[py?] } *Insert sequence (OIR)*
A: [Yee]uh
(0.3 sec.)
B: No ← *base second pair-part*
A: Nno:? } *Post expansion*
B: No

As even these simplified examples make clear, the structures that can be built with these building blocks can be quite complex. In fact, using insert sequences, far deeper center-embedding is found in conversational structure than can be found in language syntax, as illustrated next (Levinson, 2013, 155, ex. 14, after Merritt 1976):

(14) Merritt (1976)

C: Do you have master carbons?
(Q(prerequest: 0)
S: (pause) Yes, I think we do
(A:0)
Wat kind do you want?
(Q:1)
C: How many kinds do you have?
(Q:2)
S: Well, there are carbons for gelatin
(A:2)
duplicators, and carbons for spirits
C: Well, I'll take the carbons for spirits, please
(A:1)
S: ((goes to get))
(Action: 0)

A recent study (Kendrick et al., 2014) showed that virtually all these complex patterns of usage can be found in a dozen different languages of 11 distinct unrelated language families. So once again, these patterns seem to have a "natural" origin.

We have here reviewed three dimensions of conversational organization—turn-taking, repair, sequence organization—that have been studied in careful, controlled comparisons across 10 or more languages around the world. Another domain for which we have much comparative material is the greeting behavior that initiates verbal interaction (see Duranti, 1997). We can anticipate that many other features of the conversational niche will turn out to be equally pan-human, for example the use of laughter, gaze, and gesture, which are known to be culturally curtailed in variable ways (see, e.g., Rossano, Brown, & Levinson, 2009), but which nevertheless can be found in probably all societies (see, e.g., Sidnell, 2009). This universality stands in contrast to the extraordinary variability of human language itself across social groups, as mentioned at the outset.

What does the strong universality betoken? It could perhaps be simply a question of optimal design for the niche—a set of recurrent best solutions to recurrent problems in communication (Schegloff, 2006). But more than that seems involved. Let us consider turn-taking in a bit more detail. Would functional considerations inevitably lead communicators to take turns at speaking? A number of observations argue against a purely functional explanation, an argument for parallel cultural evolution or emergence. For example, could turn-taking be basically motivated by the difficulty of listening and speaking at the same time? Miller (1947) showed that one voice is in fact poor masking for another—only when there are many background voices is intelligibility deeply compromised. Perhaps a more persuasive point is that turn-taking in the sign languages of the deaf appears to be identical to that of spoken languages (de Vos, Torreira, & Levinson, 2015): here there are two distinct channels, vision and action, and hence no channel masking occurs, but nevertheless normal conversational turn-taking occurs. Another functional motivation might be what Sacks et al. (1974) called the "proof procedure" offered by the turn-taking system: if the speaker issues only a short utterance, he can find out rapidly from the response whether he was immediately understood. That might motivate rapid turn-taking as a rapid system of checks. However, that predicts a counterfactual, namely that people would respond as soon as they can predict what speech act is intended. We know that speech act interpretation is remarkably early, sometimes if context permits as early as the first word or two of the incoming turn (Gisladottir, Chwilla, & Levinson, 2015), yet responders clearly aim to delay their turn until the incoming turn is completed. In fact, overlap occurs in only 5% of the speech stream (Levinson & Torreira, 2015). A final argument against a functional account of the universal patterns is that other turn-taking systems are not only conceivable, but actual—they occur in special institutional settings, such as classrooms, law courts, or parliaments. These in contrast to conversational turn-taking vary widely across cultures and indeed institutions, and appear functional within those contexts. So the universality of the conversational mode of turn-taking requires an independent explanation.

In the following sections, we produce some prima facie evidence for a much deeper explanation of these universal patterns, tracing their origin in ontogenesis and phylogenesis.

## 4. Ontogenesis of Interactive Skills

The phenomenon of *protoconversation* first came under scientific scrutiny in the 1970s (Bates, Camaioni, & Volterra, 1975; Bateson, 1971; Bruner, 1975; Trevarthen, 1977): infants begin early to show contingent responses to a caretaker's communicative displays. Let us first consider imitation. The ability to imitate has been thought to be instinctual, on the basis of manifestation in the first hours after birth (Meltzoff & Moore, 1977), but recent studies challenge this (Oostenbroek et al., 2016): it is at least an early-learned and complex mapping from visual image to motor output. By one month infants are cooing, and by three months infants can do vocal imitation of vowels, although the output more closely matches adult categories by five months (Kuhl & Meltzoff, 1996). This vocal imitation (apparently unique in the primates) must have an instinctual basis, but the match to target requires maturation of the vocal tract and learning of the local target categories.

Early proto-conversation involves more than the contingent matching of signals, it also requires control of timing—the turn-taking of conversation requires coordination between two or more parties, the one taking the floor, the other desisting. A series of recent studies has explored the development of this coordination from three months to three-and-a-half years and beyond. Three-month-old infants respond with the same latency as their mothers (median ~550 ms), but interestingly this response time gets greater with age up until nine months (1100 ms), when it decreases again (Hilbrink, Gattis, & Levinson, 2015); however, over the same time interval, the amount of overlapping turns produced by the infant diminish from roughly 40% to ~25%, but overlaps remain short in duration (between 600 and 500 ms). We interpret this as an instinctual tendency for response, but with temporal control developing over time.

From 18 months old, when language production begins in earnest, children respond to questions with latencies of about 870 ms, with this decreasing over the next two years to about 530 ms (Casillas, Bobb, & Clark, 2016). This is slow by adult standards and remains slow through to middle childhood as the complexity of the language increases both in input and output (Stivers, Sidnell, & Bergen, 2018). The delay in response is fairly clearly due to output difficulties, since children from two-and-a-half years of age already use linguistic cues to anticipate the ends of turns like adults do (Lammertink, Casillas, Benders, Post, & Fikkert, 2015).

Part of early interactional abilities involves the understanding of another universal interactional system, the repair system. From their first words, children perform self-repair quickly, thus forestalling other-initiated repair, and by 2;0 years old children respond to other-initiated repairs or requests for clarification about 75% of the time, mostly with helpful reformulations rather than repeats (Casillas, 2014). Their repair system appears to have all the hallmarks of the adult system remarkably early in development.

Turn-taking and repair have to do with the "mechanics" of interactive communication, but what about their understanding of the intentional background to communication? We earlier pointed to the universality of conversational sequences involving paired action types, such as questions and answers. Infants start producing "protoimperatives" (reaching for something while looking at their caretakers) and "protodeclaratives" (showing and giving of objects) by about 10 months old (Bates et al., 1975; Casillas & Hilbrink, 2018), and then pointing by 12 months (Liszkowski, Brown, Callaghan, Takada, & de Vos, 2012). These simple gestures are interpreted by caregivers as requests and declarations of interest, but what the infant actually has in mind is harder to determine. This brings us back to Grice's intentional analysis of meaning: The speaker or actor intends to cause an effect in the recipient just by getting the recipient to recognize that intention. This requires understanding and reconstructing other's mental states, but it has even been doubted that infants have any clear understanding of other's minds—children, after all, fail classic theory of mind tasks right until age four years (Baron-Cohen, Leslie, & Frith, 1985; but see Onishi & Baillargeon, 2005, who actually show false-belief understanding from 15 months). On the other hand, by 9 months infants begin to sustain joint attention and follow gaze, and by 12 months they understand and produce referential pointing and helpfully point to displaced objects to help adults (Liszkowski, Carpenter, Striano, & Tomasello, 2006), showing understanding of other's mental goals. The different results can be reconciled by noting that there are significant differences between explicit mind reading (required by the classical tasks) and implicit mind reading, which is indubitably early (Frith & Heyes, 2014).

If one reconsiders the Gricean analysis of meaning:

S $means_{NN}$ z by U iff:
1. S intends U to cause z in H
2. S intends (1) to be achieved by H recognizing (1)

there is little doubt that by the first year of life, infants grasp the first clause—they understand that communicating involves causing effects in other minds. What is less clear is whether they grasp the second clause, that communication involves the intention to make this mind-changing intention overt. Tomasello, Carpenter, and Liszkowski (2007) argued that the child does grasp this in the second year, at which time he or she is able to discount accidental behaviors and correct misunderstandings even when, for example, a request has been satisfied. Csibra (2010) suggested that human infants may be born with an innate sensitivity to an inflection as it were (motherese, gaze, and such) that marks an action as communicative, so guaranteeing the ability to recognize the second Gricean clause.

But regardless of the age at which this higher-level intention is recognized, it is important to see that there is a simple bridge to this higher-level intention recognition provided by the ritualization of behavior—where Y is an initial move followed by X, producing Y may elicit X. The infant may reach toward an object, and have it provided, without having a clear understanding of request behavior at all: it is just an observed contingency. Success at that level will lead to repetition, by which time reaching is a kind of tool that works by manipulating the recipient (Grice's condition 1). This level may be achieved not only by infants in the last months of the first year, but also by infant apes (see section 5). The next level is for the child to realize that the tool only works when she has the recipient's attention and willingness and the tool itself manifests its intention—the recipient has to discern the desire for recognition, and now we have the second clause or something close to it.

The early availability and universality of the interaction system predicts that children growing up in a linguistic vacuum would nevertheless learn to communicate, and to do so in remarkably similar ways—a prediction borne out by the phenomenon of homesign, where deaf children construct gestural signs in the absence of a cultural tradition (Goldin-Meadow, 2003).

To sum up this section, it is clear that the human infant is inducted into the interactive world remarkably early. The comparison to other primates (see section 5) suggests that the sheer rapidity, automaticity, and almost exceptionless success of this induction has an innate basis, along the lines of Darwin's (1871, pp. 55–56) "instinctive tendency to acquire an art." Within three months of birth, the infant is already an interactive being, acquiring communicative abilities proper within or soon after the first year of life. All this happens long before the child is producing language, which argues that it is not language that induces communicative and interactive abilities, but that these interactive instincts are the engine that facilitates the learning of language.

## *5. The Phylogenetic Background*

In trying to understand the evolution of our interactive abilities, we turn inevitably to our nearest cousins, the other primates. For various reasons, the record is patchy—there has been a tendency for example to observe "focal animals," so yielding ethograms for individual animals rather than interacting pairs. Nevertheless there are significant traces of the origins of our interactive abilities in the primate record.

Vocal turn-taking, for example, has been reported patchily across all the major clades of the primate order (see figure 14.2). Thus we have accounts of turn-taking from among the lemurs, such as *Lepilemur edwardsi* (Mendez-Cardenas & Zimmermann, 2009); from New World monkeys, such as the common marmoset, *Callithrix jacchus* (Chow, Mitchell, & Miller, 2015; Takahashi, Narayanan, & Ghazanfar, 2013), the pygmy marmoset, *Cebuella pygmaea* (Snowdon & Cleveland, 1984), the coppery titi, *Callicebus cupreus* (Müller & Anzenberger, 2002), and squirrel monkeys of the *Saimiri* genus (Symmes & Biben, 1988); from the Old World monkeys such as Campbell's monkey, *Cercopithecus campbelli* (Lemasson et al., 2011); and finally from the lesser apes, the siamangs, *Hylobates syndactylus* (Geissmann & Orgeldinger, 2000; Haimoff, 1981). Homoplasy—parallel evolution—cannot be ruled out, but the situation may be just as in the birds where vocal learning is discontinuously represented, suggesting easy gain, loss, and regain (Jarvis et al., 2014).

Particularly interesting are various reports on the learning of this turn-taking system by youngsters—just like human infants, juvenile Campbell's monkeys have the instinct to reciprocate, but the timing and avoidance of overlap have to be learned (Lemasson et al., 2011). Marmoset mothers reinforce the learning by

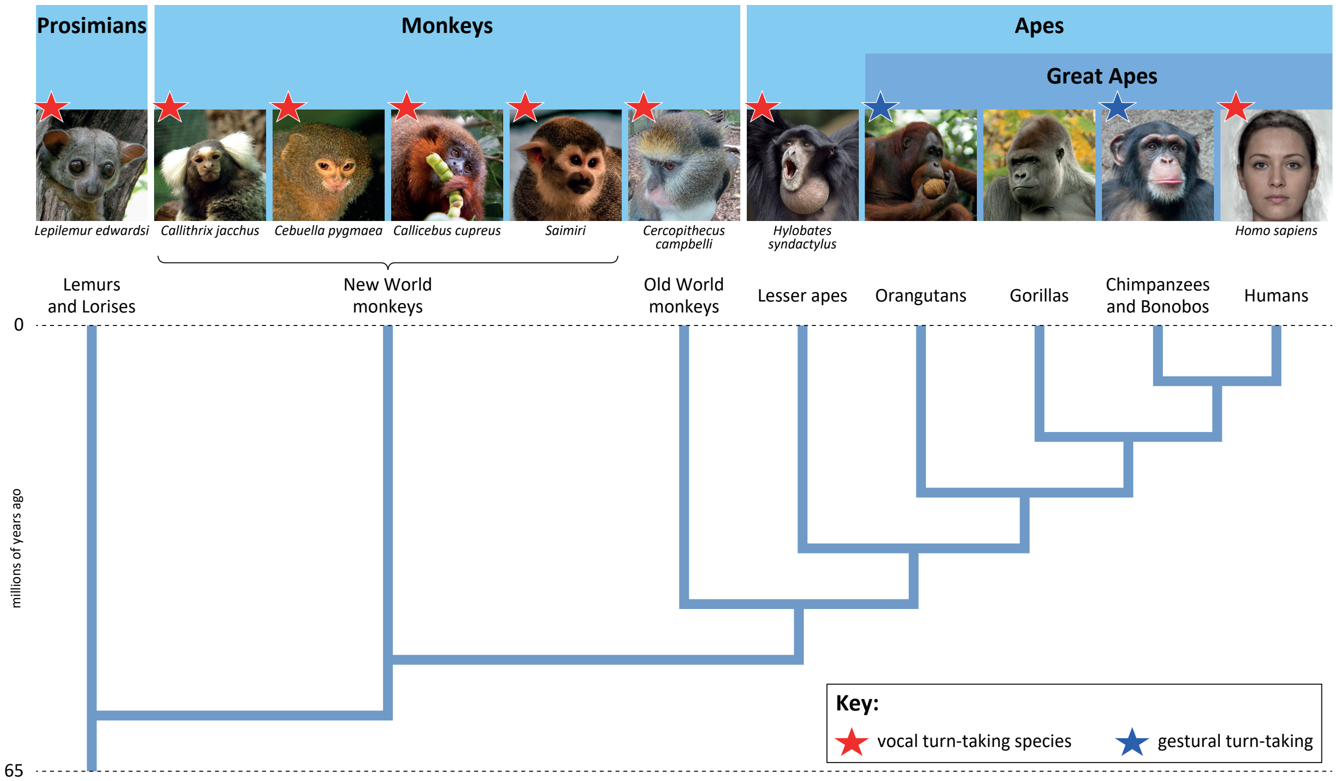


FIGURE 14.2 Vocal turn-taking represented across the primate order (from Levinson, 2016).

penalizing overlaps by the infant with nonresponse, and interrupting their infants when the response is of the wrong type (Chow et al., 2015; Takahashi, Fenley, & Ghazanfar, 2016). Vocal turn-taking seems, on the current evidence, to be particularly associated with pair-bonding species.

However, the great apes, apart from us, stand apart; their vocalizations have been described as involuntary and show little evidence of fast turn-taking (Call & Tomasello, 2007). But they use gesture in a flexible way to negotiate leaving together, food exchange, and the like (Pika, Liebal, Call, & Tomasello, 2005). The shape of the gestures is plausibly derived from ritualization of instrumental actions (Halina, Rossano, & Tomasello, 2013). Gestural turn-taking in bonobos has in fact a similar rapidity to human vocal alternation (Rossano, 2013). But the most interesting findings emerging from work on the great apes' gestural communication concerns parallels to the action sequences we reviewed in section 3. Orangutans, for instance, request and offer food using gestures of various kinds (approaching for mouth-to-mouth transfer, open-hand begging, apparent reaching), with success between a quarter and half the time (Rossano & Liebal, 2014). Requests may often be repeated, pursuing the goal unless rejected, with gaze maintained. A study of joint travel initiations among bonobos and their infants showed similar joint attention and sequences of eliciting and responding actions (Hutchins & Johnson, 2009; Rossano, 2013). In a larger scale comparison of bonobo and chimpanzee travel initiations, Fröhlich et al. (2016) showed that these sequences are systematic and show differences across the two closely related species, with bonobos showing more humanlike gaze engagement and speed of response.

All in all, there is increasing evidence of precursors to human interactional behavior among our nearest phylogenetic relatives.

## 6. The Interaction Engine Hypothesis

This chapter has brought together some systematic arguments for positing a distinct interactional ability that underlies our communicative behavior and language in particular. Communicative abilities can be dissociated from language: they are antecedent in ontogeny and phylogeny, and they permit humans to communicate effectively without language where deprived of it. These abilities are in their crucial properties universal—relatively invariant—across cultural groups, in strong contrast to the cultural specificity of linguistic coding. Furthermore, it is this interactional capacity that underlies the very possibility of language: languages are learned in the interactional niche, within the scaffolding that this provides, with joint attention to the referents being named, with exercise of other-initiated repair, and with the whole functional framework that sequences of actions (speech acts) provide. It is this interactional ability that I have dubbed "the interaction engine."

The objection may be made that this interactional ability is not one thing, but rather an assemblage of various talents and proclivities, with different phylogenetic origins and different ontogenetic patterns of development. All that is indubitable—the interaction engine is not a Fodorean module, or a Chomskyan single miraculous mutation, or a psychological faculty. It is rather a loose assemblage of various abilities, instincts, and motivations that work together to make possible the miracle of human communication. It is the bric-a-brac of useful oddments assembled through a long phylogeny, part of which may be reconstructed by comparison across species and cultures and evolutionary forbears. The sequence may have gone as follows (Levinson & Holler, 2014), perhaps cashing out Darwin's (1871, p. 106) observation that "the half-art and half-instinct of language still bears the stamp of its gradual evolution."

a. The fact that vocal turn-taking behavior is found scattered across all the major clades of the primates suggests that this may have been the foundational element in this accumulation. As we have seen it puts tremendous pressure on language processing (see also Levinson, 2016), and children are not able to match adult standards until well into middle childhood. The suggestion is that we have inherited a specific tempo, with the 1–2 s bursts separated by 2–300 ms, which was an ecology that language grew into and had to adapt to—early vocal turn-taking may have had the simplest functions, like the signals of proximity and well-being in marmosets (Takahashi et al., 2013). The complexity of linguistic signals reflects this compression into preexisting slots.
b. In the Hominidae, gesture has played an important role in communication (Call & Tomasello, 2007), implicating gaze, and humans show anatomical adaptation to the importance of the visibility of gaze direction in the white sclera of our eyes (Kobayashi & Kohshima, 2001) and their horizontal elongation (Mayhew & Gómez, 2015), which may be one of the earliest physiological adaptations for human communication. Gaze is important for establishing mutual attention, but gaze and gesture also directly afford spatial communication, suggesting that this may have played a crucial role in early hominin communication (Levinson & Holler, 2014). The face-to-face pose that would make this useful is the basis for what Goffman (1961, p. 18) called the "eye-to-eye ecological huddle" characteristic of human communication. Sustained joint attention makes possible the joint action that provides the working proof of other minds and the possibility of working with them.
c. Simple contingent action sequencing is found in gestural communication in the great apes (Rossano & Liebal, 2014), showing that for some actions such as requesting and offering, more complex communication media are not required (often gestures alone suffice in human interaction; Rossi, 2014). This precursor for human language seems restricted to certain social relations, mother-infant in particular, so one of the human tricks has been to generalize maternal empathy across our social relationships.
d. Vocal communication must slowly have developed in complexity in this interactional niche formed by turn-taking, gesture, and gaze. *Homo erectus* seems to have lacked voluntary breath control, but by the time of the common ancestor to Neanderthals and modern humans (well over half a million years ago) this special rewiring of the peripheral nervous system for elaborate speech was in place (Dediu & Levinson, 2013), responding no doubt to the adaptive pressures for expressive vocal language. The antecedent fast-paced turn-taking system explains the highly compressed nature of conversational speech.

Because this suite of capacities, the interaction engine, has probably been acquired en route through hominin evolution does not entail that it does not act as a package—even bipedalism required multiple adaptations over a long period, from adjustments to pelvis, vertebral column, and its articulation to the skull, not to mention balance and respiration. Some evidence for the package comes from *autism* (Frith, 2003; Frith & Hill, 2003), where the classical syndrome exhibits gaze aversion, difficulties with face recognition and emotion displays, and difficulties with joint attention and theories of mind tasks. Combined with this is lack of motivation for communication, problems with understanding gestures, and where, as often, language is well-formed, remarkable difficulties in the pragmatics of usage: there is no search for the underlying action or intent. Turn-taking involves the switching of deictic roles, so what was *you* becomes *I* and what was *there* becomes *here*, but these switches are difficult for autistic children, who often echo utterances without deictic adjustment.

Autistic viewers of conversation fail, unlike normal children from two years old and up (Lammertink et al., 2015), to anticipate speaker switches by switching gaze to the recipient or to the direction of pointing gestures (Klin, Jones, Schultz, Volkmar, & Cohen, 2002). Interestingly, these deficits cannot be attributed directly to low-level physiological causes such as failure of detection of looking direction—they seem to depend crucially on failures to understand joint attention and "theory of mind" (Baron-Cohen, 1995). Comparisons with other clinical populations are also revealing: Down's syndrome children, for instance, have full social and communicative competence despite their mental retardation (Baron-Cohen et al., 1985). A further recent development is the demonstration that genetic factors predisposing to autism also correlate with lesser interpersonal and communication skills in the general population (Robinson et al., 2016).

To conclude, the hypothesis is that the signal achievements of our species—language and cultural accumulation—rely crucially on a relatively neglected substrate of human interactional abilities, which seem largely instinctual in character, and which together have formed the crucible for human development both over the individual lifetime and the lifetime of the species.


## *Acknowledgments*

This report builds on extensive work on interaction structure conducted in my department from 2005: I am grateful to all my colleagues, and the Max Planck Gesellschaft and the European Research Council for generous funding. This chapter benefitted greatly from comments by Susan Goldin-Meadow on a draft.


## REFERENCES


Baron-Cohen, S. (1995). *Mindblindness and the language of the eyes: An essay in evolutionary psychology.* Cambridge, MA: MIT Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a 'theory of mind'? *Cognition, 21*(1), 37–46.

Bates, E., Camaioni, L., & Volterra, V. (1975). Performatives prior to speech. *Merrill-Palmer Quarterly, 21,* 205–226.

Bateson, M. C. (1971). The interpersonal context of infant vocalization. *Quarterly Progress Report of the Research Laboratory of Electronics, 100,* 170–176.

Berwick, R. C., & Chomsky, N. (2016). *Why only us: Language and evolution.* Cambridge, MA: MIT Press.

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports, 5,* 12881.

Bruner, J. S. (1975). The ontogenesis of speech acts. *Journal of Child Language, 2,* 1–19.

Call, J., & Tomasello, M. (Eds.) (2007). *The gestural communication of apes and monkeys.* Mahwah, NJ: Lawrence Erlbaum.

Casillas, M. (2014). Turn-taking. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 53–70). Amsterdam: Benjamins.

Casillas, M., Bobb, S. C., & Clark, E. V. (2016). Turn-taking, timing, and planning in early language acquisition. *Journal of Child Language, 43*(6), 1310–1337.

Casillas, M., & Hilbrink, E. (2018). *Communicative act development.* Manuscript in preparation.

Chow, C. P., Mitchell, J. F., & Miller, C. T. (2015). Vocal turn-taking in a non-human primate is learned during ontogeny. *Proceedings of the Royal Society B, 282*(1807), 20150069.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31,* 489–558.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences, 39,* e62.

Connolly, B., & Anderson, R. (1987). *First contact.* New York: Film Makers Library.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind and Language, 25,* 141–168.

Darwin, C. (1871). *The descent of man.* London: John Murray.

Dediu, D., & Levinson, S. C. (2012). Abstract profiles of structural stability point to universal tendencies, family-specific factors, and ancient connections between languages. *PLOS ONE, 7*(9), e45198.

Dediu, D., & Levinson, S. C. (2013). On the antiquity of language: The reinterpretation of Neanderthal linguistic capacities and its consequences. *Frontiers in Language Sciences, 4,* 397.

De Ruiter, J. P. A., Noordzij, M. L., Newman-Norlund, S., Hagoort, P., Levinson, S. C., & Toni, I. (2010). Exploring the cognitive infrastructure of communication. *Interaction Studies, 11,* 51–77.

de Vos, C., Torreira, F., & Levinson, S. C. (2015). Turn-timing in signed conversations: Coordinating stroke-to-stroke turn boundaries [Special issue]. *Frontiers in Psychology, 6*(268) (on turn-taking in human communication).

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., … Enfield, N. J. (2015). Universal principles in the repair of communication problems. *PLOS ONE, 10*(9), e0136100.

Dingemanse, M., Torreira, F., & Enfield, N. J. (2013). Is "Huh?" a universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PLOS ONE, 8*(11), e78273.

Duranti, A. (1997). Universal and culture-specific properties of greetings. *Journal of Linguistic Anthropology, 7*(1), 63–97.

Enfield, N. J., Dingemanse, M., Baranova, J., Blythe, J., Brown, P., Dirksmeyer, T., … Torreira, F. (2013). Huh? What?: A first survey in 21 languages. In M. Hayashi, G. Raymond, & J. Sidnell (Eds.), *Conversational repair and human understanding* (pp. 343–380). New York: Cambridge University Press.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive Science. *Behavioral and Brain Sciences, 32*(5), 429–492.

Floyd, S., Manrique, E., Rossi, G., & Torreira, F. (2016). Timing of visual bodily behavior in repair sequences: Evidence from three languages. *Discourse Processes, 53*(3), 175–204.

Frith, C., & Heyes, C. (2014). The cultural evolution of mind reading. *Science*, *344*, 6190.

Frith, U. E. (2003). *Autism: Explaining the enigma* (2nd ed.). Oxford: Blackwell.

Frith, U. E., & Hill, E. E. (2003). *Autism: Mind and brain*. Oxford: Oxford University Press.

Fröhlich, M., Kuchenbuch, P., Müller, G., Fruth, B., Furuichi, T., Wittig, R. M., & Pika, S. (2016). Unpeeling the layers of language: Bonobos and chimpanzees engage in cooperative turn-taking sequences. *Scientific Reports*, *6*, 25887.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, *25*, 737–767.

Geissmann, T., & Orgeldinger, M. (2000). The relationship between duet songs and pair bonds in siamangs. *Hylobates syndactylus. Animal Behaviour*, *60*, 805–809.

Gisladottir, R. S., Chwilla, D. J., & Levinson, S. C. (2015). Conversation electrified: ERP correlates of speech act recognition. *PLOS ONE*, *10*(3), *e0120068*.

Goffman, E. (1961). *Encounters: Two studies in the sociology of interaction*. New York: Bobbs-Merrill.

Goodwin, C. (1995). Co-constructing meaning in conversations with an aphasic man. *Research on Language and Social Interaction*, *28*(3), 233–260.

Goldin-Meadow, S. (2003). *The resilience of language: What gesture creation in deaf children can tell us about how all children learn language*. Oxford: Psychology Press.

Grice, H. P. (1957). Meaning. *Philosophical Psychology*, *67*, 377–388.

Grice, H. P. (1989). *Studies in the ways of words*. Cambridge, MA: Harvard University Press.

Haimoff, E. H. (1981). Video analysis of siamang *(Hylobates syndactylus)* songs. *Behaviour*, *76*, 128–151.

Halina, M., Rossano, F., & Tomasello, M. (2013). The ontogenetic ritualization of bonobo gestures. *Animal Cognition*, *16*(4), 653–666.

Hammarström, H., Forkel, R., Haspelmath, M., & Bank, S. (2016). *Glottolog 2.7*. Jena, Germany: Max Planck Institute for the Science of Human History. Retrieved from http://glottolog.org.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*, 555–568.

Hilbrink, E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: A longitudinal study of mother-infant interaction. *Frontiers in Psychology*, *6*, 1492.

Hutchins, E., & Johnson, C. 2009. Modeling the emergence of language as an embodied collective cognitive activity. *Topics in Cognitive Science*, *1*(3), 523–546

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., … Suh, A. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, *346*(6215), 1320–1331.

Kendrick, K. H., Brown, P., Dingemanse, M., Floyd, S., Gipper, S., Hayano, K., … Levinson, S. C. (2014). *Sequence Organization: A Universal Infrastructure for Action*. Paper presented at the 4th International Conference on Conversation Analysis, University of California at Los Angeles.

Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: A quantitative study. *Discourse Processes*, *52*(4), 255–289.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, *59*(9), 809–816.

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: Comparative studies on external morphology of the primate eye. *Journal of Human Evolution*, *40*(5), 419–435.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, *100*(4), 2425–2438.

Lammertink, I., Casillas, M., Benders, T., Post, B., & Fikkert, P. (2015). Dutch and English toddlers' use of linguistic cues in predicting upcoming turn transitions. *Frontiers in Psychology*, *6*, 495.

Lee, N., Mikesell, L., Joaquin, A., Mates, A., & Schumann, J. (2009). *The interactional instinct: The evolution and acquisition of language*. Oxford: Oxford University Press.

Lemasson, A., Glas, L., Barbu, S., Lacroix, A., Guilloux, M., Remeuf, K., & Koda, H. (2011). Youngsters do not pay attention to conversational rules: Is this so for nonhuman primates? *Scientific Reports*, *1*, 22.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Levinson, S. C. (2006). On the human 'interaction engine'. In N. J. Enfield, & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 39–69). Oxford: Berg.

Levinson, S. C. (2013). Recursion in pragmatics. *Language*, *89*, 149–162.

Levinson, S. C. (2016). Turn-taking in human communication, origins, and implications for language processing. *Trends in Cognitive Sciences*, *20*(1), 6–14.

Levinson, S. C., Hammarström, H., & Roberts, S. (2018). *How is the design space for language actually populated?* Manuscript in preparation.

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B*, *369*, 20130302.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *7*, 731.

Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Liszkowski, U. (2013). Using theory of mind. *Child Development Perspectives*, *7*(2), 104–109.

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & De Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cognitive Science*, *36*, 698–713.

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- and 18-month-olds point to provide information for others. *Journal of Cognition and Development*, *7*(2), 173–187.

Mayhew, J. A., & Gómez, J. C. (2015). Gorillas with white sclera: A naturally occurring variation in a morphological trait linked to social cognitive functions. *American Journal of Primatology*, *77*(8), 869–877.

Meltzoff, A., & Moore, M. K. (1977). Imitation of facial and manual gestures of human neonates. *Science*, *198*, 75–78.

Mendez-Cardenas, M. G., & Zimmermann, E. (2009). Duetting: A mechanism to strengthen pair bonds in a dispersed pair-living primate *(Lepilemur edwardsi)*. *American Journal of Physiological Anthropology*, *139*, 523–532.

Merritt, M. W. (1976). *Resources for saying in service encounters.* (Unpublished doctoral dissertation). University of Pennsylvania, Philadelphia.

Miller, G. A. (1947). The masking of speech. *Psychological Bulletin*, *44*, 105–129.

Müller, A. E., & Anzenberger, G. (2002). Duetting in the Titi monkey *Callicebus cupreus*: Structure, pair specificity and development of duets. *Folia Primatologica*, *73*, 104–115.

Noordzij, M. L., Newman-Norlund, S. E., de Ruiter, J. P., Hagoort, P., Levinson, S. C., & Toni, I. (2009). Brain mechanisms underlying human communication. *Frontiers in Human Neuroscience*, *3*, 14.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258.

Oostenbroek, J. M., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., … Slaughter, V. (2016). Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Current Biology*, *26*(10), R412–R414.

Pika, S., Liebal, K., Call, J., & Tomasello, M. (2005). The gestural communication of apes. *Gesture*, *5*(1/2), 41–56.

Pomerantz, A., & Heritage, J. (2013). Preference. In T. Stivers & J. Sidnell (Eds.), *Handbook of Conversation Analysis* (pp. 210–228). Chichester, UK: Wiley-Blackwell.

Robinson, E. B., St. Pourcain, B., Anttila, V., Kosmicki, J. A., Bulik-Sullivan, B., Grove, J., … & Martin, J. (2016). Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nature Genetics*, *48*(5), 552–555.

Rossano, F. (2013). Sequence organization and timing of bonobo mother-infant interactions. *Interaction Studies*, *14*, 160–189.

Rossano, F., Brown, P., & Levinson, S. C. (2009). Gaze, questioning and culture. In J. Sidnell (Ed.), *Conversation analysis: Comparative perspectives* (pp. 187–249). Cambridge: Cambridge University Press.

Rossano, F., & Liebal, K. (2014). "Requests" and "offers" in orangutans and humans. In P. Drew & E. Couper-Kuhlen (Eds.), *Requesting in social interaction* (pp. 333–362). Amsterdam: John Benjamins.

Rossi, G. (2014). When do people not use language to make requests? In P. Drew & E. Couper-Kuhlen (Eds.), *Requesting in social interaction* (pp. 303–334). Amsterdam: John Benjamins.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, *50*(4), 696–735.

Schegloff, E. (2006). Interaction: The infrastructure for social institutions, the natural ecological niche for language, and the arena in which culture is enacted. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 123–148). Oxford: Berg.

Schegloff, E. (2007). *Sequence organization in interaction: A primer for conversation analysis: Sequence organization.* Cambridge: Cambridge University Press.

Schegloff, E., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–382.

Schelling, T. (1960). *The strategy of conflict.* Cambridge, MA: MIT Press.

Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, *305*(5691), 1779–1782.

Seyfeddinipur, M., & Gullberg, M. (Eds.) (2014). *From gesture in conversation to visible action as utterance: Essays in honor of Adam Kendon.* Amsterdam: John Benjamins.

Sidnell, J. (2009). *Conversation analysis: Comparative perspectives.* Cambridge: Cambridge University Press.

Snowdon, C. T., & Cleveland, J. (1984). "Conversations" among pygmy marmosets. *American Journal of Primatology*, *7*, 15–20.

Sperber, D., & Wilson, D. (1986). *Relevance.* Oxford: Basil Blackwell.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592.

Stivers, T., Sidnell, J., & Bergen, C. (2018). Children's responses to questions in peer interaction: A window into the ontogenesis of interactional competence. *Journal of Pragmatics*, *124*, 14–30.

Symmes, D., & Biben, M. (1988). Conversational vocal exchanges in squirrel monkeys. In D. Todt, P. Goedeking, & D. Symmes (Eds.), *Primate vocal communication* (pp. 123–132). Berlin: Springer.

Takahashi, D. Y., Fenley, A. R., & Ghazanfar, A. A. (2016). Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Philosophical Transactions of the Royal Society B*, *371*, 20150370.

Takahashi, D. Y., Narayanan, D. Z., & Ghazanfar, A. A. (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. *Current Biology*, *23*, 2162–2168.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*(3), 705–722.

Trevarthen, C. (1977). Descriptive analyses of infant communicative behavior. In H. R. Schaffer (Ed.), *Studies in mother-infant interaction* (pp. 227–270). London: Academic Press.

Van Deusen-Phillips, S. B., Goldin-Meadow, S., & Miller, J. P. (2001). Enacting stories, seeing worlds: Similarities and differences in the cross-cultural narrative development of linguistically isolated deaf children. *Human Development*, *44*, 311–336.