PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

rstb.royalsocietypublishing.org

Opinion piece

Cite this article: Levinson SC, Holler J. 2014 The origin of human multi-modal communication. *Phil. Trans. R. Soc. B* **369**: 20130302.
http://dx.doi.org/10.1098/rstb.2013.0302

One contribution of 12 to a Theme Issue 'Language as a multimodal phenomenon: implications for language learning, processing and evolution'.

Subject Areas:
behaviour, cognition, evolution

Keywords:
language evolution, intentional vocalization, gesture, multi-modal communication, deixis, iconicity

Author for correspondence:
Stephen C. Levinson
e-mail: stephen.levinson@mpi.nl

Royal Society Publishing

The origin of human multi-modal communication

Stephen C. Levinson[1,2] and Judith Holler[1]

[1]Language and Cognition Department, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, The Netherlands
[2]Linguistics Department, Radboud University Nijmegen, The Netherlands

One reason for the apparent gulf between animal and human communication systems is that the focus has been on the presence or the absence of language as a complex expressive system built on speech. But language normally occurs embedded within an interactional exchange of multi-modal signals. If this larger perspective takes central focus, then it becomes apparent that human communication has a layered structure, where the layers may be plausibly assigned different phylogenetic and evolutionary origins—especially in the light of recent thoughts on the emergence of voluntary breathing and spoken language. This perspective helps us to appreciate the different roles that the different modalities play in human communication, as well as how they function as one integrated system despite their different roles and origins. It also offers possibilities for reconciling the 'gesture-first hypothesis' with that of gesture and speech having evolved together, hand in hand—or hand in mouth, rather—as *one* system.

1. Introduction

Human communication is unusual in the animal world on at least two principal counts: it has an unrivalled complexity and expressivity on the one hand, and an unparalleled inter-group variation on the other. The combination is extraordinary, because the variation within an unusually genetically homogeneous species excludes a fully biological explanation. In this paper, we take the view that human communication is evolutionarily stratified, composed of layers of abilities of different types and different antiquity. A wide range of scholars of different perspectives seem to subscribe to such a general view (e.g. [1,2]). But in this paper we suggest that viewing language as embedded in its full pragmatic, interactive and multi-modal context transforms this stratificational perspective. Unpeeling the layers can then help us see the different contributions of the distinct systems that underlie the peculiarities of human communication.

It is often hard for the literate world to remember that the core ecology for language use is in face-to-face interaction—this is the niche in which languages are learnt and where the great bulk of language use occurs. In this niche, language production always occurs with the involvement of not only the vocal tract and lungs, but also the trunk, the head, the face, the eyes and, normally, the hands. Our upright posture allows the whole ventral surface of the body to be used in communication. The speaker produces a multi-modal display, part semiotic, part entrained by the effort of vocal production. When thinking about the origins of human language, it is essential to bear this ensemble of linked systems in mind. The ease with which human language switches the main channel or modality carrying lexical material from mouth to hands, as in the sign languages of the deaf, should be a constant reminder that human communication is a system of systems, where the burden of information can be shifted from one part to another (see also [3–6]).

2. Human communication as an evolutionarily stratified system

The modern human communication system is, on a biological time-scale, a recent innovation. The standard line is that the emergence of language followed, perhaps with a 100 000 year lag, by the emergence of anatomically modern humans about 200 000 years ago. Clearly, however, it must predate the great diaspora of modern humans thought to date to 60 000 years ago. In a recent meta-study surveying all the genetic, anatomical and archaeological evidence, Dediu & Levinson [7] argue that a compelling case can be made for a much earlier origin of modern vocal language at over half a million years ago, dating back to the common ancestor (often identified as *Homo heidelbergensis*) of modern humans and Neandertals (see [8] for additional evidence supporting the idea of an early origin of modern human language). The development of speech required a by-pass of the involuntary breathing control involved in primate calls and a direct connection to the primary motor cortex, on the one hand, and to the intercostal muscles via the vertebral column, on the other [9,10]. This enlarged enervation is missing from the vertebral column of *Homo erectus* in the one well-preserved fossil that has come down to us, with an age of 1.4 Ma [11]. Although pre-adaptive changes in the thoracic vertebrae may have occurred for other reasons such as singing [12] or running [13], they would have rapidly come to play a crucial role in speech [14]. So, it seems probably that the origin of modern speech, and the prevalence of the vocal tract in human communication, would seem to lie somewhere between 1.4 Ma and *ca* 600 000 years ago. This remains, of course, a recent date, compared to, for example, the 50 Myr of evolution behind birdsong or bat echolocation.

(a) Pragmatics as the foundation of human communication

But speech is just one system in the 'system of systems' that constitutes human communication, and given its recent origins, there is every reason to expect that it was a late layer on other systems. This would fit well with a gestural theory of language origins [2,15–24]. Call & Tomasello [25] show that among our closest relatives in the Hominidae, intentional communicative acts are signalled largely by non-vocal means, especially with the hands (but see [26,27] on the intentional nature of at least some great ape vocalizations). Many of these gestures have an iconic or indexical meaning, for example, an offering of something presented with the hand or other body parts [28], a request made with an open hand [29], or a request to be picked up by an infant with raised arms [30]. Such semiotic functions can be readily associated by ontogenetic ritualization—the use of an initial gesture in a sequence coming to indicate the full sequence of actions [31]. In the great apes, ritualized gestures emerge in the following way: A does something XYZ, which routinely leads B to do something else; B comes to respond on the basis of only the first step, X, of A's action; subsequently, A anticipates B's response and only performs X in order to elicit the response [25, pp. 7–8]. In a broader sense, ritualization is the major route by which all animal communication systems seem to arise [32], but in the great apes, including us, the process takes place by learning between interacting dyads, rather than by biological evolution—what is biologically required are the motivation and ability to predict an action sequence from an initial part, and to conventionalize it as a means of expression. By the standard phylogenetic reasoning, we should have inherited this system, and indeed we almost certainly have, but have supplemented it with speech among other capacities. Indeed, careful observation of human interaction reveals the very same process in constant operation—if I reach as if to get the water, you pass it to me. Much of the facility with which we interact and cooperate depends on precisely this kind of nonce gesture [33].

Researchers in linguistic pragmatics have long held that language is the tip of an iceberg riding on a deep infrastructure of communicational abilities [34,35]. Simple utterances are rarely interpreted at face value—thus *Are you using that pencil?* is likely to be read as a request, *Do you want another beer?* as an offer, *What are you doing tonight?* as a prelude to an invitation, and so on. Grice [36] emphasized the presumption of cooperation and the reciprocal reconstruction of communicative intentions, while Clark [3] has emphasized the joint activities in which communication is embedded, and conversation analysts [37] have stressed the complex interactional conventions or practices that impose semiotic significance onto utterances (as when 'forty three' is an answer to a specific question about someone's age or house number, etc.). These different viewpoints all contribute to our understanding of the large inferential background to human communication, and its special role in the origin of human communication has been argued for by Tomasello [2,38].

Recently, it has become clear that many aspects of this pragmatic background to human communication are universal, or at least highly similar across cultures. The inferential system that Grice called 'generalized conversational implicature' seems to be universal in principle and generates small inferences from (sometimes language-specific) contrasts in the lexicon [39]. There are clear tendencies for a core inventory of speech acts like questions, greetings, requests and so on, to recur across cultures, even though many actions beyond the core vary [40]. Above all, the interactional system that is essential for human communication seems to have a strong common foundation across all cultures. For example, the core niche for language use in all cultures is a speech and gesture exchange system in which participants take short, rapidly alternating turns. It is in this context that the great bulk of language production occurs—extrapolating from earlier studies, we each seem to produce about 16 000 words in about 1200 turns a day in conversation [41]. Turns on average across languages have only about 200 ms between them [42], which is extraordinary if one bears in mind that the latencies involved in producing even a single word are of the order of 600 ms [43]. Consequently, most language use in conversation is based on a predictive look-ahead (e.g. [44–46]). Turn-taking appears remarkably early in infancy, certainly within the first year and well before the first words [47]. It is not a simple reflex of the limitations of the vocal–auditory channel, as it occurs equally in manual sign languages. Interestingly, vocal alternation also occurs (if patchily) right across the primate order, from lemurs [48] to marmosets [49,50], from Campbell's monkeys [51] to gibbons [52]. Although it is, curiously, not a major feature of communication among the great apes, the trait is perhaps persistent enough to suggest homology, though convergent evolution cannot be ruled out [53,54].

As noted, many speech acts seem universal in character, and so do the sequences of actions they construct—for example, pairs of questions and answers, offers and acceptances, greetings and greetings, and so forth. These sequences, though typically expressed in language, are also embodied in other ways: a request (perhaps visual) may prompt a visible action, a wave another wave back, a passing of a needed item a reciprocal grasping and so forth. They also appear to have a largely universal structure, and recent work suggests that precursors of some of these very same visual, bodily action sequences can be seen in apes [30,55]. Sequences can be embedded in other sequences, so offering recursive structures that may be the ultimate origin of recursion in language syntax [40].

Human interactional ethology, despite the presence of precursor elements in other primate species, as a whole ensemble is entirely distinctive. For example, the toleration or indeed expectation of mutual gaze is of paramount importance in humans [56] but occurs less in other primates (e.g. [57], see also [58]), and the white sclera of the human eye has almost certainly evolved to enhance gaze detection [59]—remarkably, human infants are sensitive to the difference between direct and averted gaze from just 2–5 days after birth [60]. The rapid turn-taking despite indefinitely varying contents of turns is again without parallel; the sustained multi-modal deployment of vocal and visual signals on hands, face and body and the sheer amount of time and effort invested in communication seem without parallel among the other primates. The capacities that lie behind this unique ethology have been called the *interaction engine* [61], a term we use later.

(b) Indexicality and iconicity in language evolution

One communicational peculiarity of humans is pointing, especially with the index finger [2,62,63]. It appears in human infancy before language, before the first year and—as far as we know—across cultures [64]. On the one hand, it may arise from ritualization of a reaching gesture [65], but this will only account for the so-called imperative pointings. Although apes display begging behaviour with the open hand [29,66], offer food in the palm [28] and may perform what appear to be deictic gestures with the whole hand [67] or localized referential actions with deictic elements [68], they do not point with the index finger (but see [69] for some possible exceptions in captive apes). Declarative pointing, on the other hand, whose sole function is to draw the other's attention to something, is human-specific in both form and function—it typically involves the index finger extended and has an abstract symbolic function: the recipient is meant not to look at the finger, but in the vector indicated beyond the finger, and find some referent that is probably what the pointer had in mind as worthy of attention. It would therefore appear to be tightly linked to social cognitive development [70], including Theory of Mind [71], and for that reason has been heralded as an ontogenetic milestone on the way to language (and, indeed, occurs in the majority of cases with vocalizations [72]). Pointing is thus an extremely powerful device, not because it accurately denotes (as Wittgenstein ([73], paragraph 33) noted, a point at, say, pieces of paper could be indicating the colour, the shape or even the number) but because it invites the recipient to locate a referent of mutual interest (a social process that recruits reward-related neurocircuitry in humans [74]). In some cultures, pointing plays a fundamental role in communication, especially when there are conventions of accurate orientation: Guugu Yimithirr speakers, for example, will inspect a gesture for precise directional 'compass' veracity, which allows the elliptical reference to places and persons without verbal reference at all [75]. Village sign languages are systems that have arisen *de novo* over four or more generations in remote areas where there are significant pockets of inherited deafness; these typically make extensive use of pointing, to the extent, for example, where the language has no need for place-names [76].

While pointing has a strong symbolic element with its arbitrary and species-specific use of the index finger, many of the details of pointing are indexical—for example, the varied hand shapes, orientations and elevations of the arm are adapted to the relevant aspects of the surroundings, as when the distance of an unseen referent is indicated with the vertical angle of the arm, probably because more distant objects are higher in the field of view on a plane surface. Since the kinds of gestures that derive from ritualization are also largely indexical in character, indexicality may play a much greater role in the gestural and visual modalities than is usually thought, where iconicity is often given pride of place.

Iconic representation is nevertheless another important element in the accumulation of communicative capacities [20,77]. The apparently sudden appearance of iconic representations like cave paintings in the archaeological record some 30 000 years ago has even led some scholars to suggest that they index the birth of the modern mind [78,79]. However, since Saussure's insistence on the arbitrariness of the linguistic sign, iconicity in language has been thought to lie largely at the borders of the linguistic system, for example in sound-symbolism. Some spoken languages nevertheless bundle considerable expressive power in a specific grammatical category, traditionally called expressives or ideophones (words conveying, for example, the quality of a sensory experience by exploiting not only consistent meaning–sound relationships but also structural analogy, as in reduplication [80,81]). Despite this lexical recognition, these words resist deep incorporation into the grammatical machinery of a language. This resistance reflects a tension between grammatical integration and richness of iconic form and meaning—once integrated, expressives become subject to the normal processes of reduction and regularization and lose most of their resonance [82]. This puts limits on the extent to which spoken language can fully exploit its own iconic potential, even though this is visible throughout linguistic systems (see [77] for a review), as for example in prosody or in the sequential reading of Caesar's *veni, vidi, vici* (I came, I saw, I conquered), which relies on the parallel between linguistic time and event time (for wide-ranging principles of this type operating in grammar, see [83]).

Where iconicity comes into its own is in gesture. Iconic gestures are the gestures that mimic motion, depict size, trace shapes or sketch the spatial relations between things [5,84], and they have intrigued researchers. The prime role that iconicity plays in gesture is the special affordances of gesture to indicate spatial relations. Spatial thinking seems central to human cognition [75] and plays a role in structuring many other cognitive domains, like time, kinds of social relationship or set relations. The relative position of entities with respect to each other is frequently depicted by iconic gestures [85,86], because the two hands provide a natural means of conveying the spatial relations between two (or more) things. And whereas spoken language is, through its finite lexicon, invariably coarse on spatial relations, gesture affords accurate depictions of angle, orientation and shape: the two together offer the complementarity of 'digital' and 'analogue' channels.

The ability of the gestural modality to depict spatial relations has implications beyond the spatial domain, for iconic gestures and signs are well suited to depicting transitivity, and thus agents and patients [87,88]. In the evolution of language, this facility to indicate space must have given gesture a special importance (for a detailed discussion of the role of the visual modality in the emergence of complex syntax, see [2]). In current human communication, gestures depicting spatial relations still carry a great deal of the communicative import of informal conversation.

All this suggests a long phylogenetic acquisition of layers of this system. At the deepest stratum, one may assume a capacity for ad hoc gestural symbols ritualized from early elements of action sequences, as found among the present great apes. Gestures of this kind may have been early enriched with simple vocalizations that, among other things, would draw visual attention to the signaller. A stratum layered on top of this, before complex speech evolved, might have consisted of a pointing-enriched manual sign system (something similar to a 'gestural protolanguage', as has been proposed by others, e.g. [16,89–91]) and may have been associated with the speechless *H. erectus*, the first human species to exit Africa some 1.5 Ma spreading throughout Eurasia and carrying a highly developed technology including Acheulian (Mode 2) handaxes and the control of fire. These handaxes are stone tools regularly flaked on both sides to produce symmetrical, flattened pear-shaped tools with sharp cutting edges, demonstrating advanced tool-making skills, which can only be replicated by months of practice under modern tuition [92]. This technology indexes a level of cultural replication that is pretty inconceivable without a complex and expressive communication system, perhaps approximating the complexity of village sign or homesign systems. We think it likely that many aspects of human communication ethology date back this far. Earlier varieties of humans, like *Homo habilis*, who was active over 2 Ma, had much simpler tool assemblages, but nevertheless this was, compared to the modern great apes, an advanced culture-carrying species that almost certainly had some relatively complex communication system.

(c) Phylogeny of a multi-modal, sequential communication system

As indicated above, modern human communication exhibits a specialized ethology not found in other animals: its face-to-face character, which affords the full deployment of multiple articulators, the frequent deployment of mutual gaze and the sustained exchange of short but complex communicative turns are characteristics. In addition, there is the cognitive background of joint attention, common ground, collaboration and the reasoning about communicative intent, as noted by many commentators on human communication (e.g. [2,3,36]). This whole assemblage, which we have called the *interaction engine* [61], is visible early in human development (e.g. in the 'proto-conversation' of six-month-old infants), and, in contrast to linguistic detail, strongly universal across cultures [42], and allows us to communicate without spoken words, as when I indicate with a gesture that you have signs of your breakfast on your chin. In modern humans, it is this system that enables the acquisition or enlargement of spoken or signed languages through 'ritualization' or the re-use of nonce signals that have been used successfully before. Our hypothesis is that, given its language independence, the interaction engine is phylogenetically older than language, and perhaps characterized the communication of early *Homo* before complex speech evolved. This system is itself no doubt phylogenetically layered, and one can speculate that the more ethological elements (e.g. mutual gaze, gesture and turn-taking) may have partially preceded and driven the underlying cognitive capacity for evaluating other minds, since there are plausible precursors in the primate lineage (e.g. gestures among the great apes, and turn-taking in other primate clades). The development of this system may in turn have been pushed by ecological changes to which increased cooperation was a necessary response [38]. We will here treat the interaction engine as a unified package, which is a simplification for current purposes. What is clear however from this package is that early human communication was highly visual and thus at least partially gestural, since it is otherwise hard to account for the white sclera and the unusual toleration and exploitation of mutual gaze.

Once an interactional system of gesture (including pointing) is combined with some capacity for iconic representation, itself a natural affordance of gesture, a system emerges that allows for communication about events happening elsewhere or in the past or future (Hockett's [93] design feature of 'displacement') [94]. Further, once the spatial representations of gesture have been exploited for relational encoding, we have the basis for the emergence of something like a basic sign language (similar to a homesign or village sign system), likely accompanied by simple vocalizations.

And, finally, as already indicated, it seems that complex vocal communication can be traced back somewhere before *H. heidelbergensis*, at over half a million years ago, when voluntary breathing was in place. A high degree of breath control is required for modern speech, since the depth of inhalation must correlate with the length of what is to be expressed, and the timing with every point of stress. This required, as mentioned, a rewiring of both the central and the peripheral nervous systems. There would be over 0.8 m years between *H. erectus* and *H. heidelbergensis* for the evolution of this system, and it was presumably acquired gradually in tandem with the increasing reliance on the vocal channel.

Table 1 summarizes the suggestions made in this section. We should add that much speculation on the evolution of language has gone on to suggest accumulated layers within the spoken language system (e.g. [2,16,22,95]), suggestions that attempt to spell out the steps between first simple vocalizations and complex modern spoken language. A review of these extensive developments lies beyond our immediate purposes here.

The layered acquisition of our communicative competences, as summarized in table 1, suggests that systems have been added but not replaced. The reason is that each layer has its natural affordances: ritualized bodily reaching is still a current practice used, for example, to make routine requests at the dinner table, pointing usefully augments speech and can be dovetailed with speech through the innovation of demonstrative words, iconic gestures can capture the movements, shape, size or spatial arrangement of entities better than a thousand words, and so forth. The gestural modality is especially well adapted to communicating about these sorts of visuo-spatial dimensions, which lie close to human preoccupations and are central to human cognition [75]. Speech itself has many advantages: the small articulators permit rapid and low-effort encoding, allow wide and distant signal

Table 1. Hypothetical layers of communicative competencies as they evolved.

time-scale	taxa	ritualized gestures	'interaction engine'	pointing	iconic gestural representations	voluntary vocal utterances	modern language capacities
6 m+	Hominidae	+					
2 m	early *Homo*	+	+	+	+		
1–0.6 m	immediate ancestor of *H. heidelbergensis*	+	+	+	+	+	
0.2	*H. sapiens* (incl. Neandertals)	+	+	+	+	+	+

broadcast, communication without sight and so on. Thus, a multi-modal communication system that combines both gesture and speech, and thus their complementary strengths and weaknesses, seems to meet human communication needs rather optimally.

The proposed sequential accumulation of layers of a communication system maps roughly onto the stages we observe during the development of communication in human children—proto-gestures deriving from ritualized action sequences such as stretching one's arms up in the air in order to be picked up occur very early (albeit in an already more generalized form than with young apes, [96]) and act as a form of pre-linguistic turn-taking, followed by pointing, followed by both kinds of brachio-manual action becoming integrated with speech to function as co-speech deictic and co-speech iconic (or other representational) gestures.

The details of this development are telling. For example, turn-taking before language can be quite fast, with the infant responding in under three-quarters of a second [97], approaching adult norms. But later, when children are trying to respond with more complex language at say 3 years of age, the response times can be twice as slow [98], converging with adult norms only in middle childhood. This suggests that the natural rhythm of conversation is independent of spoken language, and children have to gradually learn to compress complex material into the short rapid bursts of speech that adults use. The adult turn-taking speed puts extraordinary pressure on language production and comprehension: since it takes between 600 and 1500 ms to plan a response, and the gaps between turns are only on average 200 ms, this forces those engaged in dialogue to be already planning responses long before the other speaker has completed his or her utterance. Comprehension and production must thus work in parallel, with the next speaker predicting the ongoing turn by the other, in order to achieve precisely timed behavioural alternation. The whole system suggests an evolution from an original rapid exchange of very simple gestural or vocal material, into a system where the complexity of the linguistic and gestural material that is crammed into these short bursts has grown to the very limits that human cognition can process.

3. One multi-modal communication system or separate systems?

The evidence is that, despite the modern human communication system having evolved in layers (see [22,90] for one account of how the communication burden may gradually have shifted from hand to mouth), what results is one integrated multi-modal communication system, as suggested by many details of the whole assemblage. Evidence for this hypothesis is plentiful. For example, hand and mouth are closely connected in the somatotopic organization of the human motor cortex (e.g. [99,100]), and a very similar connection is also evident in the monkey motor cortex [101]. The hand–mouth connection is further evidenced by overt human behaviour such as drawing or cutting something, which is frequently accompanied by intricate movements of the tongue, lips or jaws. And although the hands are the major articulators in sign languages, the mouth and face are always also involved. Neuroanatomical asymmetries in the brains of non-human primates and the lateralization of both their vocal and their gestural communicative signals [102–104] further corroborate the notion of an early evolutionary link between hand and mouth (but see [105]). In addition, congenitally blind individuals gesture while they speak despite never having seen a single gesture [106,107]. And further evidence consists of the fact that neurons coding for manual goal-directed transitive movements occupy areas in the monkey brain that correspond to brain areas critical for processing language in the human brain—the putative mirror neuron system [108–110].

It may then be that there was pre-adaptation for an integrated multi-modal communication system based on a close marriage between hands and mouth, which was only fully exploited when the changes in cortical organization occurred that made voluntary breathing and intentional spoken communication possible. Given the large time-scales we are envisaging, the gradual co-evolution of vocal language with a pre-existing gestural mode of communication may have taken place over nearly a million years, so that the different modalities are deeply intertwined. This view may therefore not be as diametrically opposed as it seems to McNeill's [111] proposal that speech and gesture coevolved from the beginning. In the stratificational model we are proposing, our cousins the great apes suggest an early reliance on the gestural modality, but co-occurring simple vocalizations may rapidly have emerged as a way of drawing attention to the signals. Indeed, it is possible that multimodality is actually present in our great ape cousins and thus in the common ancestor, since the extent to which their gestural signals co-occur with vocalizations is still a largely unexplored domain. In any case, from those initial multi-modal seeds our fully-fledged multi-modal communication system would have evolved in the layered manner suggested. During the course of this evolution, the communicational burden has progressively but only partially shifted from hand to mouth, but sign languages offer a constant reminder that the roots of the system are 'hand + mouth' and where the motivation exists the burden can be readily shifted back.

Crucially, this perspective is at odds with the view that gesture was a 'bridging modality' that withered away once conventionalized spoken language had emerged (e.g. [2,21,109], see also [111]). Such a view reduces the role of gesture to a scaffolding function for the evolution and ontogenetic development of speech, with speech, once fully developed, completely eclipsing the multi-modal origins and leaving gesture as an almost functionless relic. This would be a fundamentally misleading picture. As other scholars have suggested (first and foremost [5,6,84,111–115]), *together* speech and gesture constitute adult human language—the two modalities are different components of one and the same system. Owing to their differing nature, these two components make perfect complements, each taking over what the other is less well suited to perform. During their coevolution, speech and gesture have mutually and maximally adapted to one another, resulting in a default mode for modern human communication characterized by highly efficient and informationally rich visual–verbal utterances. The system remains highly flexible, allowing us to shift the burden from words to gestures as required by the current communicative needs, as when hunting or on a noisy factory floor, with sign languages forming just one end of the spectrum [116].

Even in linguistically fully developed individuals, gestures remain a prominent part of human communication; adults accompany most clauses [5] or idea units [113] with one or more gestures, and children's use of gestures increases (rather than decreases) over the course of their acquisition of spoken language (with their gesture rate reaching adult levels by the age of 4–5 years [5]). Further evidence for the tight speech–gesture connection in adult human communication is the striking flexibility of this two-pronged communication system. Without speech, the gestural modality is able to take over much of the communicative burden, as evidenced by sign languages themselves, and when hearing adults are asked not to speak but to communicate only by means of their hands, they turn to creating sequentially structured gestural utterances [4,117]. Conversely, when preventing people from using gestures, the imagistic content of their messages decreases radically and speech becomes hesitant, further corroborating the assumption of gestures being integral to the human language system. Finally, during child development, the emergence of gesture and speech is intricately connected as evidenced by the two modalities' parallel developmental trajectories and patterns [5,118]. This connection is maintained in adulthood and evident in the precise temporal, semantic and pragmatic relationship of speech and gesture [5], and further supported by the finding that semantic information conveyed by speech and gesture is processed in the same brain areas [119–121].

Despite the remarkable flexibility that allows us to shift between visual and verbal modes of communication as needs require, the neglected modality is rarely fully repressed. Thus, deaf signers mouth and even vocalize, while speakers continue to gesture frequently even when the gestures are not visually accessible to our addressee, such as on the telephone [122], supporting the notion of an evolutionarily tight connection between gesture and speech. Interestingly, the use of unseen gestures is specifically associated with interactional uses of language [122,123] but not with monologue-like settings [124,125]. This observation is very much in line with the idea that gestural communication is strongly embedded in collaborative action [2,3] and that dialogic social engagement with another (as in the interaction engine) is foundational to human communication [61].

4. Conclusion

We have argued that despite the tight integration of the different modalities into modern human communication, the whole ensemble should be seen as a system of systems that has accumulated over the two and a half million years that humans have been a cognitively advanced, tool-using species. The accumulations can be thought of as strata, and peeling away the strata successively can give us some insights into the probable evolution of the whole complex system. This holistic account of our communicational capacities also helps to bridge the gulf between the articulate species and our inarticulate cousins, allowing us to see precursor adaptations in, for example, the turn-taking widely if sparsely represented in current primates and the gestural skills of great apes.

Acknowledgements. We thank the editor and two anonymous reviewers for their invaluable feedback on an earlier version of this manuscript.

Funding statement. The authors were supported through ERC Advanced grant no. 269484 INTERACT awarded to Prof. Levinson and through the Max Planck Society.

References

1. Hauser MD, Chomsky N, Fitch WT. 2002 The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569–1579. (doi:10.1126/science.298.5598.1569)
2. Tomasello M. 2008 *Origins of human communication*. Cambridge, UK: MIT Press.
3. Clark HH. 1996 *Using language*. Cambridge, UK: Cambridge University Press.
4. Goldin-Meadow S, McNeill D, Singleton J. 1996 Silence is liberating: removing the handcuffs on grammatical expression in the manual modality. *Psychol. Rev.* **103**, 34–55. (doi:10.1037/0033-295X.103.1.34)
5. McNeill D. 1992 *Hand and mind: what gestures reveal about thought*. Chicago, IL: Chicago University Press.
6. Kendon A. 2004 *Gesture: visible action as utterance*. Cambridge, UK: Cambridge University Press.
7. Dediu D, Levinson SC. 2013 On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. *Front. Psychol.* **4**, 397. (doi:10.3389/fpsyg.2013.00397)
8. Schepartz LA. 1993 Language and modern human origins. *Yearb. Phys. Anthropol.* **36**, 91–126. (doi:10.1002/ajpa.1330360607)
9. Jürgens U. 2002 Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.* **26**, 235–258. (doi:10.1016/S0149-7634(01)00068-9)
10. Simonyan K, Horwitz B. 2011 Laryngeal motor cortex and control of speech in humans.

Neuroscientist **17**, 197–208. (doi:10.1177/1073858410386727)

11. MacLarnon A, Hewitt G. 2004 Increased breathing control: another factor in the evolution of human language. *Evol. Anthropol.* **13**, 181–97. (doi:10.1002/evan.20032)
12. Fitch WT. 2009 Fossil cues to the evolution of speech. In *The cradle of language* (eds RP Botha, C Knight), pp. 112–134. Oxford, UK: Oxford University Press.
13. Bramble DM, Lieberman DE. 2004 Endurance running and the evolution of *Homo*. *Nature* **432**, 345–352. (doi:10.1038/nature03052)
14. Draper MH, Ladefoged P, Whitteridge D. 1960 Expiratory pressures and air flow during speech. *Br. Med. J.* **1**, 1837–1843. (doi:10.1136/bmj.1.5189.1837)
15. Tylor E. 1865 *Researches into the early history of mankind and the development of civilization*. London, UK: Murray.
16. Hewes G. 1973 Primate communication and the gestural origin of language. *Curr. Anthropol.* **14**, 65–84. (doi:10.1086/201401)
17. Kendon A. 1975 Gesticulation, speech, and the gesture theory of language origins. *Sign Lang. Stud.* **9**, 349–373. (doi:10.1353/sls.1975.0016)
18. Kendon A. 1991 Some considerations for a theory of language origins. *Man* **26**, 199–221. (doi:10.2307/2803829)
19. Armstrong DF, Wilcox S. 2007 *The gestural origin of language*. Oxford, UK: Oxford University Press.
20. Sterelny K. 2012 Language, gesture, skill: the co-evolutionary foundations of language. *Phil. Trans. R. Soc. B* **367**, 2141–2151. (doi:10.1098/rstb.2012.0116)
21. Corballis MC. 2002 *From hand to mouth: the origins of language*. Princeton, NJ: Princeton University Press.
22. Arbib MA. 2005 From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav. Brain Sci.* **28**, 105–124. (doi:10.1017/S0140525X05000038)
23. Gentilucci M, Corballis MC. 2006 From manual gesture to speech: a gradual transition. *Neurosci. Biobehav. Rev.* **30**, 949–960. (doi:10.1016/j.neubiorev.2006.02.004)
24. Corballis MC. 2009 The evolution of language. *Ann. NY Acad. Sci.* **1156**, 19–43. (doi:10.1111/j.1749-6632.2009.04423.x)
25. Call J, Tomasello M. 2007 *The gestural communication of apes and monkeys*. Manhaw, NJ: LEA.
26. Gruber T, Zuberbühler K. 2013 Vocal recruitment for joint travel in wild chimpanzees. *PLoS ONE* **8**, e76073. (doi:10.1371/journal.pone.0076073)
27. Schel AM, Townsend SW, Machanda Z, Zuberbühler K, Slocombe KE. 2013 Chimpanzee alarm call production meets key criteria for intentionality. *PLoS ONE* **8**, e76674. (doi:10.1371/journal.pone.0076674)
28. Liebal K, Pika S, Tomasello M. 2006 Gestural communication of orangutans (*Pongo pygmaeus*). *Gesture* **6**, 1–38. (doi:10.1075/gest.6.1.02lie)
29. Pollick AS, de Waal FB. 2007 Ape gestures and language evolution. *Proc. Natl Acad. Sci. USA* **104**, 8184–8189. (doi:10.1073/pnas.0702624104)
30. Rossano F. 2013 Gaze in conversation. In *The handbook of conversation analysis* (eds T Stivers, J Sidnell), pp. 308–29. Chichester, UK: Wiley-Blackwell.
31. Thorpe WH. 1966 Ritualization in ontogeny. I. Animal play. *Phil. Trans. R. Soc. Lond. B* **251**, 311–319. (doi:10.1098/rstb.1966.0015)
32. Scott-Phillips TC, Blythe RA, Gardner A, West SA. 2012 How do communication systems emerge? *Proc. R. Soc. B* **279**, 1943–1949. (doi:10.1098/rspb.2011.2181)
33. Clark HH. 2005 Coordinating with each other in a material world. *Discourse Stud.* **7**, 507–525. (doi:10.1177/1461445605054404)
34. Levinson SC. 1983 *Pragmatics*. Cambridge, UK: Cambridge University Press.
35. Sperber D, Wilson D. 1986 *Relevance: communication and cognition*. Cambridge, UK: Cambridge University Press.
36. Grice HP. 1989 *Studies in the way of words*. Cambridge, MA: Harvard University Press.
37. Schegloff EA. 2007 *Sequence organization in interaction: a primer in conversation analysis*, vol. 1. Cambridge, UK: Cambridge University Press.
38. Tomasello M. 2014 *A natural history of human thinking*. Cambridge, MA: Harvard University Press.
39. Levinson SC. 2000 *Presumptive meanings: the theory of generalized conversational implicature*. Cambridge, UK: MIT Press.
40. Levinson SC. 2013 Recursion in pragmatics. *Language* **89**, 149–162. (doi:10.1353/lan.2013.0005)
41. Mehl MR, Vazire S, Ramírez-Esparza N, Slatcher RB, Pennebaker JW. 2007 Are women really more talkative than men? *Science* **317**, 82. (doi:10.1126/science.1139940)
42. Stivers T *et al*. 2009 Universals and cultural variation in turn-taking in conversation. *Proc. Natl Acad. Sci. USA* **106**, 10 587–10 592. (doi:10.1073/pnas.0903616106)
43. Indefrey P, Levelt WJ. 2004 The spatial and temporal signatures of word production components. *Cognition* **92**, 101–144. (doi:10.3389/fpsyg.2011.00255)
44. DeLong KA, Urbach TP, Kutas M. 2005 Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* **8**, 1117–1121. (doi:10.1038/nn1504)
45. Pickering MJ, Garrod S. 2007 Do people use language production to make predictions during comprehension? *Trends Cogn. Sci.* **11**, 105–110. (doi:10.1016/j.tics.2006.12.002)
46. Magyari L, de Ruiter JP. 2012 Prediction of turn-ends based on anticipation of upcoming words. *Front. Psychol.* **3**, 376. (doi:10.3389/fpsyg.2012.00376)
47. Snow CE. 1977 The development of conversation between mothers and babies. *J. Child Lang.* **4**, 1–22. (doi:10.1017/S0305000900000453)
48. Méndez-Cárdenas MG, Zimmermann E. 2009 Duetting—a mechanism to strengthen pair bonds in a dispersed pair-living primate (*Lepilemur edwardsi*)? *Am. J. Phys. Anthropol.* **139**, 523–532. (doi:10.1002/ajpa.21017)
49. Snowdon CT, Cleveland J. 1984 'Conversations' among pygmy marmosets. *Am. J. Primatol.* **7**, 15–20. (doi:10.1002/ajp.1350070104)
50. Takahashi DY, Narayanan DZ, Ghazanfar AA. 2013 Coupled oscillator dynamics of vocal turn-taking in monkeys. *Curr. Biol.* **23**, 2162–2168. (doi:10.1016/j.cub.2013.09.005)
51. 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? *Sci. Rep.* **1**, 22. (doi:10.1038/srep00022)
52. Inoue Y, Sinun W, Yosida S, Okanoya K. 2013 Intergroup and intragroup antiphonal songs in wild male Muellers gibbons (*Hylobates muelleri*). *Interact. Stud.* **14**, 24–43. (doi:10.1075/is.14.1.03ino)
53. Geissmann T. 2002 Duet-splitting and the evolution of gibbon songs. *Biol. Rev.* **77**, 57–76. (doi:10.1017/S1464793101005826)
54. Haimoff EH. 1986 Convergence in the duetting of monogamous Old World primates. *J. Hum. Evol.* **15**, 51–59. (doi:10.1016/S0047-2484(86)80065-3)
55. Wilkinson R, Leudar I, Pika S. 2012 Requesting behaviors within episodes of active sharing. In *Developments in primate gesture research*, vol. 6 (eds S Pika, K Liebal), pp. 199–221. Amsterdam, The Netherlands: John Benjamins.
56. Argyle M, Cook M. 1976 *Gaze and mutual gaze*. Cambridge, UK: Cambridge University Press.
57. Bard KA, Myowa-Yamakoshi M, Tomonaga M, Tanaka M, Costall A, Matsuzawa T. 2005 Group differences in the mutual gaze of chimpanzees (*Pan troglodytes*). *Dev. Psychol.* **41**, 616–624. (doi:10.1037/0012-1649.41.4.616)
58. Nettle D, Cronin KA, Bateson M. 2013 Responses of chimpanzees to cues of conspecific observation. *Anim. Behav.* **86**, 595–602. (doi:10.1016/j.anbehav.2013.06.015)
59. Kobayashi H, Kohshima S. 2001 Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *J. Hum. Evol.* **40**, 419–435. (doi:10.1006/jhev.2001.0468)
60. Farroni T, Csibra G, Simion F, Johnson MH. 2002 Eye contact detection in humans from birth. *Proc. Natl Acad. Sci. USA* **99**, 9602–9605. (doi:10.1073/pnas.152159999)
61. Levinson SC. 2006 On the human 'interaction engine'. In *Roots of human sociality: culture, cognition and interaction* (eds NJ Enfield, SC Levinson), pp. 39–69. Oxford, UK: Berg.
62. Bates E. 1976 *Language and context: the acquisition of pragmatics*, vol. 13. New York, NY: Academic Press.
63. Tomasello M, Carpenter M, Liszkowski U. 2007 A new look at infant pointing. *Child Dev.* **78**, 705–722. (doi:10.1111/j.1467-8624.2007.01025.x)
64. Liszkowski U, Brown P, Callaghan T, Takada A, de Vos C. 2012 A prelinguistic gestural universal of

human communication. *Cogn. Sci.* **36**, 698–713. (doi:10.1111/j.1551-6709.2011.01228.x)

65. Carpendale JI, Carpendale AB. 2010 The development of pointing: from personal directedness to interpersonal direction. *Hum. Dev.* **53**, 110–126. (doi:10.1159/000315168)
66. Pika S, Liebal K, Tomasello M. 2005 Gestural communication in subadult bonobos (*Pan paniscus*): repertoire and use. *Am. J. Primatol.* **65**, 39–61. (doi:10.1002/ajp.20096)
67. Hobaiter C, Leavens DA, Byrne RW. 2013 Deictic gesturing in wild chimpanzees (*Pan troglodytes*)? Some possible cases. *J. Comp. Psychol*. **128**, 82–87. (doi:10.1037/a0033757)
68. Pika S, Mitani J. 2006 Referential gestural communication in wild chimpanzees (*Pan troglodytes*). *Curr. Biol.* **16**, R191–R192. (doi:10.1016/j.cub.2006.02.037)
69. Leavens DA, Hopkins WD, Bard KA. 1996 Indexical and referential pointing in chimpanzees (*Pan troglodytes*). *J. Comp. Psychol.* **110**, 346. (doi:10.1037/0735-7036.110.4.346)
70. Matthews D, Behne T, Lieven E, Tomasello M. 2012 Origins of the human pointing gesture: a training study. *Dev. Sci.* **15**, 817–829. (doi:10.1111/j.1467-7687.2012.01181.x)
71. Camaioni L, Perucchini P, Bellagamba F, Colonnesi C. 2004 The role of declarative pointing in developing a theory of mind. *Infancy* **5**, 291–308. (doi:10.1207/s15327078in0503_3)
72. Cochet H, Vauclair J. 2010 Pointing gestures produced by toddlers from 15 to 30 months: different functions, hand shapes and laterality patterns. *Infant Behav. Dev.* **33**, 431–441. (doi:10.1016/j.infbeh.2010.04.009)
73. Wittgenstein L. 1969 *On certainty*. Oxford, UK: Basil Blackwell.
74. Schilbach L, Wilms M, Eickhoff SB, Romanzetti S, Tepest R, Bente G, Shah NJ, Fink GR, Vogeley K. 2010 Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *J. Cogn. Neurosci.* **22**, 2702–2715. (doi:10.1162/jocn.2009.21401)
75. Levinson SC. 2003 *Space in language and cognition: explorations in cognitive diversity*. Cambridge, UK: Cambridge University Press.
76. De Vos C. 2012 Sign-spatiality in Kata Kolok: how a village sign language in Bali inscribes its signing space. PhD thesis, Radboud University Nijmegen, Nijmegen, The Netherlands.
77. Perniss P, Thompson RL, Vigliocco G. 2010 Iconicity as a general property of language: evidence from spoken and signed languages. *Front. Psychol.* **1**, 227. (doi:10.3389/fpsyg.2010.00227)
78. Mithen S. 1996 *The prehistory of the mind: the cognitive origins of art, religion and science*. London, UK: Thames and Hudson.
79. Marshack A. 1976 Some implications of the paleolithic symbolic evidence for the origin of language. *Curr. Anthropol.* **17**, 274–282. (doi:10.1086/201716)
80. Slobin DI. 1968 Antonymic phonetic symbolism in three natural languages. *J. Pers. Soc. Psychol.* **10**, 301–305. (doi:10.1037/h0026575)
81. Hinton L, Nichols J, Ohala J (eds) 1994 *Sound symbolism*. Cambridge, UK: Cambridge University Press.
82. Dingemanse M. In press. Expressiveness and system integration. On the typology of ideophones, with special reference to Siwu. *STUF—Language Typology and Universals*.
83. Haiman J. 1985 *Natural syntax: iconicity and erosion*. Cambridge, UK: Cambridge University Press.
84. McNeill D. 1985 So you think gestures are nonverbal? *Psychol. Rev.* **92**, 350–371. (doi:10.1037/0033-295X.92.3.350)
85. Holler J, Beattie G. 2002 A micro-analytic investigation of how iconic gestures and speech represent core semantic features in talk. *Semiotica* **142**, 31–69. (doi:10.1515/semi.2002.077)
86. Holler J, Beattie G. 2003 How iconic gestures and speech interact in the representation of meaning: are both aspects really integral to the process? *Semiotica* **146**, 81–116. (doi:10.1515/semi.2003.083)
87. Armstrong DF, Stokoe WC, Wilcox SE. 1995 *Gesture and the nature of language*. Cambridge, UK: Cambridge University Press.
88. Beattie G, Shovelton H. 2002 An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *Br. J. Psychol.* **93**, 179–192. (doi:10.1348/000712602162526)
89. Stokoe WC. 2001 *Language in hand: why sign came before speech*. Washington, DC: Gallaudet University Press.
90. Arbib MA. 2012 *How the brain got language: the mirror system hypothesis*. Oxford, UK: Oxford University Press.
91. Fitch WT. 2010 *The evolution of language*. Cambridge, UK: Cambridge University Press.
92. Stout D. 2011 Stone toolmaking and the evolution of human culture and cognition. *Phil. Trans. R. Soc. B* **366**, 1050–1059. (doi:10.1098/rstb.2010.0369)
93. Hockett C. 1960 The origin of speech. *Sci. Am.* **203**, 88–111. (doi:10.1038/scientificamerican0960-88)
94. Perniss P, Vigliocco G. 2014 The bridge of iconicity: from a world of experience to the experience of language. *Phil. Trans. R. Soc. B* **369**, 20130300. (doi:10.1098/rstb.2013.0300)
95. Jackendoff R. 1999 Possible stages in the evolution of the language capacity. *Trends Cogn. Sci.* **3**, 272–279. (doi:10.1016/S1364-6613(99)01333-9)
96. Marentette P, Nicoladis E. 2012 Does ontogenetic ritualization explain early communicative gestures in human infants? In *Developments in primate gesture research*, vol. 6 (eds S Pika, K Liebal), pp. 33–54. Amsterdam, The Netherlands: John Benjamins.
97. Jaffe J, Beebe B, Feldstein S, Crown CL, Jasnow MD, Rochat P, Stern DN. 2001 Rhythms of dialogue in infancy: coordinated timing in development. *Monogr. Soc. Res. Child Dev.* **66**, i–viii, 1–132.
98. Garvey C. 1984 *Children's talk*. Cambridge, MA: Harvard University Press.
99. Aflalo TN, Graziano MS. 2006 Possible origins of the complex topographic organization of motor cortex: reduction of a multidimensional space onto a two-dimensional array. *J. Neurosci.* **26**, 6288–6297. (doi:10.1523/JNEUROSCI.0768-06.2006)
100. Meier JD, Aflalo TN, Kastner S, Graziano MS. 2008 Complex organization of human primary motor cortex: a high-resolution fMRI study. *J. Neurophysiol.* **100**, 1800–1812. (doi:10.1152/jn.90531.2008)
101. Graziano MS, Aflalo TN. 2007 Mapping behavioral repertoire onto the cortex. *Neuron* **56**, 239–251. (doi:10.1016/j.neuron.2007.09.013)
102. Vauclair J. 2004 Lateralization of communicative signals in nonhuman primates and the hypothesis of the gestural origin of language. *Interact. Stud.* **5**, 363–384. (doi:10.1075/is.5.3.04vau)
103. Meguerditchian A, Vauclair J. 2006 Baboons communicate with their right hand. *Behav. Brain Res.* **171**, 170–174. (doi:10.1016/j.bbr.2006.03.018)
104. Meguerditchian A, Vauclair J, Hopkins WD. 2010 Captive chimpanzees use their right hand to communicate with each other: implications for the origin of the cerebral substrate for language. *Cortex* **46**, 40–48. (doi:10.1016/j.cortex.2009.02.013)
105. Fitch WT, Braccini SN. 2013 Primate laterality and the biology and evolution of human handedness: a review and synthesis. *Ann. NY Acad. Sci.* **1288**, 70–85. (doi:10.1111/nyas.12071)
106. Iverson JM, Goldin-Meadow S. 1998 Why people gesture when they speak. *Nature* **396**, 228. (doi:10.1038/24300)
107. Iverson JM, Goldin-Meadow S. 2001 The resilience of gesture in talk: gesture in blind speakers and listeners. *Dev. Sci.* **4**, 416–422. (doi:10.1111/1467-7687.00183)
108. Rizzolatti G, Fadiga L, Gallese V, Fogassi L. 1996 Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* **3**, 131–141. (doi:10.1016/0926-6410(95)00038-0)
109. Rizzolatti G, Arbib MA. 1998 Language within our grasp. *Trends Neurosci.* **21**, 188–194. (doi:10.1016/S0166-2236(98)01260-0)
110. Rizzolatti G, Craighero L. 2004 The mirror-neuron system. *Annu. Rev. Neurosci.* **27**, 169–192. (doi:10.1146/annurev.neuro.27.070203.144230)
111. McNeill D. 2012 *How language began: gesture and speech in human evolution*. Cambridge, UK: Cambridge University Press.
112. Kendon A. 1972 Some relationships between body motion and speech. An analysis of an example. In *Studies in dyadic communication* (eds A Siegman, B Pope), pp. 177–210. Elmsford, NY: Pergamon Press.
113. Kendon A. 1980 Gesticulation and speech: two aspects of the process of utterance. In *The relationship of verbal and nonverbal communication* (ed. MR Key), pp. 207–227. The Hague, The Netherlands: Mouton and Co.
114. Kendon A. 2000 Language and gesture: unity or duality. In *Language and gesture* (ed. D McNeill), pp. 47–63. Cambridge, UK: Cambridge University Press.

115. McNeill D. 2005 *Gesture and thought*. Chicago, IL: University of Chicago Press.
116. Kendon A. 1985 Some uses of gesture. In *Perspectives on silence* (eds D Tannen, M Saville-Troike), pp. 215–234. Norwood, NJ: Ablex.
117. Gershkoff-Stowe L, Goldin-Meadow S. 2002 Is there a natural order for expressing semantic relations? *Cogn. Psychol.* **45**, 375–412. (doi:10.1016/S0010-0285(02)00502-9)
118. Goldin-Meadow S. 2003 *Hearing gesture: how our hands help us think*. Cambridge, MA: Harvard University Press.
119. Skipper JI, Goldin-Meadow S, Nusbaum HC, Small SL. 2007 Speech-associated gestures, Broca's area, and the human mirror system. *Brain Lang.* **101**, 260–277. (doi:10.1016/j.bandl.2007.02.008)
120. Willems RM, Hagoort P. 2007 Neural evidence for the interplay between language, gesture, and action: a review. *Brain Lang.* **101**, 278–289. (doi:10.1016/j.bandl.2007.03.004)
121. Willems RM, Özyürek A, Hagoort P. 2007 When language meets action: the neural integration of gesture and speech. *Cerebral Cortex* **17**, 2322–2333. (doi:10.1093/cercor/bhl141)
122. Bavelas JB, Gerwing J, Sutton C, Prevost D. 2008 Gesturing on the telephone: independent effects of dialogue and visibility. *J. Mem. Lang.* **58**, 495–520. (doi:10.1016/j.jml.2007.02.004)
123. Holler J, Tutton M, Wilkin K. 2011 Co-speech gestures in the process of meaning coordination. In *Proc. 2nd GESPIN—Gesture and Speech in Interaction Conf., Bielefeld, 5–7 September 2011*. See http://hdl.handle.net/11858/00-001M-0000-0012-1BB3-D.
124. Alibali MW, Heath DC, Myers HJ. 2001 Effects of visibility between speaker and listener on gesture production: some gestures are meant to be seen. *J. Mem. Lang.* **44**, 169–188. (doi:10.1006/jmla.2000.2752)
125. Jacobs N, Garnham A. 2007 The role of conversational hand gestures in a narrative task. *J. Mem. Lang.* **56**, 291–303. (doi:10.1016/j.jml.2006.07.011)