Review

Visual bodily signals as core devices for coordinating minds in interaction

Judith Holler[1,2]

[1]Max-Planck-Institut für Psycholinguistik, Nijmegen, The Netherlands
[2]Donders Centre for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

JH, 0000-0003-0671-6651

The view put forward here is that visual bodily signals play a core role in human communication and the coordination of minds. Critically, this role goes far beyond referential and propositional meaning. The human communication system that we consider to be the explanandum in the evolution of language thus is not spoken language. It is, instead, a deeply multimodal, multilayered, multifunctional system that developed—and survived—owing to the extraordinary flexibility and adaptability that it endows us with. Beyond their undisputed iconic power, visual bodily signals (manual and head gestures, facial expressions, gaze, torso movements) fundamentally contribute to key pragmatic processes in modern human communication. This contribution becomes particularly evident with a focus that includes non-iconic manual signals, non-manual signals and signal combinations. Such a focus also needs to consider meaning encoded not just via iconic mappings, since kinematic modulations and interaction-bound meaning are additional properties equipping the body with striking pragmatic capacities. Some of these capacities, or its precursors, may have already been present in the last common ancestor we share with the great apes and may qualify as early versions of the components constituting the hypothesized interaction engine.

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

1. The richness of utterance: beyond iconicity, towards multimodal pragmatics

When wondering about the role of gesture in the evolution of human language, opinions have been plentiful and diverse. Some consider its role as negligible if relevant at all (e.g. [1,2]). Scholars who ascribe a central role of gestures to how we have acquired language fall, broadly speaking, into two camps: those arguing for 'gesture-supplanted-by-speech' scenarios of language evolution, where gesture fulfilled a bridging function and its relevance largely withered away once a fully fledged linguistic system had been acquired, eventually leading to the speaking species that we are (e.g. [3–8]); and those who argue for language having evolved as a multimodal system, with the vocal and gestural modalities intertwined from the very beginning, and playing an integral role in communication also in modern human language (e.g. [9–19]). According to the former, spoken language is the evolutionary end product and the ultimate explanandum, while the latter considers this to be the unified speech-gesture system.

My aim here is to further corroborate this latter view by highlighting the substantial contribution visual bodily signals make to human pragmatic communication. Much of the focus to date has been on gestural iconicity and the encoding of semantic information as part of communicating referential meaning or propositional content when discussing the integral role of gesture in human language, *de novo* language creation and evolution (e.g. [12,13,20–37]). This may have led us to significantly underestimate the role of the body in the process of

laying the foundations for the emergence of human language. Conveying propositional information, arguably, is one of the core building blocks of human communication. Studies focusing on iconicity (and deixis) in the emergence of language thus have revealed fundamental insights into the nature of human language, the evolution of symbolic communication, its learnability, conventionalization, systematization and how we transmit thoughts that go beyond the here and now. My point is certainly not that such a focus is to be argued with—quite the opposite; work on the role of gestural iconicity in language evolution and emergence has undisputable merit. However, the strong emphasis on iconicity, and consequentially on semantics, have distracted us from devoting attention to the role of visual bodily signals in pragmatic communication. Broadening our focus of enquiry will help us to recognize the full scope of the body's involvement in human communication.

This entails considering its contribution to a host of key ingredients required to make communication successful. Such an approach will facilitate connecting behavioural observations with theoretical frameworks recently put forward, such as the Interaction Engine Hypothesis [38,39]. The notion of the interaction engine refers to a set of behavioural and cognitive predispositions that constitute the roots of the advanced social interaction capabilities of the human species. These include signalling that an act is meant to communicate and that it is recognized as such, and that it can be deciphered for what it is supposed to mean—not just in terms of the semantic 'bits' of information encoded but in terms of what the speaker intends to achieve with communicating them at a particular moment in time (i.e. the pragmatic meaning and illocutionary force needs to be derived). An essential prerequisite for this is that speakers tailor messages to their specific addressees' requirements. Equally important are mechanisms that allow addressees to signal that they have understood or have trouble understanding, as well as the interactional structures that facilitate these processes, such as socially contingent actions and a turn-taking system. Section 2, therefore, reviews evidence which demonstrates the contribution of visual bodily signals to the following core pragmatic processes:

(i) signalling the intent to communicate;
(ii) communicating specific intentions and pragmatic meanings;
(iii) tailoring information to an interlocutor's knowledge (recipient design, common ground);
(iv) signalling understanding (grounding);
(v) signalling and repairing trouble in understanding; and
(vi) calibrating communicative action to the interaction in progress, moment-by-moment (coordination of sequential actions, turn-taking).

To gain these insights, we require an approach that goes beyond a focus on the purely semantic information in the visual modality. This means: (i) focusing on the pragmatic functions of gestures even when these themselves are iconic or deictic in nature; (ii) considering manual gestures other than iconic and pointing gestures, such as those defined as pragmatic and interactive gestures; (iii) analysing visual signals carried by articulators other than the hands, especially the face, and the combination of signals into visual–visual or multimodal composites; (iv) measuring the contribution to communication not only based on depiction and deixis (the what), but also based on analyses of movement

kinematics and structural aspects of signal form (the how); and (v) applying a socially contextualized perspective able to capture the detailed interactional and reciprocal embeddings of visual bodily signals.

Such an approach may also offer new avenues for cross-species comparisons (§3). As noted by Fröhlich et al. [9], many of the gestures typically theorized about have rather little in common with those used by non-human primates. An interactionally-grounded approach that puts the pragmatics of multimodal behaviour centre-stage may thus offer new avenues for advancing our understanding of how we coordinate minds in human communication and how this ability may have evolved (§4).

2. Visual bodily signals as coordination devices in human communication

This section aims to illustrate how broadening our focus in the manner described above facilitates appreciating the contribution of visual bodily signals to core components of human communication—that is, their function as coordination devices in joint actions [19]. For each of the six components listed in the previous section (here constituting §§2a–2f), some evidence for the contribution of visual bodily signals is given, ranging from manual gestures, torso movements, head gestures and eye gaze, to facial signals, as well as composites (i.e. signal combinations) involving more than one type of articulator or method of signalling [19].

One disclaimer is necessary at this juncture: what follows is a non-exhaustive summary of some of the ways in which the body forms part of a selection of processes key to human communication. In its brevity, it does not do justice to the full breadth of relevant extant research, and there will certainly be omissions owing to the author's own shortcomings. Nor does it do justice to the rich insights that can be gained from microscopic, detailed and holistic analyses of situated human interaction yielded by conversation analytical techniques, since much more in-depth treatments are necessary to reveal them. Nevertheless, I hope it will become clear from this brief review how manifold the functions of visual bodily gestures are and how vital their contribution is to the pragmatic processes that they form an inherent part of. [1]

(a) Visually signalling the intent to communicate

To facilitate coordination, interlocutors produce behaviours in such a way that allows them to be identified as being intended to communicate, thus contrasting them with behaviours that are not.

(i) Manual gestures

In [40], participants performed manual gestures and actions either for an addressee to learn from them or in a more self-serving context. Movements in the more communicative context were larger and more complex in structure than less communicative movements. Pointing gestures, too, differ kinematically depending on whether they are more or less communicatively intended, with communicative pointing gestures being produced more slowly (duration from onset to the maximum point of extension) and with less velocity than less communicative pointing gestures, and being held longer in their pointing position [41]. Since these studies
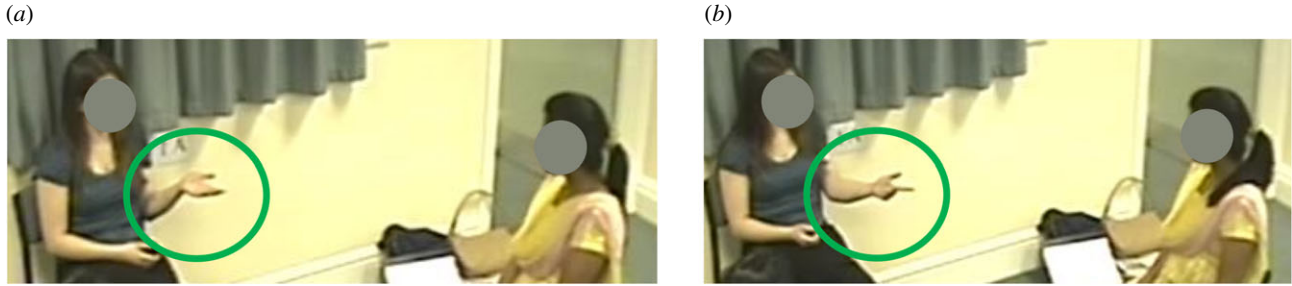
Figure 1. Example of an OHS (open hand supine, or palm-up-open-hand) gesture (*a*), and example of a pointing gesture directed at the addressee as a way of marking information constituting mutually shared knowledge which the speaker is referring to verbally (common ground pointing gesture (CG-point)) (*b*).

manipulated the extent to which the signals were relevant for addressees, there is reason to believe that kinematic variation of this kind is produced by speakers with the intent that their interlocutor perceives the behaviour as ostensive and thus as communicatively relevant and intended [42].

(ii) Torso movements

Looking beyond the hands, body orientation is an important give-away as to the communicative intent of another agent. The body being oriented towards someone signals address, the intent to initiate and maintain interaction [43–47].

(iii) Gaze

Person-directed gaze is another signal conveying the intent to communicate or initiate interaction [40,48–53] (and person-directed face orientation can have a similar function).

(iv) Composites

Gaze direction also interacts with the directional positioning of other bodyparts (torso, legs, head) resulting in complex gestalts. These are core to signalling the intent to engage in or maintain interaction and the dynamic coordination of participation frameworks [43,54–58]. Speaker gaze directed towards co-speech gestures also highlights their communicative intendedness for the addressee [59–63].

Thus, one of the most primordial of processes—signalling the intent to communicate and initiating interaction—is done bodily in various ways, using a range of different articulators, their orientation, and kinematic variations.

(b) Communicating pragmatic meaning and illocutionary force with visual bodily signals

The ultimate prerequisite for communication to be successful is for an interlocutor to recognize the message as what the speaker intends it to mean [64,65]. A look at conversation, the natural home of human language use, highlights the pragmatic meaning contribution of visual bodily signals.

(i) Manual gestures

Pointing gestures can communicate indirect requests both in the presence and in the absence of speech [66,67]. Manual gestures also contribute to communicating illocutionary force [65] through palm orientation. Kendon [68] proposed the notion of 'gesture families', i.e. groupings of gestures which share kinesic features (such as hand shape, orientation and movement pattern) and are related in meaning and pragmatic function. One example is the 'open hand supine' (OHS) family consisting

of gestures performed with an open, flat hand and the arm in the supine position (also 'palm-up-open-hand gestures' [69]) (figure 1). OHS gestures are associated with the themes of giving, offering, showing/presenting, requesting or receiving—i.e. the metaphorical embodiment of a non-present entity being placed on the flat palm (see also [70–72]). OHS gestures may be directed towards the space in front of the speaker (e.g. to offer an idea or explanation), towards the interlocutor (e.g. to request a response) or to entities or locations in the environment (e.g. to present or display a person or object). Another group of manual gestures is characterized by the palm facing down or forwards, typically combined with lateral motion or motion away from the body, clearing the body space from unwanted entities in a metaphorical way [68,73,74]. Such gestures act as embodied forms of negation, declination, refusal and negative assessments.

These examples illustrate manual gestures contributing to the illocutionary force of utterances by clarifying if an utterance should be interpreted as an offer, a request, a declination and such like. Critically, the illocutionary force of such recurrent gesture forms often remains implicit when considering just the accompanying speech [75], thus highlighting gestures' core pragmatic contribution to the conversation. Aside from illocutionary force, manual gestures contribute to pragmatics also by adding emphasis and contrast to elements in speech (e.g. beats [34]), something Kendon referred to as the 'parsing' function of gestures [68].

(ii) Torso movements

The body's contribution to communicating pragmatic meaning is just as evident when we look beyond manual gestures. Research in the domain of human sensorimotor communication and joint action has revealed that point-light displays of body movements (in the absence of speech) allow observers to decode information about specific intentions, such as ordering someone to do something, requesting or offering an object [76,77]. In conversational context, torso movements were found to differ between categories of social actions, with questions expressing a stance or sentiment being accompanied by larger torso movements than requests for information, for example. This may be related to the act of distancing oneself from or showing agreement with certain attitudes [78].

(iii) Head gestures

The fundamental pragmatic functions of affirming and negating are often accomplished through highly conventionalized movements of the head, such as nods and shakes, and addressees use head nods to signal agreement and to confirm receipt and understanding [79,80]. These may

perhaps be considered the prototypical ones, but the rich repertoire of complex head movements in conversation has long been acknowledged, and the wide range of pragmatic functions they fulfil carefully documented (see e.g. [81]; Heylen [82,83]; Wagner *et al.* [84]).

(iv) Gaze

Gaze direction, too, plays a significant role. For example, gaze aversion in conversation can mark dispreferred responses (such as declining an offer) and may occur considerably earlier than the verbal response itself [85]. Also, rolling eyes can mark statements as ironic [86].

(v) Facial signals

One of the main bearers of pragmatic information during conversation is the face. For many decades, facial expressions were the focus of emotion research, but recent years have seen this starting to shift. Eyebrow movements are a frequent question marker in spoken and signed languages alike [87–89]. Frowns, too, are a common occurrence [90] and can foreshadow trouble in understanding [91] and form a contrast with eyebrow raises to mark different forms of clarification requests [92]. Squints also often mark questionhood [93,94]. Eyebrow raises, eyebrow flashes and expressive mouth movements have been associated with utterances to be interpreted as sarcasm and irony [86,95].

(vi) Composites

Pragmatic manual gestures often combine with other visual signals, such as manual negation gestures co-occurring with headshakes, and palm-up gestures with shoulder shrugs and facial signals [68]. The combination of signals acting as gestalts is particularly prevalent in the face. Squints, for example, frequently co-occur with eyebrow frowns in marking questionhood [90]. The 'not-face' consists of an interplay of several signals even, most commonly eyebrow frown and pushed up chin muscles, plus pulled back corners of the mouth or pressed lips, expressing negative moral judgement [96]. The 'facial shrug' involves one corner of the mouth being pulled back plus an eyebrow flash (but mouth and eyebrow components can also occur on their own), translating as 'I don't know', 'oh well' or expressions of dislike [97,98]. The 'thinking face' involves a blank, puzzled or concentrated face with furrowed brows or raised eyebrows combined with averted gaze [98,99], typically accompanying word searches.

In summary, a wide range of visual bodily signals contributes information that specifies the pragmatic meaning and illocutionary force of an utterance. Importantly, the information the body contributes is often not encoded in the verbal modality and frequently occurs before or at the utterance beginning, thus underlining their communicative import.

(c) Producing recipient-designed utterances with visual bodily signals

The notion of recipient or audience design in communication refers to speakers' sensitivity to the co-participants in an interaction and to the adaptation in communicative behaviour this leads to [100,101]. Crucially, this adaption entails speakers designing their utterances such that they match the knowledge, beliefs and assumptions of the participants

they want their message to reach. Designing utterances with the intended recipient(s) in mind is a fundamental pragmatic process, and successful communication depends on it [7,19]. There is plenty of evidence that speakers adapt their signalling in response to environmental factors (such as noise or visibility), but even more convincing is the versatile ability of manual gestures in crafting messages that are tailored to fit addressees' informational needs.

(i) Manual gestures

A number of studies have shown that speakers draw on the gestural modality in a manner that takes into account common ground—that is, the knowledge, beliefs and assumptions that interlocutors mutually share [19]. Typically, this involves a reduction in gesture frequency (and sometimes gesture rate [102]) and information content [103–105], when common ground between interlocutors exists (but see [106]). Speakers also mark reference to shared knowledge through gesture form. Gestures are often smaller [104,107,108], less precise [103,108] or occur lower in gesture space [109] when referring to information that already is in the interlocutors' common ground (with some exceptions [102]). Similarly, speakers use gesture duration to contrast given with new information in referential communication (see [110], and references relating to other ways of gestures marking information status therein).

Further, some non-iconic, interactive palm-up gestures serve the purpose of marking what someone is saying as being a citation of something said earlier [111]. And interactive pointing gestures towards the addressee (figure 1) can mark information as references to common ground, often without explicit verbal reference to the information's common ground status [112].

(ii) Facial signals

Much less is known about facial signalling of common ground. In some signed languages, squints can mark information as common ground and its degree of accessibility [113]. Whether squints and other facial signals function similarly together with spoken language is a highly interesting question for future research.

In summary, the fundamental process of designing messages appropriate for specific recipients is a multimodal phenomenon. Much of this process rests on kinematic form changes in gesture being pragmatically instrumental, in addition to or *in lieu* of adaptions in speech, and it goes beyond the use of iconic manual gestures.

(d) Grounding information with the body

Achieving mutual understanding in interaction is foundational to human communication. Core to this is the process of grounding, i.e. reaching the mutual belief that a conversational contribution has been understood by all participants [114]. One common way of grounding information is with minimal verbal responses, such as 'mhm' or 'yeah', signalling understanding and allowing a conversation to proceed [115]. Producing a next speaking turn that pragmatically fits the previous one, too, signals understanding, as does responding with a fitting instrumental action if warranted by the context [19,116,117].

(i) Manual gestures

Interlocutors use iconic depictions to signal understanding through alignment in gesture form ([118–122], see [18] for a theoretical account linking such behavioural with cognitive alignment). For example, a speaker may refer to a man holding a big grocery bag by depicting the action of holding such a large bag themselves. Their interlocutor may reuse that very gesture later on in the dialogue to refer to the same character. Such repeated gestures give the clear impression that interlocutors have mutually shared understanding of whom they are referring to. The gestures may even do so more unambiguously than verbal references, especially when words are not aligned, are not produced at all, or change in order to propose a different verbal conceptualization, where the gestures then provide the 'visual anchor' of alignment [118, p. 143]. Reusing an interlocutor's gestures may be a particularly efficient form of grounding, and it is deemed intentional, thus differentiating this process from other forms of behaviour matching such as foot tapping [118]. Note that repeated gestures are not about semantic information being conveyed or iconically mapped afresh—the reuse of the same gestural form is critical (with some leeway [118,119]) so that it is recognized as a reference to something said earlier by the other interlocutor. The gestures thus obtain their grounding function through repetition across interlocutors.

Manual gestures with interactive functions, typically performed as palm-up gestures, also play a role in grounding, in two different ways [111]. One function is seeking a response from the addressee that would indicate understanding (thus functioning similarly to the words 'you know?'). The other relates to palm-up interactive gestures that occur in a different sequential position: Bavelas et al.'s [123] micro-analysis of the grounding process has exposed not just two steps (speaker's contribution and addressee's signal of understanding) but also a third, namely the speaker acknowledging the addressee's understanding, claimed to be a crucial step in establishing *mutually* shared understanding. Interactive palm-up gestures in this third-step-position thus paraphrase as 'I see that you understood me' [123, p. 397]. Critically, the information conveyed by the interactive grounding gestures is typically not verbalized.

(ii) Head gestures

Head gestures are incredibly frequent in conversation, especially as part of the addressee's signal repertoire [79,84,124]. Malisz et al. [79] found that the majority of listener head gestures were nods (often consisting of multiple nodding cycles), used for different pragmatic functions, such as displaying continued attention and confirming understanding. The form of the head gestures seemed to disambiguate between those two functions, with head movements consisting of an upward motion followed by a downward motion (i.e. reverse nods) being more specifically used for confirming understanding. Both of these types of head movements occurred in the second position, i.e. as the addressee's response to information conveyed by the speaker. Head gestures also play a prominent role in the third position of grounding sequences, i.e. as acknowledgements of the addressee signalling understanding, thus functioning similarly to the third position palm-up interactive gestures described above [123].

(iii) Facial signals

Smiles, too, can function as grounding signals [125] and frequently form part of the three-step grounding process (see [123] for examples). Eye blinks, too, can contribute to the process of grounding [126,127]. From what we know so far, blinks occur mainly in the second position, i.e. as addressees' responses to speakers' utterances, especially during extended speaking turns [127]. Importantly, the subtle kinematics of eye-blinks appear to make a difference, with long blinks (duration greater than 410 ms) functioning more clearly as grounding signals than short blinks, at least in experimental settings [126]. Eyebrow flashes, too, can function as signals for grounding. In Yélî Dnye (a Papuan language isolate spoken on Rossel Island), eyebrow flashes are conventionalized, meaning 'yes, continue', or function as an affirmative response [128,129].

(iv) Composites

Noteworthy is the frequent co-occurrence of blinks with nods [126,127], or blinks with eyebrow flashes [129], which may lead to an amplified communicative effect. Other facial signals may easily combine into more powerful grounding devices, too, such as shoulder shrugging with a cute face or rolling eyes [123]. Addressee facial 'displays' can also serve the grounding process, such as wincing in response to someone describing a painful experience, or a facial expressions of concern or fear in response to elements of close call stories [130,131]. Such 'specific listener responses' [131] signal understanding of the message (as well as empathy), and possibly more unambiguously so than 'generic listener responses' (e.g. 'mhm', nods).

While gaze may not function as a grounding signal in itself, the significance of visual signals for grounding is underlined by the fact that speakers actively elicit them by looking at the addressee, thus creating brief time windows of mutual gaze during which the addressee responses occur [131]. Thus, manual gestures, head movements and facial signals play an important role in the process of grounding, a *sine qua non* in conversational interaction [19].

(e) Visual bodily signals for initiating repair and responding to repair initiations

Despite being able to draw upon mechanisms for grounding that are beneficial for achieving mutual understanding in conversation, troubles in understanding are common [132]. Causes include trouble in hearing and in grasping the actual content of an utterance. Part of the process of achieving mutual understanding is repair mechanisms for such instances in which communication goes awry [133]. Vocally, interlocutors can refashion their own utterances to undertake repairs (self-repair), or issue requests for repair by asking 'huh', 'what?', 'who?' and so forth (other-initiated repair).

(i) Manual gestures

In face-to-face interaction, visual signals form an integral part of the repair strategies interlocutors have at their disposal [134]. For example, speakers' rate of self-repair and time spent manually gesturing is positively correlated [135], and when encountering dysfluencies associated with miscommunications, speakers produce manual gestures higher up in gesture space [136], presumably to foreground the gestural information in resolving the communicative trouble. When attempting to self-repair, speakers also use gestures to elicit helping responses from their addressees [111,137,138], usually in the form of interactive gestures expressing that the speaker is in some kind of
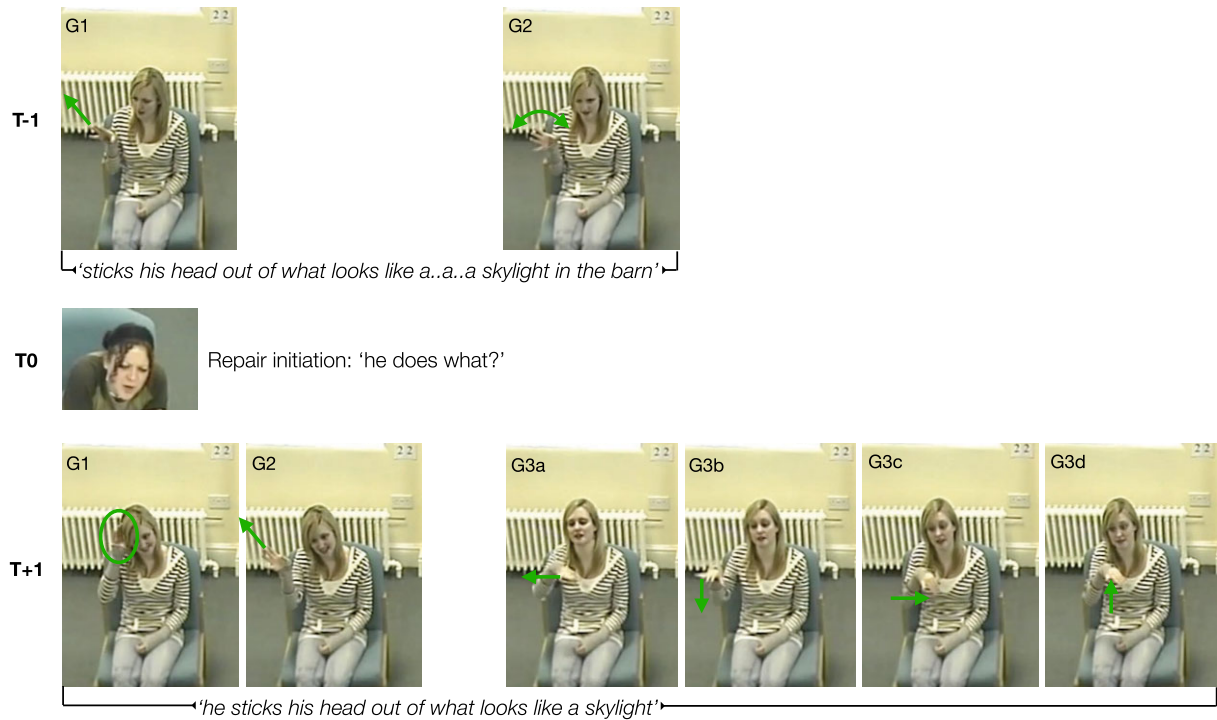
Figure 2. Illustration of a multimodal repair sequence. T − 1, trouble source; T0, repair initiation; T + 1, repair solution [139]; G, gesture (G3a − G3d = gestural subcomponents). The speaker says (T − 1) '[*he goes up another level and*] *sticks his head out of what looks like a a a skylight in the barn*' to describe a little boy who climbed up to the top of a barn looking out of a window in the roof. In response to the addressee initiating repair (T0) by asking '*he does what?*', the original speaker says (T + 1) '*he sticks his head out of what looks like a skylight*'. Note that the speech at T − 1 and T + 1 differs marginally, the change consisting mainly of the last three words (in the barn) having been dropped and the utterance not being dysfluent. The accompanying gestures, however, do most of the repair work. The utterance at T − 1 includes two gestures. The first involves the flat hand, palm down, moving diagonally upwards and forwards depicting the boy's head moving up through the roof window. The second gesture is a pragmatic one and involves the hand, palm facing down and fingers spread, turning in the wrist (from left to right, 3×) which seems to convey uncertainty about whether the term 'skylight' is the best fit. Both gestures get revised as part of the utterance in T + 1: the first gesture now involves the hand, held in the same way as before, moving first to the speaker's own head before then moving upwards and forwards. This added movement clarifies that the upward moving hand is meant to depict a head moving up. The second gesture changes from a pragmatic to an iconic one, now outlining the square shape of the opening in the roof, presumably to illustrate its window-like features.

communicative trouble (e.g. palm-up gestures or rubbing index finger and thumb together, held up in interactive space to invite a word suggestion).

When responding to repair initiations, the manual modality also plays a role. While speakers may not necessarily gesture more in response to clarification requests [61,138], they do adjust their manual gestures to address the recipient's informational requirements. In the example above (figure 2) the speaker responds multimodally to a repair initiation from the addressee. Verbally, the speaker actually more or less repeats what was said (i.e. the trouble source). The burden of the repair work is accomplished gesturally by adjusting the form, and consequently the meaning, of the two gestures that formed part of the original utterance. In addition to adaptations in the semantic information encoded, manual gestures speakers use to respond to clarification requests can also differ in their kinematic realization; typically they then become more precise, larger or are performed in more visible areas of gesture space [61]. Only a consideration of the situated, pragmatic use and function of the gestures and the change in gestural form reveals such contributions to the repair process.

Addressees also make substantial use of manual gestures when *initiating* the repair. Healey *et al.* [138] showed that addressees' hand movements not only increased and became faster, but their behaviour also changed from more generic feedback (e.g. nods) to content specific manual gestures when issuing clarification requests. Also, Mortensen

[140] described addressees using a specific type of iconic gesture, a cupped hand placed behind the ear, pragmatically in order to initiate repairs.

(ii) Torso movements

Movements of the torso, too, accompany repair initiations. Addressees may lean forward to intervene with questions, seek clarification, holding the lean until a response or clarification has been obtained [141,142]. Forward torso leans also occur when speakers attempt to provide repair solutions, especially as pursuits (i.e. when understanding in response to initial repair attempts is not displayed) [143].

(iii) Head movements

Speakers producing self-repairs nod more, and the primary and secondary (in multi-party conversation) addressee(s) also increase their nodding to help solve the trouble in utterance formulation [144]. Addressees also use head movements to *initiate* repair, including lateral turn or tilts and forward movements of the head to signal problems in understanding and to request clarification [142,145,146].

(iv) Facial signals

The face plays a prevalent role in addressees' repair initiations also, particularly eyebrows [92,146]. There even appears to be a fine-grained association between eyebrow raises and

open repair initiations (e.g. 'huh?', 'what?') and eyebrow frowns with restricted repair initiations (e.g. 'he did what?'), thus hinting at intricate connections between the precise form of eyebrow movement and the type of response desired [92]. In some cultures, eyebrow flashes are used to confirm a speaker's successful repair [129].

(v) Composites

Eyebrow movements can form part of more complex facial displays for repair initiations, such as raised eyebrows combined with averted gaze and corners of the mouth pulled down [146].

In addition to the presence of the various bodily movements being associated with repair, their *absence*, too, can carry a signalling function. When addressees initiate repair, they may temporarily pause any bodily movement (including gaze, head, torso lean, eyebrows or manual movements) keeping them in a static hold position [147], which can result in something like the 'freeze-look' [148]. Interlocutors respond to the embodied absence of responses, for instance by expanding and refashioning their preceding turn to repair the communication problem [149].

The above shows that the body may be used in various ways to address problems in understanding, including addressees signalling trouble in understanding and speakers repairing troubles in speaking and understanding.

(f) Coordinating sequential interaction with visual bodily signals

The vehicle for exchanging contributions in conversation is by taking turns at talking [101]. While producing speech in an alternating fashion may seem trivial at first sight, a closer look at what it entails reveals the complexities of the process. The gaps between consecutive turns are so short that the next speakers must start planning their contribution in parallel with the current turn that is still underway in order to launch it on time. This requires the prediction of upcoming content, as well as its approximate ending [150]. In face-to-face interaction, the body is very much involved in these processes.

(i) Manual gestures

Gestures conveying meaning are timed such that they significantly precede corresponding speech. This endows them with the power to project turn content, which may help early next turn planning (i.e. what to say) [151]. Further, manual gestures contribute to the alternation and timing of turns (i.e. when to speak). Duncan [152] showed that when the hands are actively involved in gesticulation they act as strong signals suppressing potential next speakers from taking the floor, while not gesturing yields the floor. In line with the latter, there is evidence that gesture retractions are associated with turn transitions rather than turn continuations [153], and retractions prior to turn end lead to shorter transitions than gestures retracted after turn completion [154]. Interestingly, certain gestures do not act as attempt-suppressing signals but in the opposite way, such as palm-up gestures offering the turn [68,71,111,155]. Moreover, participants in the conversation can gesturally signal their intention to speak next, such as by pointing to targets in a shared visual space while another speaker's turn is still underway [156].

(ii) Gaze

Some studies suggest that speakers look at addressees when they wish to hand over the turn, or that such behaviour solicits turn-taking attempts from the addressee [152,157]. In multiparty interactions, gaze direction can be an effective signal for the next speaker selection [51,158]. However, others have argued that gaze is organized at the level of sequences of action, such as gaze being directed at the addressee during the initiation of a question and averted only on completion of the response, rather than the end of the question turn [159,160]. Gaze thus seems to play a crucial role in the sequential coordination of talk in face-to-face interaction and may do so at various levels.

(iii) Composites

Visual bodily signals may also combine in intricate ways. Duncan proposed that the more floor yielding signals co-occur, the more likely are attempts from the addressee to take the floor [152], but that the floor-keeping signal of gesturing trumps the occurrence of any turn yielding signals. However, certain interactive gestures offering or handing over the turn must be an exception to this, and their effect may be enhanced when speakers gaze or bodily orient towards their addressee. Prosodic information and speech content are important cues in the verbal modality [150,161] and interact with visual signals, presumably leading to complex multimodal gestalts that govern turn-taking in face-to-face interaction.

3. The quest for similarities between human and non-human primate communication

There is clear evidence for some degree of iconic and deictic gestural behaviour in non-human primates [162–167], and such discoveries are highly interesting. However, looking for points of comparison based on human pragmatic and interactive signalling allows us to potentially see further in terms of the evolutionary precursors to human language. As in humans, visual signalling in non-human primates is not confined to the hands. A wide range of articulators form an integral part of the repertoire, including eyes/gaze, mouth, head, arms, torso, chest, back, feet, legs or the entire body at once (e.g. [168–172]). Recently, the face has also been increasingly studied in great apes. While facial signals have been assumed to be rather fixed and reflexive, accumulating evidence points towards considerable complexity, flexibility and intentionality [173–175]. The following paragraphs hypothetically suggest some parallels that arise when considering communication by means of non-iconic visual signals and kinematic variation. They also point to recent studies that have adopted a more fine-grained analytic approach focused on interactive, reciprocal processes to further our understanding of non-human primate communication. An approach that moves beyond iconicity and involving detailed analyses of visual signalling in interactive *situ* may shed more light on some interesting parallels and potential precursors of core components foundational to pragmatic communication in humans (§§3a–3g).

(a) Visually signalling an intent to communicate

Similar to humans, body and face orientation as well as gaze direction can function as address and interaction-inviting signals in non-human primates [14,176–180]. Further similarities

are observable when looking at movement kinematics. For example, gaze of prolonged duration can signal communicative desires in gorillas [14], and a study on gibbons showed that the duration of facial expressions was longer in social contexts and when they could be visually perceived by a conspecific [175]. The amplitude of movements can also be linked to communicative motivations, such as observed in the 'directed scratch', a kinematically exaggerated scratching action understood as a request for grooming which differs from the purely instrumental action of scratching oneself [181]. Modulations of kinematic features such as signal duration or movement amplitude in connection with social and communicative motivations closely resemble the kinematic modulation of human behaviour, as seen above [40,41,110,126,127].

(b) Communicating pragmatic meaning and illocutionary force with visual bodily signals

One striking case of similarity in visual bodily signalling is the use of palm-up gestures in humans [68,69], orangutans [182] and chimpanzees [183] when making requests (e.g. food), one of the most fundamental of communicative actions. Deictic gestures, too, can be used by bonobos and chimpanzees to direct attention and make requests [162,163,184], as in humans [19,66,67].

The pragmatic functions of head gestures also bear some close resemblance to those of humans, such as bonobos preventing others' actions with head shakes, a presumed potential precursor to human visual signals of negation [185,186].

There is also evidence for the combination of movements into facial displays in non-human primates. The development of facial behaviour coding tools (e.g. ChimpFACS [187], OrangFACS [188] and GibbonFACS [189]) have propelled this research significantly forwards by allowing for fine-grained kinematic analyses [190]. This has yielded insights into configurational patterns of muscle activity forming facial expressions comparable to humans, at least in their morphological basis [187], as well as differences across species [188]. Composite signals can also be found beyond facial expressions. Oña et al. [191] investigated the combination of two different arm gestures with facial expressions in chimpanzees. While both gestures elicited affiliative responses when used on their own, the combination with a bared teeth facial expression enhanced this affiliative effect when paired with a stretched arm gesture, but eliminated the affiliative responses when combined with a bent arm gesture. Moreover, when embracing a multimodal view of non-human primate communication (e.g. [183,192–196]), evidence of cross-modal signal combinations can be found. Genty [194] found that contest hoot vocalisations were more frequently combined with gestures classed as 'soft' (e.g. touch) than 'rough' (e.g. kick) when used during friendly play, but the reverse applied to contest hoots used in agonistic contexts. Further, Eckert et al. [197] refer to reports on the sign-trained gorilla Koko, who used facial signals (the play-face) to contextualize what appeared to be intentionally produced wrong manual signs as acts of teasing.

Findings from these studies not only offer important new insights into the compositionality of non-human primates' communicative acts, a core feature of human language. They also show that compositional differences can map onto differences in pragmatic function, which is another striking resemblance (e.g. [90,93–95]).

(c) Producing recipient-designed utterances with visual bodily signals

Great ape gestural communication boasts some of the basic ingredients for recipient design, such as awareness of the recipient's visual attentional state, as evidenced by apes moving around an experimenter so they can see their gestures [198] and by young chimpanzees' cross-modal gestural adjustments (silent-visual versus audible-or-contact gestures) for their visually attentive/inattentive mothers [199]. Expanding such analyses to further domains of pragmatic behaviour and cognition, therefore, seems worthwhile. For example, looking for kinematic modulations of gestural behaviour when communicating about information that is or is not in common ground (e.g. [200,201]) paralleling those found in humans (size, space, precision, as seen above) may shed new light on the extent to which great apes consider their recipients' knowledge or state of mind. Interesting, too, would be analyses focused on incremental common ground (which builds up as a consequence of shared interaction history [19]). Recent findings show that bonobos are able to take into account this form of common ground, using more established (and successful) signals with interlocutors who share interaction history with them than with interlocutors who do not, while the reverse pattern was found for the use of novel signals [202]. An important question is whether such repeated signals bear any similarities to when and how humans repeat their gestures, and whether repetitions of the same gesture forms across interlocutors [118–122] occur in apes at all. If they do, one may wonder if such repetitions act in any way as a form of grounding (indicating communicative 'success' and mutual understanding) as they seem to do in human communication [118].

(d) Visual bodily signals for grounding and repair

Further similarities between human and non-human primate communication can be gleaned from reciprocal behaviour in the light of communicative success or lack thereof. Response waiting followed by repetition or adjustment of gestural behaviour when communication was unsuccessful—behaviours observable in great apes [9,14,195,202–205]—resembles humans' close observation of their interlocutors for evidence of understanding and communicative success [116] and gestural adjustments as strategies for repair when needed (figure 2; [61]). A particularly interesting example is the observation of orangutans trying to address problems in understanding by adapting their signalling to the degree of (mis)understanding evident from their interlocutor's behaviour. Partial misunderstanding led to adaptations that involved narrowing down the range of signals used, focusing on a subset of signals previously employed, but complete misunderstanding was met with broadening the range of signals used and avoiding unsuccessful ones [205]. It is possible that the ability for such adaptations may be the evolutionary foundation on which the capacity for detailed semantic and pragmatic gestural adaptations observable in humans (e.g. figure 2) may have built. Future studies investigating the link between specific informational requirements for solving misunderstandings and non-human primates' detailed gestural adaptations (or lack thereof) to address these may shed more light on this issue.

While specific vocal repair initiations in great apes may not exist, changes in the dynamics of movement that hint at

problems in uptake—including holds or tacit forward movements of head or torso, as seen in human communication [141,147]—are much more of a possibility. One interesting question is whether the observation that repair by non-human primates tends to be self- rather than other-initiated (Heesen *et al.* [206]) is confirmed if we apply a micro-analytic perspective including fine-grained movement kinematics. An interesting question, too, is whether it may be possible to find kinematic correlates of goal achievement in ape behaviour following responses to instances of persistence or elaboration, similar to the release of bodily articulators that were held in a static or forward position [147]. Equally interesting would be a focus on analogies (or differences) to the three-step grounding process (A–B–A, i.e. interactants acknowledging each other's communicative moves and ratifying them as satisfactory) evident in human interaction, as seen above [123].

(e) Coordinating sequential interaction with visual bodily signals

One domain in which a sensitivity for sequential, reciprocal action has been observed throughout much of the animal kingdom is that of turn-taking, vocally and gesturally [207,208]. Similarly to humans, turn sequences in apes are frequently accompanied by gaze [180,208], an association also seen in collaboratively breeding marmoset monkeys (Burkart *et al.* [209]). One interesting focus for future cross-species comparisons is the micro-organization of articulators and signals in the sequential coordination of interaction, such as the precise temporal interplay of gesture and gaze (and perhaps vocalizations).

We already know from recent studies that gestural behaviour in non-human primate interactions is often characterized by a high degree of mutual calibration on a temporally fine-grained scale, creating the opportunity for moment-by-moment reciprocal feedback and adaptability, similar to the behavioural processes observable in human interaction. Recent work has suggested that, at least for chimpanzees, such processes may be much more characteristic than communication based on a fixed species-specific repertoire of gestures, especially in mother–infant interactions [164,210,211]. The Social Negotiation Hypothesis [210,211] is based on data suggestive of idiosyncratic and flexible, dyad-specific and context-specific gestural communication and proposes that 'interactants mutually shape—or [*sic.*] 'socially negotiate'—the outcome of each gestural interaction in real time' [211, pp. 557–558]. For example, analyses of great ape interactions show that gestures may be adapted based on factors such as the attitudinal state of the infant (coordinative/discoordinative) resulting in different intensities of a push gesture [164], or the mother's consideration of whether the context requires the need for carrying the infant [210].

Detailed analyses of behaviour among non-human primates, also applying methods informed by conversation analysis [58,206,212,213] reveal their interactions to be orderly, reciprocal, micro-coordinated sequential activities. This involves establishing participation frameworks, awareness of conspecifics' attentional states, the monitoring of responses and moment-by-moment adjustment of signals to coordinate entry and exit phases of interactions, suspend and resume joint actions, request and respond to requests, and so forth, in contexts involving play, grooming, travelling or sharing food (e.g. [176,180,214,215]). Often, these finely coordinated interaction sequences involve a range of different visual articulators, including manual gestural actions, body orientation and gaze, and their detailed interplay has been illustrated nicely in these recent studies. Further advances may here be made by focusing on the subtle (and not so subtle) variation in the kinematic properties of these individual visual signals, as well as on the signals' intricate temporal co-organization and potential differences in interactive meaning this may entail—an area that still requires considerable work in humans, too.

The above examples show that a focus beyond iconicity, including compositionality of multimodal utterances, movement kinematics and analyses of the role of visual signals for coordinating interactional processes, lays bare new potential precursors to human multimodal language and the dispositions that constitute the hypothesized interaction engine in modern humans [38,39]. To what extent exactly those aspects of visual behaviour appearing similar in humans reflect similarities in cognitive abilities, or their evolutionary roots, goes beyond the scope of the present review, and is to a large extent still a matter of future research. Detailed interaction analyses supplemented with experimental manipulation to causally relate behaviour to cognitive and social abilities may be the most promising way forward in trying to consolidate the meaning of observed parallels between human and non-human primates.

4. Human multimodal language and its potential precursors

While a range of theories about the origins of human language cast gesture as a central element, the weight and role attributed to the gestural modality varies considerably. Some theories ascribe a bridging function to gesture, serving as a transitional stage in the development of a fully-fledged vocal language system [3–8]. Others have cast the role of gesture as having emerged from an inherent biomechanical connection between vocalizations and bodily movements, suggesting that 'hand gestural movements may thus have evolved as an embodied innovation for vocal control' [210, p. 11366]. Both approaches can be considered speech-centric in that gesture fulfils functions that assist vocal language in achieving its full potential, be it via (proto-)symbolic representation or acoustic modulation.

The view put forward here contrasts with those approaches in that it considers visual signalling as being a fundamental component of modern human language. Thus, the present account is in line with existing ones arguing that human language evolved as a multimodal phenomenon (e.g. [9–18]) and that spoken language alone does not represent the evolutionary pinnacle and therefore, by itself, is not the explanandum. A multimodal communication system which evolved with the visual and the vocal channels inextricably intertwined (and on a par) is extremely effective as it offers deeply integrated but at the same time complementary modes of communication with different affordances for encoding information (linear-sequential versus holistic [34]).

Critically, such a multimodal, multilayered system allows for parallel information streams with rich, composite signals and multimodal utterances, which substantially expands expressive potential. This boosts the system's compositional power and affords complex but efficiently packaged—and by all likelihood more quickly recognizable—communicative
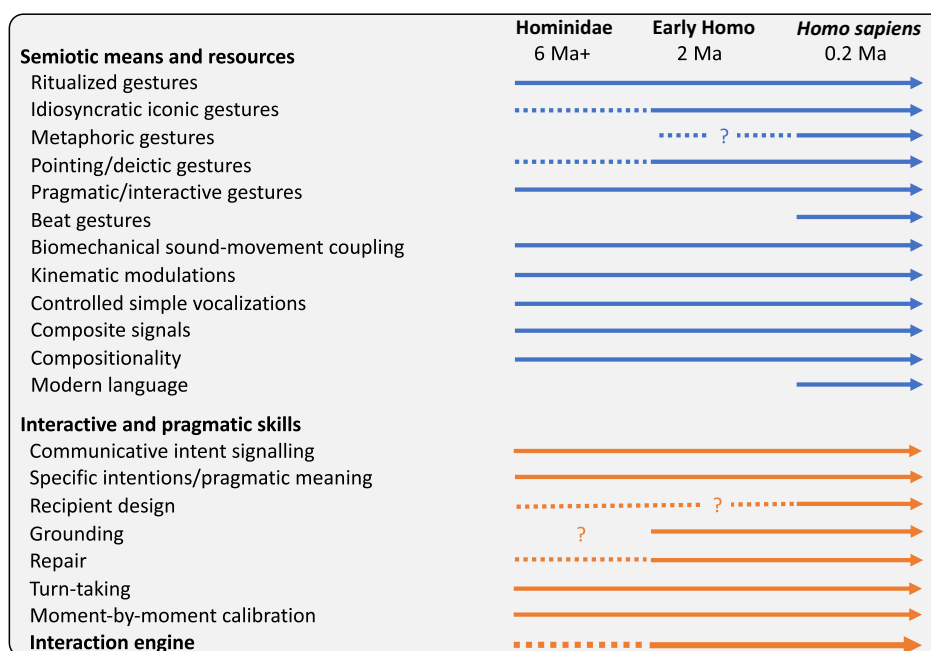
Figure 3. Hypothetical timelines for the evolution of hallmarks of human communicative behaviour. Solid arrows, present; dotted lines, precursor present; question marks, domain of uncertainty. Note that the gesture types refer to existing descriptions and labels (iconic, metaphoric, beats: McNeill [34]; pragmatic: Kendon [68]; interactive: Bavelas *et al.* [111]). Reference to the interaction engine is in bold since this hypothesized construct entails the interactive and pragmatic skills listed above it.

gestalts. This rich expressive potential crucially rests on the diverse set of articulators humans have at their disposal (the hands, head, face, gaze and torso) and the wide range of visual signals they are able to convey and meaningfully combine. All this contributes to the versatile set of human pragmatic abilities. It also significantly augments the capacity for fine-tuning and flexibly adapting communicative messages to interlocutors and their interactive behaviours, a much-needed prerequisite for creating mutual understanding in social interaction.

As Levinson [39] has pointed out, some non-human primate behaviours may be considered precursors to the human abilities that create a special capacity for social interaction, and thus the foundation for the emergence of modern human language. A multimodal view of human pragmatics and a consideration of visual bodily signals in the coordination of minds in interaction makes the gap between humans and non-human primate communication appear smaller than it may seem at first sight. As we saw above, such an approach points towards some of the key features of the modern human communication system already being present in our last common ancestor with chimpanzees around 6 Ma, or in even earlier last common ancestors on the hominid line (figure 3). Iconicity will of course have played a fundamental role in the further evolution of referential and propositional communication in both gesture and speech (e.g. [12,13,17,26,29,33]). The potential for early iconicity is already evident in non-human primates [29,165–167], plus perhaps even the seeds for some basic gesture–prosodic connection [216]. In line with the Interaction Engine Hypothesis, the main driver of creating the foundations for the emergence of complex language are likely to have been social-interactional abilities and motivations. Here I have highlighted that a focus on pragmatic, socially situated visual bodily signalling can reveal potentially even earlier evolutionary precursors to modern human language than commonly assumed.

The sections above have detailed a range of core contributions the body makes to communication. Delving deeper into investigating signals and their functions that scaffold the pragmatic processes elementary to coordinating minds in interaction is a fruitful way forward in our endeavour to understand both human and non-human primate communication in their own right. It also facilitates discovering potential precursors and the evolutionary path that led to the emergence of modern human multimodal language and its foundational interactional capacities.

5. Conclusion

Human communication is fundamentally multimodal. If we ask how verbal language evolved, then the answer has got to be not by supplanting gesture. Human visual signalling is anything but a temporary bridge to spoken language or a primitive mode of communication that was replaced by spoken words. Only a scenario in which the vocal and the visual modalities co-evolved into a tightly integrated system can explain how our species ended up with the highly adaptable and flexible multimodal toolbox that allows us to achieve the coordination of minds so effortlessly, effectively and efficiently. The multimodal nature of the human communication system combined with the socio-cognitive abilities that constitute the hypothesized interaction engine has equipped the human species with an unparalleled system for communicating not only semantic but also pragmatic meaning. A wide range of visual bodily signals going considerably beyond iconic manual gestures forms an integral component of the process of achieving mutual understanding in face-to-face interaction.

Endnote

[1]Although this review is based on data and examples from speakers of different languages and cultures, these differences are glossed over in the present manuscript, as are individual differences and atypical populations. While all immensely important aspects in their own right, their discussion would go beyond the scope of the present review. Future analyses are required to devote detailed attention to these issues and gain an understanding of the extent to which current views may need to be refined.

References

1. Berwick RC, Chomsky N. 2015 *Why only us: language and evolution*. Cambridge, MA: MIT Press.

2. Seyfarth RM, Cheney DL. 2014 The evolution of language from social cognition. *Curr. Opin. Neurobiol.* **28**, 5–9. (doi:10.1016/j.conb.2014.04.003)

3. Arbib MA. 2012 *How the brain got language: the mirror system hypothesis*. New York, NY: Oxford University Press.

4. Corballis MC. 2017 *The truth about language: what it is and where it came from*. Chicago, IL: University of Chicago Press.

5. Corballis MC. 2002 *From hand to mouth: the origins of language*. Princeton, NJ: Princeton University Press.

6. Hewes GW. 1973 Primate communication and the gestural origin of language. *Curr. Anthropol.* **14**, 5–24.

7. Tomasello M. 2008 *Origins of human communication*. Cambridge, MA: MIT Press.

8. Wundt WM. 1900 *Völkerpsychologie: Bd., 1.-2. T. Die Sprache. 1900*. Leipzig, Germany: W. Englemann.

9. Fröhlich M, Sievers C, Townsend SW, Gruber T, van Schaik CP. 2019 Multimodal communication and language origins: integrating gestures and vocalizations. *Biol. Rev.* **94**, 1809–1829. (doi:10.1111/brv.12535)

10. Kendon A. 2017 Reflections on the 'gesture-first' hypothesis of language origins. *Psychon. Bull. Rev.* **24**, 163–170. (doi:10.3758/s13423-016-1117-3)

11. Levinson SC, Holler J. 2014 The origin of human multi-modal communication. *Phil. Trans. R. Soc. B* **369**, 20130302. (doi:10.1098/rstb.2013.0302)

12. McNeill D. 2012 *How language began: gesture and speech in human evolution*. Cambridge, UK: Cambridge University Press.

13. Perlman M. 2017 Debunking two myths against vocal origins of language: language is iconic and multimodal to the core. *Interact. Stud.* **18**, 376–401. (doi:10.1075/is.18.3.05per)

14. Prieur J, Barbu S, Blois-Heulin C, Lemasson A. 2020 The origins of gestures and language: history, current advances and proposed theories. *Biol. Rev.* **95**, 531–554. (doi:10.1111/brv.12576)

15. Goldin-Meadow S. 2017 What the hands can tell us about language emergence. *Psychon. Bull. Rev.* **24**, 213–218. (doi:10.3758/s13423-016-1074-x)

16. Zlatev J, Wacewicz S, Zywiczynski P, Weijer J van de. 2017 Multimodal-first or pantomime-first? *Interact. Stud.* **18**, 465–488. (doi:10.1075/is.18.3.08zla)

17. Vigliocco G, Perniss P, Vinson D. 2014 Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Phil. Trans. R. Soc. B* **369**, 20130292. (doi:10.1098/rstb.2013.0292)

18. Lister CJ, Fay N. 2017 How to create a human communication system: a theoretical model. *Interact. Stud.* **18**, 314–329. (doi:10.1075/is.18.3.02lis)

19. Clark HH. 1996 *Using language*. Cambridge, UK: Cambridge University Press.

20. Bohn M, Kachel G, Tomasello M. 2019 Young children spontaneously recreate core properties of language in a new modality. *Proc. Natl Acad. Sci. USA* **116**, 26 072–26 077. (doi:10.1073/pnas.1904871116)

21. Cartmill EA, Rissman L, Novack MA, Goldin-Meadow S. 2017 The development of iconicity in children's co-speech gesture and homesign. *Lang. Interact. Acquis.* **8**, 42–68. (doi:10.1075/lia.8.1.03car)

22. Fay N, Lister CJ, Ellison TM, Goldin-Meadow S. 2014 Creating a communication system from scratch: gesture beats vocalization hands down. *Front. Psychol.* **5**, 354. (doi:10.3389/fpsyg.2014.00354)

23. Motamedi Y, Schouwstra M, Smith K, Culbertson J, Kirby S. 2019 Evolving artificial sign languages in the lab: from improvised gesture to systematic sign. *Cognition* **192**, 103964. (doi:10.1016/j.cognition.2019.05.001)

24. Murgiano M, Motamedi Y, Vigliocco G. 2021 Situating language in the real-world: the role of multimodal iconicity and indexicality. *J. Cogn.* **4**, 38. (doi:10.5334/joc.113)

25. Perniss P, Thompson R, 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)

26. 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)

27. Ortega G. 2017 Iconicity and sign lexical acquisition: a review. *Front. Psychol.* **8**, 1280. (doi:10.3389/fpsyg.2017.01280)

28. Ortega G, Özyürek A. 2020 Types of iconicity and combinatorial strategies distinguish semantic categories in silent gesture across cultures. *Lang. Cogn.* **12**, 84–113. (doi:10.1017/langcog.2019.28)

29. Perlman M, Cain AA. 2014 Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture* **14**, 320–350. (doi:10.1075/gest.14.3.03per)

30. Pyers J, Senghas A. 2020 Lexical iconicity is differentially favored under transmission in a new sign language: the effect of type of iconicity. *Sign Lang. Linguist.* **23**, 73–95. (doi:10.1075/sll.00044.pye)

31. Sandler W. 2009 Symbiotic symbolization by hand and mouth in sign language. *Semiotica* **2009**, 241–275. (doi:10.1515/semi.2009.035)

32. Thompson RL, Vinson DP, Woll B, Vigliocco G. 2012 The road to language learning is iconic: evidence from British Sign Language. *Psychol. Sci.* **23**, 1443–1448. (doi:10.1177/0956797612459763)

33. Zlatev J, Żywiczyński P, Wacewicz S. 2020 Pantomime as the original human-specific communicative system. *J. Lang. Evol.* **5**, 156–174. (doi:10.1093/jole/lzaa006)

34. McNeill D. 1992 *Hand and mind: what gestures reveal about thought*. Chicago, IL: University of Chicago Press.

35. Holler J, Beattie G. 2002 A micro-analytic investigation of how iconic gestures and speech represent core semantic features in talk. *Semiotica* **2002,** 31–69. (doi:10.1515/semi.2002.077)

36. 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)

37. Goldin-Meadow S. 2003 *Hearing gesture: how our hands help US think*. Cambridge, MA: Belknap Press of Harvard University Press.

38. 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.

39. Levinson SC. 2019 Interactional foundations of language: the interaction engine hypothesis. In *Human language: from genes and brain to behavior* (ed. P Hagoort), pp. 189–200. Cambridge, MA: MIT Press.

40. Trujillo JP, Simanova I, Bekkering H, Ozyurek A. 2018 Communicative intent modulates production and comprehension of actions and gestures: a Kinect study. *Cognition* **180**, 38–51. (doi:10.1016/j.cognition.2018.04.003)

41. Peeters D, Chu M, Holler J, Hagoort P, Özyürek A. 2015 Electrophysiological and kinematic correlates of communicative intent in the planning and production of pointing gestures and speech. *J. Cogn. Neurosci.* **27**, 2352–2368. (doi:10.1162/jocn_a_00865)

42. Sperber D, Wilson D. 1986 *Relevance: communication and cognition*. Cambridge MA: Harvard University Press.

12

43. Goodwin C. 1981 Conversational organization: interaction between speakers and hearers. New York, NY: Academic Press.

44. Kendon A. 1990 Conducting interaction: patterns of behavior in focused encounters. New York, NY: Cambridge University Press.

45. Mortensen K, Hazel S. 2014 Moving into interaction—social practices for initiating encounters at a help desk. J. Pragmat. 62, 46–67. (doi:10.1016/j.pragma.2013.11.009)

46. Nagels A, Kircher T, Steines M, Straube B. 2015 Feeling addressed! The role of body orientation and co-speech gesture in social communication. Hum. Brain Mapp. 36, 1925–1936. (doi:10.1002/hbm.22746)

47. Robinson JD. 1998 Getting down to business: talk, gaze, and body orientation during openings of doctor-patient consultations. Hum. Commun. Res. 25, 97–123. (doi:10.1111/j.1468-2958.1998.tb00438.x)

48. Cary MS. 1978 The role of gaze in the initiation of conversation. Soc. Psychol. 41, 269–271. (doi:10.2307/3033565)

49. Goffman E. 1966 Behavior in public places: notes on the social organization of gatherings. Reissue edition. New York, NY: Free Press.

50. Goodwin C. 1980 Restarts, pauses, and the achievement of a state of mutual gaze at turn beginning. Sociol. Inq. 50, 272–302.

51. Lerner GH. 2003 Selecting next speaker: the context-sensitive operation of a context-free organization. Lang. Soc. 32, 177–201. (doi:10.1017/S004740450332202X)

52. Holler J, Schubotz L, Kelly S, Hagoort P, Schuetze M, Özyürek A. 2014 Social eye gaze modulates processing of speech and co-speech gesture. Cognition 133, 692–697. (doi:10.1016/j.cognition.2014.08.008)

53. Holler J, Kokal I, Toni I, Hagoort P, Kelly SD, Özyürek A. 2015 Eye'm talking to you: speakers' gaze direction modulates co-speech gesture processing in the right MTG. Soc. Cogn. Affect. Neurosci. 10, 255–261. (doi:10.1093/scan/nsu047)

54. Goodwin C. 2000 Action and embodiment within situated human interaction. J. Pragmat. 32, 1489–1522. (doi:10.1016/S0378-2166(99)00096-X)

55. Mondada L. 2014 Bodies in action: multimodal analysis of walking and talking. Lang. Dialogue 4, 357–403. (doi:10.1075/ld.4.3.02mon)

56. Mondada L. 2016 Challenges of multimodality: language and the body in social interaction. J. Socioling. 20, 336–366. (doi:10.1111/josl.1_12177)

57. Schegloff EA. 1998 Body torque. Soc. Res. 65, 535–596.

58. Mondada L, Meguerditchian A. 2022 Sequence organization and embodied mutual orientations: openings of social interactions between baboons. Phil. Trans. R. Soc. B 377, 20210101. (doi:10.1098/rstb.2021.0101)

59. Goodwin C. 1986 Gestures as a resource for the organization of mutual orientation. Semiotica 62, 29–50. (doi:10.1515/semi.1986.62.1-2.29)

60. Gullberg M, Holmqvist K. 1999 Keeping an eye on gestures: visual perception of gestures in face-to-face communication. Pragmat. Cogn. 7, 35–63. (doi:10.1075/pc.7.1.04gul)

61. Holler J, Wilkin K. 2011 An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. J. Pragmat. 43, 3522–3536. (doi:10.1016/j.pragma.2011.08.002)

62. Streeck J. 1993 Gesture as communication I: its coordination with gaze and speech. Commun. Monogr. 60, 275–299. (doi:10.1080/03637759309376314)

63. Streeck J. 1994 Gesture as communication II: the audience as co-author. Res. Lang. Soc. Interact. 27, 239–267. (doi:10.1207/s15327973rlsi2703_5)

64. Grice HP. 1957 Meaning. Philos. Rev. 66, 377–388. (doi:10.2307/2182440)

65. Austin JL. 1962 How to do things with words. The William James Lectures delivered at Harvard University in 1955. Oxford, UK: Oxford University Press.

66. Enrici I, Adenzato M, Cappa S, Bara BG, Tettamanti M. 2011 Intention processing in communication: a common brain network for language and gestures. J. Cogn. Neurosci. 23, 2415–2431. (doi:10.1162/jocn.2010.21594)

67. Kelly SD, Barr DJ, Breckenridge Church R, Lynch K. 1999 Offering a hand to pragmatic understanding: the role of speech and gesture in comprehension and memory. J. Mem. Lang. 40, 577–592. (doi:10.1006/jmla.1999.2634)

68. Kendon A. 2004 Gesture: visible action as utterance. Cambridge, UK: Cambridge University Press.

69. Müller C. 2004 Forms and uses of the Palm Up Open Hand: a case of a gesture family? In The semantics and pragmatics of everyday gestures (eds R Posner, C Müller), pp. 234–256. Berlin, Germany: Weidler.

70. Cooperrider K, Abner N, Goldin-Meadow S. 2018 The palm-up puzzle: meanings and origins of a widespread form in gesture and sign. Front. Commun. 3, 23. (doi:10.3389/fcomm.2018.00023)

71. Streeck J. 2009 Gesturecraft: the manufacture of meaning. Amsterdam, The Netherlands: John Benjamins Publishing Company.

72. Streeck J, Hartge U. 1992 Previews: gestures at the transition place. In The contextualization of language (eds P Auer, A Di Luzio), pp. 135-157. Amsterdam, The Netherlands: John Benjamins.

73. Bressem J, Müller C. 2014. The family of away gestures: negation, refusal, and negative assessment. In Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of linguistics and communication science (HSK) 38/2, vol. 2, pp. 1592–1604. Berlin, Germany: De Gruyter Mouton.

74. Harrison S. 2010 Evidence for node and scope of negation in coverbal gesture. Gesture 10, 29–51. (doi:10.1075/gest.10.1.03har)

75. Bressem J, Müller C. 2014 A repertoire of German recurrent gestures with pragmatic functions. In Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of linguistics and communication science (HSK) 38/2, pp. 1575–1591. Berlin, Germany: De Gruyter Mouton.

76. Manera V, Schouten B, Becchio C, Bara BG, Verfaillie K. 2010 Inferring intentions from biological motion: a stimulus set of point-light communicative interactions. Behav. Res. Methods 42, 168–178. (doi:10.3758/BRM.42.1.168)

77. Vesper C, Sevdalis V. 2020 Informing, coordinating, and performing: a perspective on functions of sensorimotor communication. Front. Hum. Neurosci. 14, 168. (doi:10.3389/fnhum.2020.00168)

78. Trujillo JP, Holler J. 2021 The kinematics of social action: visual signals provide cues for what interlocutors do in conversation. Brain Sci. 11, 996. (doi:10.3390/brainsci11080996)

79. Malisz Z, Włodarczak M, Buschmeier H, Skubisz J, Kopp S, Wagner P. 2016 The ALICO corpus: analysing the active listener. Lang. Resour. Eval. 50, 411–442. (doi:10.1007/s10579-016-9355-6)

80. Bousmalis K, Mehu R, Pantic M. 2012 Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: a survey of related cues databases and tools. Image and Vision Computing 31, 203–221. (doi:10.1016/j.imavis.2012.07.003)

81. Birdwhistle RL. 1970 Kinesics and Context - Essays on Body Motion Communication. Philadelphia: University of Pennsylvania Press.

82. Heylen D. 2005 Challenges ahead: Head movements and other social acts in conversations. In Proceedings of AISB 2005, pp. 45–52.

83. Heylen D. 2008 Listening heads. In Modeling Communication with robots and virtual humans, pp. 241–259. Berlin, Heidelberg: Springer.

84. Wagner P, Malisz Z, Kopp S. 2014 Gesture and speech in interaction: an overview. Speech Commun. 57, 209–232. (doi:10.1016/j.specom.2013.09.008)

85. Kendrick KH, Holler J. 2017 Gaze direction signals response preference in conversation. Res. Lang. Soc. Interact. 50, 12–32. (doi:10.1080/08351813.2017.1262120)

86. Aguert M. 2021 Paraverbal expression of verbal irony: vocal cues matter and facial cues even more. J. Nonverbal Behav. 46, 45–70. (doi:10.1007/s10919-021-00385-z)

87. Ekman P. 1979 About brows: emotional and conversational signals. In Human ethology (eds M Cranach, K Foppa, W Lepenies, D Ploog), pp. 169–249. Cambridge, UK: Cambridge University Press.

88. Enfield NJ et al. 2013 Huh? What? – a first survey in twenty-one languages. In Conversational repair and

human understanding (eds M Hayashi, G Raymond, J Sidnell), pp. 343–380. Cambridge, UK: Cambridge University Press.

89. de Vos C, van der Kooij E, Crasborn O. 2009 Mixed signals: combining linguistic and affective functions of eyebrows in questions in sign language of the Netherlands. *Lang. Speech* **52**, 315–339. (doi:10.1177/0023830909103177)

90. Nota N, Trujillo JP, Holler J. 2021 Facial signals and social actions in multimodal face-to-face interaction. *Brain Sci.* **11**, 1017. (doi:10.3390/brainsci11081017)

91. Kaukomaa T, Peräkylä A, Ruusuvuori J. 2014 Foreshadowing a problem: turn-opening frowns in conversation. *J. Pragmat.* **71**, 132–147. (doi:10. 1016/j.pragma.2014.08.002)

92. Hoemke P. 2019 The face in face-to-face interaction. Doctoral thesis, Radboud University Nijmegen, Nijmegen, The Netherlands.

93. Borràs-Comes J, Prieto P. 2011 'Seeing tunes.' The role of visual gestures in tune interpretation. *Lab. Phonol.* **2**, 355–380. (doi:10.1515/labphon.2011. 013)

94. Crespo Sendra V, Kaland C, Swerts M, Prieto P. 2013 Perceiving incredulity: the role of intonation and facial gestures. *J. Pragmat.* **47**, 1–13. (doi:10.1016/ j.pragma.2012.08.008)

95. Tabacaru S, Lemmens M. 2014 Raised eyebrows as gestural triggers in humour: the case of sarcasm and hyper-understanding. *Eur. J. Humour Res.* **2**, 11–31. (doi:10.7592/EJHR2014.2. 2.tabacaru)

96. Benitez-Quiroz CF, Wilbur RB, Martinez AM. 2016 The not face: a grammaticalization of facial expressions of emotion. *Cognition* **150**, 77–84. (doi:10.1016/j.cognition.2016.02.004)

97. Bavelas J, Chovil N. 2018 Some pragmatic functions of conversational facial gestures. *Gesture* **17**, 98–127. (doi:10.1075/gest.00012.bav)

98. Chovil N. 1991 Discourse-oriented facial displays in conversation. *Res. Lang. Soc. Interact.* **25**, 163–194. (doi:10.1080/08351819109389361)

99. Goodwin M, Goodwin C. 1986 Gesture and coparticipation in the activity of searching for a word. *Semiotica* **62**, 51–76. (doi:10.1515/semi. 1986.62.1-2.51)

100. Clark HH, Carlson TB. 1982 Hearers and speech acts. *Language* **58**, 332–373. (doi:10.2307/414102)

101. Sacks H, Schegloff EA, Jefferson G. 1974 A simplest systematics for the organization of turn-taking for conversation. *Language* **50**, 696–735. (doi:10.2307/ 412243)

102. Holler J, Bavelas J. 2017 Multi-modal communication of common ground: a review of social functions. In *Why gesture? How the hands function in speaking, thinking and communicating* (eds RB Church, MW Alibali, SD Kelly), pp. 213–240. Amsterdam, The Netherlands: John Benjamins Publishing.

103. Gerwing J, Bavelas J. 2004 Linguistic influences on gesture's form. *Gesture* **4**, 157–195. (doi:10.1075/ gest.4.2.04ger)

104. Holler J, Stevens R. 2007 The effect of common ground on how speakers use gesture and speech to represent size information. *J. Lang. Soc. Psychol.* **26**, 4–27. (doi:10.1177/0261927X06296428)

105. Schubotz L, Özyürek A, Holler J. 2019 Age-related differences in multimodal recipient design: younger, but not older adults, adapt speech and co-speech gestures to common ground. *Lang. Cogn. Neurosci.* **34**, 254–271. (doi:10.1080/23273798.2018. 1527377)

106. Holler J, Wilkin K. 2009 Communicating common ground: how mutually shared knowledge influences speech and gesture in a narrative task. *Lang. Cogn. Process.* **24**, 267–289. (doi:10.1080/ 01690960802095545)

107. Enfield NJ, Kita S, De Ruiter JP. 2007 Primary and secondary pragmatic functions of pointing gestures. *J. Pragmat.* **39**, 1722–1741. (doi:10.1016/j.pragma. 2007.03.001)

108. Galati A, Brennan SE. 2014 Speakers adapt gestures to addressees' knowledge: implications for models of co-speech gesture. *Lang. Cogn. Neurosci.* **29**, 435–451. (doi:10.1080/01690965.2013.796397)

109. Hilliard C, Cook SW. 2016 Bridging gaps in common ground: speakers design their gestures for their listeners. *J. Exp. Psychol. Learn. Mem. Cogn.* **42**, 91–103. (doi:10.1037/xlm0000154)

110. Holler J, Bavelas JB, Wood J, Geiger M, Simons L.

111. Bavelas JB, Chovil N, Coates L, Roe L. 1995 Gestures specialized for dialogue. *Pers. Soc. Psychol. Bull.* **21**, 394–405. (doi:10.1177/ 0146167295214010)

112. Holler J. 2010 Speakers' use of interactive gestures as markers of common ground. In *Gesture in embodied communication and human-computer interaction* (eds S Kopp, I Wachsmuth), pp. 11–22. Berlin, Germany: Springer.

113. Sandler W. 2018 The body as evidence for the nature of language. *Front. Psychol.* **9**, 1782. (doi:10. 3389/fpsyg.2018.01782)

114. Clark HH, Brennan SE. 1991 Grounding in communication. In *Perspectives on socially shared cognition* (eds LB Resnick, JM Levine, SD Teasley), pp. 127–149. Washington, DC: APA Books.

115. Schegloff EA. 1982 Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences. In *Analyzing discourse: text and talk* (ed. D Tannen), pp. 71–93. Washington, DC: Georgetown University Press.

116. Clark HH, Krych MA. 2004 Speaking while monitoring addressees for understanding. *J. Mem. Lang.* **50**, 62–81. (doi:10.1016/j.jml.2003. 08.004)

117. Clark HH, Wilkes-Gibbs D. 1986 Referring as a collaborative process. *Cognition* **22**, 1–39. (doi:10. 1016/0010-0277(86)90010-7)

118. Holler J, Wilkin K. 2011 Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *J. Nonverbal Behav.* **35**, 133–153. (doi:10.1007/s10919-011-0105-6)

119. Kimbara I. 2008 Gesture form convergence in joint description. *J. Nonverbal Behav.* **32**, 123–131. (doi:10.1007/s10919-007-0044-4)

120. Kimbara I. 2006 On gestural mimicry. *Gesture* **6**, 39–61. (doi:10.1075/gest.6.1.03kim)

121. Mol L, Krahmer E, Maes A, Swerts M. 2012 Adaptation in gesture: converging hands or converging minds? *J. Mem. Lang.* **66**, 249–264. (doi:10.1016/j.jml.2011.07.004)

122. Rasenberg M, Ozyurek A, Dingemanse M. 2020 Alignment in multimodal interaction: an integrative framework. *Cogn. Sci.* **44**, e12911. (doi:10.1111/ cogs.12911)

123. Bavelas J, Gerwing J, Healing S. 2017 Doing mutual understanding. Calibrating with micro-sequences in face-to-face dialogue. *J. Pragmat.* **121**, 91–112. (doi:10.1016/j.pragma.2017.09.006)

124. Heldner M, Hjalmarsson A, Edlund J. 2013 *Backchannel relevance spaces*, pp. 137–146. Frankfurt am Main, Germany: Peter Lang Publishing Group.

125. Brunner LJ. 1979 Smiles can be back channels. *J. Pers. Soc. Psychol.* **37**, 728–734. (doi:10.1037/ 0022-3514.37.5.728)

126. Hömke P, Holler J, Levinson SC. 2018 Eye blinks are perceived as communicative signals in human face-to-face interaction. *PLoS ONE* **13**, e0208030. (doi:10. 1371/journal.pone.0208030)

127. Hömke P, Holler J, Levinson SC. 2017 Eye blinking as addressee feedback in face-to-face conversation. *Res. Lang. Soc. Interact.* **50**, 54–70. (doi:10.1080/ 08351813.2017.1262143)

128. Levinson SC. 2005 Living with Manny's dangerous idea. *Discourse Stud.* **7**, 431–453. (doi:10.1177/ 1461445605054401)

129. Levinson SC. 2015 Other-initiated repair in Yélî Dnye: seeing eye-to-eye in the language of Rossel Island. *Open Linguist.* **1**, 386–410. (doi:10.1515/ opli-2015-0009)

130. Bavelas JB, Black A, Lemery CR, Mullett J. 1986 'I show how you feel': motor mimicry as a communicative act. *J. Pers. Soc. Psychol.* **50**, 322–329. (doi:10.1037/0022-3514.50.2.322)

131. Bavelas JB, Coates L, Johnson T. 2000 Listeners as co-narrators. *J. Pers. Soc. Psychol.* **79**, 941–952. (doi:10.1037/0022-3514.79.6.941)

132. Dingemanse M *et al*. 2015 Universal principles in the repair of communication problems. *PLoS ONE* **10**, e0136100. (doi:10.1371/journal.pone. 0136100)

133. Schegloff EA, Jefferson G, Sacks H. 1977 The preference for self-correction in the organization of repair in conversation. *Language* **53**, 361–382. (doi:10.1353/lan.1977.0041)

134. Albert S, de Ruiter JP. 2018 Repair: the interface between interaction and cognition. *Top. Cogn. Sci.* **10**, 279–313. (doi:10.1111/tops.12339)

135. Howes C, Lavelle M, Healey P, Hough J, McCab R. 2016 Helping hands? Gesture and self-repair in schizophrenia. In *Proc. of LREC 2016 Workshop. Resources and processing of linguistic and extra-linguistic data from people with various forms of cognitive/psychiatric impairments (RaPID-2016)*. Bielefeld, Germany: Bielefeld University. See https:// pub.uni-bielefeld.de/record/2906042.

136. Özkan EE, Gurion T, Hough J, Healey PGT, Jamone L. 2021 Specific hand motion patterns correlate to miscommunications during dyadic conversations. In

2021 IEEE Int. Conf. on Development and Learning (ICDL), pp. 1–6. New York, NY: IEEE.

137. Holler J, Turner K, Varcianna T. 2013 It's on the tip of my fingers: co-speech gestures during lexical retrieval in different social contexts. Lang. Cogn. Process. 28, 1509–1518. (doi:10.1080/01690965.2012.698289)

138. Healey PG, Plant NJ, Howes C, Lavelle M. 2015 When words fail: collaborative gestures during clarification dialogues. In 2015 AAAI Spring Symp. Series Technical Report, vol. SS-15-07, pp. 23–29. Palo Alto, CA: Association for the Advancement of Artificial Intelligence.

139. Dingemanse M, Kendrick KH, Enfield NJ. 2016 A coding scheme for other-initiated repair across languages. Open Linguistics 2, 35–46. (doi:10.1515/opli-2016-0002)

140. Mortensen K. 2016 The body as a resource for other-Initiation of repair: cupping the hand behind the ear. Res. Lang. Soc. Interact. 49, 34–57. (doi:10.1080/08351813.2016.1126450)

141. Li X. 2014 Leaning and recipient intervening questions in Mandarin conversation. J. Pragmat. 67, 34–60. (doi:10.1016/j.pragma.2014.03.011)

142. Seo M-S, Koshik I. 2010 A conversation analytic study of gestures that engender repair in ESL conversational tutoring. J. Pragmat. 42, 2219–2239. (doi:10.1016/j.pragma.2010.01.021)

143. Rasmussen G. 2014 Inclined to better understanding—the coordination of talk and 'leaning forward' in doing repair. J. Pragmat. 65, 30–45. (doi:10.1016/j.pragma.2013.10.001)

144. Healey P, Lavelle M, Howes C, Battersby S, McCabe R. 2013 How listeners respond to speaker's troubles. In Proc. Meeting of the Cognitive Science Society, vol. 35. Seattle, WA: Cognitive Science Society.

145. Kendrick KH. 2015 Other-initiated repair in English. Open Linguist. 1, 164–190. (doi:10.2478/opli-2014-0009)

146. Kendrick KH. 2015 The intersection of turn-taking and repair: the timing of other-initiations of repair in conversation. Front. Psychol. 6, 250. (doi:10.3389/fpsyg.2015.00250)

147. Floyd S, Manrique E, Rossi G, Torreira F. 2016 Timing of visual bodily behavior in repair sequences: evidence from three languages. Discourse Process. 53, 175–204. (doi:10.1080/0163853X.2014.992680)

148. Manrique E, Enfield N. 2015 Suspending the next turn as a form of repair initiation: evidence from Argentine Sign Language. Front. Psychol. 6, 1326. (doi:10.3389/fpsyg.2015.01326)

149. Mondada L. 2011 Understanding as an embodied, situated and sequential achievement in interaction. J. Pragmat. 43, 542–552. (doi:10.1016/j.pragma.2010.08.019)

150. Levinson SC, Torreira F. 2015 Timing in turn-taking and its implications for processing models of language. Front. Psychol. 6, 731. (doi:10.3389/fpsyg.2015.00731)

151. Holler J, Levinson SC. 2019 Multimodal language processing in human communication. Trends Cogn. Sci. 23, 639–652. (doi:10.1016/j.tics.2019.05.006)

152. Duncan S. 1972 Some signals and rules for taking speaking turns in conversations. J. Pers. Soc. Psychol. 23, 283–292. (doi:10.1037/h0033031)

153. Zellers M, House D, Alexanderson S. 2016 Prosody and hand gesture at turn boundaries in Swedish. In Proc. Speech Prosody 2016, pp. 831–835. Baixas, France: International Speech Communication Association. (doi:10.21437/SpeechProsody.2016-170)

154. Holler J, Kendrick KH, Levinson SC. 2018 Processing language in face-to-face conversation: questions with gestures get faster responses. Psychon. Bull. Rev. 25, 1900–1908. (doi:10.3758/s13423-017-1363-z)

155. Bavelas JB, Chovil N, Lawrie DA, Wade A. 1992 Interactive gestures. Discourse Process. 15, 469–489. (doi:10.1080/01638539209544823)

156. Mondada L. 2007 Multimodal resources for turn-taking: pointing and the emergence of possible next speakers. Discourse Stud. 9, 194–225. (doi:10.1177/1461445607075346)

157. Kendon A. 1967 Some functions of gaze-direction in social interaction. Acta Psychol. (Amst.) 26, 22–63. (doi:10.1016/0001-6918(67)90005-4)

158. Auer P. 2021 Turn-allocation and gaze: a multimodal revision of the 'current-speaker-selects-next' rule of the turn-taking system of conversation analysis. Discourse Stud. 23, 117–140. (doi:10.1177/1461445620966922)

159. Rossano F. 2012 Gaze behavior in face-to-face interaction. Nijmegen, The Netherlands: Radboud University Nijmegen. See http://hdl.handle.net/2066/99151.

160. Streeck J. 2014 Mutual gaze and recognition: revisiting Kendon's 'Gaze direction in two-person conversation'. In From gesture in conversation to visible action as utterance: essays in honor of Adam Kendon (eds M Seyfeddinipur, M Gullberg), pp. 35–56. Amsterdam, The Netherlands: John Benjamins.

161. Holler J, Kendrick KH. 2015 Unaddressed participants' gaze in multi-person interaction: optimizing recipiency. Front. Psychol. 6, 98. (doi:10.3389/fpsyg.2015.00098)

162. Leavens DA. 2004 Manual deixis in apes and humans. Interact. Stud. 5, 387–408. (doi:10.1075/is.5.3.05lea)

163. Hobaiter C, Leavens DA, Byrne RW. 2014 Deictic gesturing in wild chimpanzees (Pan troglodytes)? Some possible cases. J. Comp. Psychol. 128, 82–87. (doi:10.1037/a0033757)

164. Perlman M, Tanner JE, King BJ. 2012 A mother gorilla's variable use of touch to guide her infant. In Developments in primate gesture research (eds S Pika, K Liebal), pp. 55–72. Amsterdam, The Netherlands: John Benjamins.

165. Tanner JE, Byrne RW. 1996 Representation of action through iconic gesture in a captive lowland gorilla. Curr. Anthropol. 37, 162–173. (doi:10.1086/204484)

166. Genty E, Zuberbühler K. 2015 Iconic gesturing in bonobos. Commun. Integr. Biol. 8, e992742. (doi:10.4161/19420889.2014.992742)

167. Douglas PH, Moscovice LR. 2015 Pointing and pantomime in wild apes? Female bonobos use referential and iconic gestures to request genito-genital rubbing. Sci. Rep. 5, 13999. (doi:10.1038/srep13999)

168. Byrne RW, Cartmill E, Genty E, Graham KE, Hobaiter C, Tanner J. 2017 Great ape gestures: intentional communication with a rich set of innate signals. Anim. Cogn. 20, 755–769. (doi:10.1007/s10071-017-1096-4)

169. Genty E, Breuer T, Hobaiter C, Byrne RW. 2009 Gestural communication of the gorilla (Gorilla gorilla): repertoire, intentionality and possible origins. Anim. Cogn. 12, 527–546. (doi:10.1007/s10071-009-0213-4)

170. Hobaiter C, Byrne RW. 2011 The gestural repertoire of the wild chimpanzee. Anim. Cogn. 14, 745–767. (doi:10.1007/s10071-011-0409-2)

171. Liebal K, Pika S, Tomasello M. 2006 Gestural communication of orangutans (Pongo pygmaeus). Gesture 6, 1–38. (doi:10.1075/gest.6.1.02lie)

172. 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)

173. Liebal K, Oña L. 2018 Mind the gap – moving beyond the dichotomy between intentional gestures and emotional facial and vocal signals of nonhuman primates. Interact. Stud. 19, 121–135. (doi:10.1075/is.17040.lie)

174. Waller BM, Caeiro CC, Davila-Ross M. 2015 Orangutans modify facial displays depending on recipient attention. PeerJ 3, e827. (doi:10.7717/peerj.827)

175. Scheider L, Waller BM, Oña L, Burrows AM, Liebal K. 2016 Social use of facial expressions in hylobatids. PLoS ONE 11, e0151733. (doi:10.1371/journal.pone.0151733)

176. Heesen R et al. 2021 Assessing joint commitment as a process in great apes. iScience 24, 102872. (doi:10.1016/j.isci.2021.102872)

177. Kaminski J, Call J, Tomasello M. 2004 Body orientation and face orientation: two factors controlling apes' behavior from humans. Anim. Cogn. 7, 216–223. (doi:10.1007/s10071-004-0214-2)

178. Tempelmann S, Kaminski J, Liebal K. 2011 Focus on the essential: all great apes know when others are being attentive. Anim. Cogn. 14, 433–439. (doi:10.1007/s10071-011-0378-5)

179. Tomonaga M, Imura T. 2010 Visual search for human gaze direction by a chimpanzee (Pan troglodytes). PLoS ONE 5, e9131. (doi:10.1371/journal.pone.0009131)

180. Rossano F. 2013 Sequence organization and timing of bonobo mother-infant interactions. Interact. Stud. Soc. Behav. Commun. Biol. Artif. Syst. 14, 160–189. (doi:10.1075/is.14.2.02ros)

181. Pika S, Mitani J. 2006 Referential gestural communication in wild chimpanzees (Pan

15

royalsocietypublishing.org/journal/rstb Phil. Trans. R. Soc. B 377: 20210094

troglodytes). *Curr. Biol.* **16**, R191–R192. (doi:10.1016/j.cub.2006.02.037)

182. Rossano F, Liebal K. 2014 'Requests' and 'offers' in orangutans and human infants. In *Requesting in social interaction* (eds P Drew, E Couper-Kuhlen), pp. 335–364. Amsterdam, The Netherlands: John Benjamins.

183. Pollick AS, de Waal FBM. 2007 Ape gestures and language evolution. *Proc. Natl Acad. Sci. USA* **104**, 8184–8189. (doi:10.1073/pnas.0702624104)

184. Halina M, Liebal K, Tomasello M. 2018 The goal of ape pointing. *PLoS ONE* **13**, e0195182. (doi:10.1371/journal.pone.0195182)

185. Bross F. 2020 Why do we shake our heads?: on the origin of the headshake. *Gesture* **19**, 269–298. (doi:10.1075/gest.17001.bro)

186. Schneider C, Call J, Liebal K. 2010 Do bonobos say NO by shaking their head? *Primates J. Primatol.* **51**, 199–202. (doi:10.1007/s10329-010-0198-2)

187. Vick S-J, Waller BM, Parr LA, Smith Pasqualini MC, Bard KA. 2007 A cross-species comparison of facial morphology and movement in humans and chimpanzees using the Facial Action Coding System (FACS). *J. Nonverbal Behav.* **31**, 1–20. (doi:10.1007/s10919-006-0017-z)

188. Caeiro CC, Waller BM, Zimmermann E, Burrows AM, Davila-Ross M. 2013 OrangFACS: a muscle-based facial movement coding system for orangutans (*Pongo* spp.). *Int. J. Primatol.* **1**, 115–129. (doi:10.1007/s10764-012-9652-x)

189. Waller BM, Lembeck M, Kuchenbuch P, Burrows AM, Liebal K. 2012 GibbonFACS: a muscle-based facial movement coding system for hylobatids. *Int. J. Primatol.* **33**, 809–821. (doi:10.1007/s10764-012-9611-6)

190. Waller BM, Julle-Daniere E, Micheletta J. 2020 Measuring the evolution of facial 'expression' using multi-species FACS. *Neurosci. Biobehav. Rev.* **113**, 1–11. (doi:10.1016/j.neubiorev.2020.02.031)

191. Oña LS, Sandler W, Liebal K. 2019 A stepping stone to compositionality in chimpanzee communication. *PeerJ* **7**, e7623. (doi:10.7717/peerj.7623)

192. Fröhlich M, van Schaik CP. 2018 The function of primate multimodal communication. *Anim. Cogn.* **21**, 619–629. (doi:10.1007/s10071-018-1197-8)

193. Genty E, Clay Z, Hobaiter C, Zuberbühler K. 2014 Multi-modal use of a socially directed call in bonobos. *PLoS ONE* **9**, e84738. (doi:10.1371/journal.pone.0084738)

194. Genty E. 2019 Vocal–gestural combinations in infant bonobos: new insights into signal functional specificity. *Anim. Cogn.* **22**, 505–518. (doi:10.1007/s10071-019-01267-0)

195. Hobaiter C, Byrne RW, Zuberbühler K. 2017 Wild chimpanzees' use of single and combined vocal and gestural signals. *Behav. Ecol. Sociobiol.* **71**, 96. (doi:10.1007/s00265-017-2325-1)

196. Fröhlich M, Bartolotta N, Fryns C, Wagner C, Momon L, Jaffrezic M, Mitra Setia T, van Noordwijk MA, van Schaik CP. 2021 Multicomponent and multisensory communicative acts in orang-utans may serve different functions. *Commun. Biol.* **4**, 1–13. (doi:10.1038/s42003-021-02429-y)

197. Eckert J, Winkler SL, Cartmill EA. 2020 Just kidding: the evolutionary roots of playful teasing. *Biol. Lett.* **16**, 20200370. (doi:10.1098/rsbl.2020.0370)

198. Liebal K, Call J, Tomasello M, Pika S. 2004 To move or not to move: how apes adjust to the attentional state of others. *Interact. Stud.* **5**, 199–219. (doi:10.1075/is.5.2.03lie)

199. Dafreville M, Hobaiter C, Guidetti M, Sillam-Dussès D, Bourjade M. 2021 Sensitivity to the communicative partner's attentional state: a developmental study on mother–infant dyads in wild chimpanzees (*Pan troglodytes schweinfurthii*). *Am. J. Primatol.* **83**, e23339. (doi:10.1002/ajp.23339)

200. Bohn M, Call J, Tomasello M. 2016 The role of past interactions in great apes' communication about absent entities. *J. Comp. Psychol.* **130**, 351–357. (doi:10.1037/com0000042)

201. Leavens DA, Reamer LA, Mareno MC, Russell JL, Wilson D, Schapiro SJ, Hopkins WD. 2015 Distal communication by chimpanzees (*Pan troglodytes*): evidence for common ground? *Child Dev.* **86**, 1623–1638. (doi:10.1111/cdev.12404)

202. Genty E, Neumann C, Zuberbühler K. 2015 Bonobos modify communication signals according to recipient familiarity. *Sci. Rep.* **5**, 16442. (doi:10.1038/srep16442)

203. Leavens DA, Russell JL, Hopkins WD. 2005 Intentionality as measured in the persistence and elaboration of communication by chimpanzees (*Pan troglodytes*). *Child Dev.* **76**, 291–306. (doi:10.1111/j.1467-8624.2005.00845.x)

204. Roberts AI, Vick S-J, Buchanan-Smith HM. 2013 Communicative intentions in wild chimpanzees: persistence and elaboration in gestural signalling. *Anim. Cogn.* **16**, 187–196. (doi:10.1007/s10071-012-0563-1)

205. Cartmill EA, Byrne RW. 2007 Orangutans modify their gestural signaling according to their audience's comprehension. *Curr. Biol.* **17**, 1345–1348. (doi:10.1016/j.cub.2007.06.069)

206. Heesen R, Fröhlich M, Sievers C, Woensdregt M, Dingemanse M. 2022 Coordinating social action: a primer for the cross-species investigation of communicative repair. *Phil. Trans. R. Soc. B* **377**, 20210110. (doi:10.1098/rstb.2021.0110)

207. Pika S, Wilkinson R, Kendrick KH, Vernes SC. 2018 Taking turns: bridging the gap between human and animal communication. *Proc. R. Soc. B* **285**, 20180598. (doi:10.1098/rspb.2018.0598)

208. Fröhlich M, Kuchenbuch P, Müller G, Fruth B, Furuichi T, Wittig RM, Pika S. 2016 Unpeeling the layers of language: bonobos and chimpanzees engage in cooperative turn-taking sequences. *Sci. Rep.* **6**, 25887. (doi:10.1038/srep25887)

209. Burkart J, Adriaense J, Brügger R, Miss F, Wierucka K, van Schaik C. 2022 A convergent interaction engine: vocal communication among marmoset monkeys. *Phil. Trans. R. Soc. B* **377**, 20210098. (doi:10.1098/rstb.2021.0098)

210. Fröhlich M, Wittig RM, Pika S. 2016 Should I stay or should I go? Initiation of joint travel in mother–infant dyads of two chimpanzee communities in the wild. *Anim. Cogn.* **19**, 483–500. (doi:10.1007/s10071-015-0948-z)

211. Pika S, Fröhlich M. 2019 Gestural acquisition in great apes: the Social Negotiation Hypothesis. *Anim. Cogn.* **22**, 551–565. (doi:10.1007/s10071-017-1159-6)

212. Schegloff EA. 2007 *Sequence organization in interaction: a primer in conversation analysis,* vol. 1. New York, NY: Cambridge University Press.

213. Stivers T, Sidnell J. 2013 *The handbook on conversation analysis*. Chichester, UK: Wiley-Blackwell.

214. Genty E, Heesen R, Guéry J-P, Rossano F, Zuberbühler K, Bangerter A. 2020 How apes get into and out of joint actions: shared intentionality as an interactional achievement. *Interact. Stud.* **21**, 353–386. (doi:10.1075/is.18048.gen)

215. Wilkinson R, Leudar I, Pika S. 2012 *Requesting behaviors within episodes of active sharing*. Amsterdam, The Netherlands: John Benjamins.

216. Pouw W, Paxton A, Harrison SJ, Dixon JA. 2020 Acoustic information about upper limb movement in voicing. *Proc. Natl Acad. Sci. USA* **117**, 11 364–11 367. (doi:10.1073/pnas.2004163117)