CellPress
OPEN ACCESS

Trends in
Cognitive Sciences

Review

Facial clues to conversational intentions

Judith Holler [1,*]

It has long been known that we use words to perform speech acts foundational to everyday conversation, such as requesting, informing, proposing, or complaining. However, the natural environment of human language is face-to-face interaction where we use words and an abundance of visual signals to communicate. The multimodal nature of human language is increasingly recognised in the language and cognitive sciences. In line with this turn of the tide, findings demonstrate that facial signals significantly contribute to communicating intentions and that they may facilitate pragmatically appropriate responding in the fast-paced environment of conversation. In light of this, the notion of speech acts no longer seems appropriate, highlighting the need for a modality-neutral conception, such as social action.

Highlights

Human language is multimodal, consisting not just of speech but also visual bodily signals.

Facial signals form part of this multimodal system and contribute to how we accomplish social actions in conversational interaction, such as requesting, informing, proposing or complaining, and the interactional responses that follow.

Such facial signalling strategies seem to be culturally specific, with some clear associations between specific facial signals (or combinations thereof) and social actions within groups, but some basic commonalities across cultures may exist.

A processing account is proposed suggesting how facial signals feed into social action attribution and response planning during conversational interaction.

The recent developments call for a radical reconceptualization of long-standing notions and concepts, first and foremost abandoning the idea of speech acts.

Rethinking intention communication in conversation

That idea that words are more than mere descriptions of the state of the world has long been acknowledged. We do things with words [1,2]. This is particularly evident when words perform actions defined by set rules, such as pronouncing two people legal partners, promising to pay dues in a financial transaction, or mandating that citizens behave in a particular way according to a constitution. But we also do things with words in everyday conversation [3–5] when we produce utterances to request, offer, warn, promise, criticise, and so forth. Traditionally, this has been referred to as performing **speech acts** (see Glossary) [1,2]. The primary thesis of this article is that a purely speech-based account does not provide a complete picture of how utterances perform social actions. This claim is rooted in the fact that the natural environment of human language is face-to-face social interaction [6], the very environment in which language has evolved, is acquired, and is used most frequently.

Among the various visual bodily articulators, the face is a prime contender for communicating intentions in conversation. With its 43 muscles, the face is a highly expressive communication tool, capable of many possible combinations of muscle movements – seven for the different eyebrow moving muscles alone [7], and >7000 combinations of voluntarily elicited muscle movements have been documented for the face as a whole [8]. This complexity results in an impressive array of facial **signals**, which can be combined with various gaze directions to create an even broader range of complex facial expressions. Recent research indicates that people make frequent use of the communicative potency of the face in conversation, producing an average of one hundred facial movements per minute [9]. This is intriguing. However, to date, facial expressions have been studied primarily in the context of emotions, where they are often deemed to be the result of universal, reflexive reactions to emotional experiences [10]. This is fundamentally different from facial expressions that convey intentions, where the facial expressions originate in a mental representation of the communicative intent rather than an emotional experience [7,11].

This Review brings together recent evidence from various disciplines, including psycholinguistics, cognitive neuroscience, and conversation analysis, showing that the facial signals produced by speakers significantly contribute to what their utterances are intended to accomplish in

[1]Donders Institute for Brain, Cognition & Behaviour, Radboud University Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

*Correspondence:
judith.holler@mpi.nl (J. Holler).

conversation, both in production and perception. This suggests we need a radically different conception of how participants in conversation convey their intentions, one which comprises both words and visual signals. As a corollary, this article advocates for replacing speech-centred terminology with modality-neutral terminology to capture the multimodal nature of intention communication. The term **social action** [3,12] offers such an alternative, while also emphasising the inherently other-directed, socially embedded nature of utterances traditionally referred to as speech acts.

This Review begins by briefly illustrating the challenges that social action communication in conversation can pose. It then reviews evidence from conversational corpora and experiments investigating the role of facial signals in this type of communication. These studies suggest some stable form-function mappings between specific social actions and facial signals in production and clear influences on social action recognition. The Review then focuses on the timing of facial signals in relation to verbal utterances. When facial signals occur early in a conversation, they may foreshadow information about the social action of the forthcoming turn. All of this has implications for the cognitive processing of utterances in conversation, as explored in the final part of the article, and may point towards predictive mechanisms with deep evolutionary roots.

Ambiguities in communicating social actions

While some regularity exists between specific grammatical forms and social actions, these connections are far from clear one-to-one mappings [4], and in terms of their surface form, many social actions come disguised as something other than what they are intended to do [13]. Even questions, one of the most common turn formats, can convey many different social actions [14]. For example, 'How old are you?' may be intended as a straightforward request for information, as an expression of surprise at someone's early achievements, or as a criticism of another's actions. Words alone thus create considerable room for social action ambiguity [5,15], resulting in the intriguing question of how interlocutors deal with it. There is evidence that **prosody** can convey critical social action-relevant information, but this is certainly not always so, thus providing part of the answer at most (Box 1). With words and prosody not telling the full story, the visual modality may contribute an important missing piece of the puzzle. The following sections detail some of the ways in which interlocutors' faces may do so.

The face as a powerful tool for pragmatic communication

In face-to-face interaction, language is deeply **multimodal** in nature. In hearing people, language consists of visual bodily signals interleaved with words and other vocal information [11,16–21]. Hence, if we aim to understand how interlocutors navigate social interaction, we must consider

Box 1. Prosody and social actions

Prosody can be instrumental in conveying different social actions in experiments where the melody of scripted utterances distinguishes between these actions [95,96]. Prosody can also be critical in free conversation, such as pitch modulations distinguishing agreements from disagreements [97]. When looking at questions in particular, speakers may deviate in their initial pitch when asking questions intended to solicit agreement compared to those seeking information [98], thus providing an early clue to the social action of the turn underway. Also, the pitch range of questions may distinguish between those seeking information and those designed to challenge what has been said [99]. However, while there is clear evidence that prosody can convey social-action-relevant information in conversation, this is not always so. Even the common idea that questions are marked by rising intonation and statements by falling intonation is highly contested [76,100–103]. For example, prosodic forms of yes-no questions in English conversations were produced with falling intonation in around half the cases (43% American English [101], 64% British English [102]), and even questions produced with declarative forms (thus lacking any syntactic marking of questionhood) often have falling intonation [102,104]. Thus, like with linguistic forms, there is ambiguity and no one-to-one mapping between prosodic patterns and social actions exists.

Glossary

Dispreferred responses: broadly speaking, preferred responses are those that perform an affiliative action and dispreferred ones those that perform a disaffiliative one. In response to an invitation, an acceptance would be perceived as more affiliative and less face-threatening than a declination, and hence preferred.
Front-loading: accumulation of information at the beginning of an utterance, or speaking turn, such as early information that marks the utterance as being a question. In many western languages, this would include lexical, syntactic, or prosodic information (e.g., WH-words, subject–verb inversion, utterance-initial pitch marking).
Multimodal language: conceptualization of human language as consisting of verbal and visual forms, including sign languages, but also the concurrent expression of mental representations in the form of speech and visual signals from various bodily articulators (hands, face, head, etc.) as jointly constituting language.
Prediction: broadly speaking, the preactivation of information before it can be processed based on (the full) input. This can relate to the preactivation of linguistic representations at various levels but also to nonlinguistic information such as a communicative intention or a physical action.
Predictive potential: visual information (e.g., facial signals or gestures) that anticipates semantically or pragmatically related information in the speech stream has the potential to predict, or facilitate the prediction of, this upcoming information. The term refers to the information the signal carries and its timing, not any certain effect in the comprehender.
Prosody: auditory information not tied to individual sounds but larger structures such as words or longer utterances, for example, intonation contours (speech melody).
Psychophysical studies: these tend to follow a more data-driven approach in that a wider range of stimuli (e.g., facial signals) are used in many per-participant trials, often combined with a reverse correlation procedure to arrive at perceptions in the absence of prespecified predictions or patterns.
Signal: here used in the sense of movements or vocalisations that form part of the effort of producing a

the involvement of the body [22]. While research in recent decades has focused on manual gestures and their semantic contributions, the face has received little attention, because, traditionally, its main function is considered emotion expression [23,24]. However, investigations into how facial signalling works in combination with spoken language, and in conversation in particular, have only taken off comparatively recently [11]. Analyses of face-to-face conversational corpora have provided insights that hint at some relatively stable mappings between specific facial movements and social actions, at least within a particular language or culture.

Much of this work has focused on question–response environments. Questions are incredibly frequent in conversation [14] (on the order of three to four per minute in Dutch casual conversation, for example [25]), thus making up a large proportion of conversational turns. Interlocutors must quickly identify questions amid other social actions, as questions will often call for a response [26]. Questions thus come with specific social contingencies, and they require the next speaker to be prepared for them. Determining which responsive social action is pragmatically appropriate requires an addressee to identify what the questioner is trying to do, and they must do so swiftly [27].

The following sections illustrate the pragmatic power of facial signals in conversation, specifically their role in composing utterances intended to accomplish particular social actions and in facilitating their understanding by addressees. Most of this work focuses on questions and responses, but some of it also on expressing rejective messages that include other types of utterances (e.g., assessments).

Some stable form-function mappings between facial movements and social actions

Gaze direction is a prominent marker of questionhood. For example, in Dutch, English, French, Italian, Catalan, Japanese, and Yélî Dnye conversations, questions to which speakers seek responses are typically uttered with gaze oriented towards the addressee [28–33]. This contrasts with gaze direction during self-directed questions [25] or statements [30]; that is, social actions that hold interactants less accountable for a response.

From several sign languages, we know that eyebrow actions, too, can mark questionhood [e.g., raised eyebrows marking polar (yes/no), and furrowed brows marking WH-questions [e.g., 34–36]]. Parallels exist in spoken communication, where, in English, French and Dutch, for example, raised and furrowed brows often occur together with questions, including those asking for clarification [7,25,30,37–39]. The pattern may be even more fine-grained than that, since open repair requests (e.g. 'what?', 'excuse me?', 'huh?') occur predominantly with eyebrow raises, and restricted repair requests (e.g. 'who?', 'he did what?') occur mainly with furrowed brows, at least in Dutch [40]. This fits with the observation that eyebrow frowns, more generally, often express a lack of certainty, doubt or perplexion [7, p.185].

In addition to insights from conversational corpora, there is evidence from experiments where participants produced scripted utterances performing different social actions. In English, Dutch, Italian, and Catalan information seeking polar questions (i.e., yes/no questions) with interrogative prosody tended to be marked by raised eyebrows [28,41,42]. In contrast, incredulity questions (expressing doubt, uncertainty or scepticism) were produced with furrowed eyebrows and squinting of the eyes in Dutch, Catalan, and Brazilian Portuguese [43–45].

Facial signals have also been studied with social actions expressing a negative or rejective attitude. In Mandarin Chinese conversation, utterances performing negative assessments (i.e. expressing a negative stance relating to the content of a previous utterance) are often produced

communicatively intended message. This does not necessarily entail awareness of the signal being produced and can include behaviours performed by habit (such as gaze aversion as a face-saving strategy).
Social action: conveying one's conversational intentions (to offer, request, complain, criticize, agree, propose and so forth), traditionally referred to as speech acts. Rather than being focused on the individual utterance or the producer's cognitive state, the notion of social action is rooted in conversational interaction where each turn either initiates or responds to another turn's social action.
Speech act: words that do not merely provide information but convey a speaker's intention by performing actions, as it were (e.g., to offer, request, complain, criticize, agree, or propose).

with eyebrow furrows, as well as pursed or pouted lips [46]. In an experimental study, furrowed eyebrows also marked utterances intended to convey (seemingly harsh) orders in Italian speakers, often together with widened eyes [42]. And in English, Czech, French, Hebrew, Mandarin, and Romanian conversation, producers of **dispreferred responses** tend to avert their gaze away from the questioner [47,48]. In English, Spanish and Mandarin Chinese speakers, experimentally elicited assertions and questions including negation were accompanied by what has been termed the 'not face', involving a combination of furrowed eyebrows, pressed lips, pulled back corners of the mouth, and/or an upward contraction of the chin muscle [49].

Together, the findings from both corpus and experimental production studies point towards some stable mappings between facial signals and social actions, especially for eyebrow movements and gaze direction.

Social action ascription based on facial signals

There is some experimental evidence that facial signals affect social action ascription in perceivers. Most of these studies asked actors to express scripted materials as naturally as possible and required participants to distinguish between different types of social actions, doing so based on auditory information alone or in conjunction with visual signals (and sometimes based on the visual signals alone). Such studies have shown, for example, that utterances are indeed more likely to be considered information seeking questions in Catalan and Dutch when accompanied by eyebrow raises [28,44], and as questions conveying doubt, scepticism or lack of certainty when accompanied by eyebrow furrows [28,43,44]. These effects were driven primarily by the visual signals alone, independently of lexical or prosodic cues. The effect was strongest in Catalan, possibly due to the lack of syntactic marking and subtler acoustic social action differentiation (but in other languages, vocal cues may outweigh visual ones, such as in Brazilian Portuguese [45]). Nota *et al*. [50,51] mapped questions and statements from naturalistic Dutch conversations onto an avatar (preserving the original speech and prosody), showing that the presence of eyebrow frowns (and to some extent brow raises) significantly improved participants' accuracy in distinguishing questions from statements. This corroborates the strong influence of conversational facial signals on social action ascription even when information from syntax, semantics and prosody is available.

Other studies have manipulated the combination of facial signals to investigate how individual movement components contribute to differences in perception. In Italian [42], utterances were perceived as questions mainly with outer brow raises or with inner brow raises+brow furrow; inner brow raises+brow furrow, in turn, were perceived as incompatible with orders, thus discriminating well between these two social actions. However, utterances were mainly perceived as orders when produced together with eyebrow furrows and eye widenings, or their combination. Eyebrow furrows+eye widenings, in turn, were perceived as significantly less compatible with questions. Assertions were not associated with any facial movements but with a blank face instead. In another study, polar questions in Dutch were perceived as primarily information seeking when combined with eyebrow raises, but when combined with head tilts or turns, the utterances were primarily perceived as seeking clarification or expressing scepticism [52]. This highlights the pragmatic power of facial (and head) movements as they appear able to fundamentally change the social action ascribed to a linguistic utterance, and it demonstrates that multimodal utterances are compositional in nature.

The combined effects of facial signals have also been demonstrated in **psychophysical studies** (though often in the absence of speech). In German speakers, the combination eyebrow furrow+squint+slightly open mouth was perceived as expressing confusion [53]. Other studies

showed that Western participants associate brow raises, eye widening, nostril dilation and mouth opening with 'acceptance messages' (including surprise), and brow lowering, wincing, nose wrinkling, and lip pinching with 'rejection messages' (including confusion, fear, disgust, anger, and sadness) [54]. Unique facial movement combinations were also found for expressions of apology, involving frowning, lateral lip stretching, smiling, and averted gaze [55].

In summary, perception experiments also provide evidence of some stable mappings between facial signals and social actions, especially for eyebrow movements, gaze direction, and some mouth movements. While the research reviewed in the preceding sections sheds light on connections between facial signals and social actions for different languages and cultures, much more systematic comparative work is needed to help us understand where the commonalities and differences lie (see Outstanding questions).

Summary and synthesis

Although insights from acted and spontaneously produced facial expressions can differ [23], the various studies on production and perception reviewed here converge on eyebrow raises and furrows being facial actions frequently associated with questions.

Moreover, eyebrow furrows seem to express problems in understanding, a lack of certainty, doubt, or scepticism, and their general function may be to seek a corrective, elaborative or justificatory response from the addressee. Non-question turns can also be accompanied by furrowed brows. Here, the furrow seems to accompany negative assessments, rejections, or orders, for example. Thus, while furrowed brows seem to be particularly typical for questions [37], they appear associated with social actions that have a negative or rejective connotation for both questions and non-questions. While this underlines that there is no one-to-one mapping between facial signals and social actions, it does hint at stable mappings between individual signals and the broader type, or valence, of social actions at least.

Where such stable mappings originate is an open question due to crosscultural and cross-species commonalities as well as differences in facial expressions (Box 2). What is interesting is that furrowed brows constitute a shared component between the human aggression face and concentration face. The concentration face has been proposed to signal that all the information the signaller can handle at that moment has been received and no more is momentarily wanted [56]. In both cases, the furrowed eyebrows might thus function as a signal that stops the progression of the ongoing interaction to give room to the signal producer's goals and needs in that moment and, as a consequence, avoid more severe conflict [56]. This includes conflict of a cognitive sort (such as an addressee's lack of understanding, uncertainty, or disagreement). Such an interpretation would fit with the idea that facial signalling, in general, is associated with conflict reduction and stronger social bonds [57,58]. Also, unique facial movement combinations were found for visual expressions of apology, including a downward-averted gaze component [55]. This seems to converge with speakers of different languages marking disaffiliative responses with downward-averted addressee [47,48], which may be interpretable as a visual form of apology for the potentially face-threatening response to come.

Taken together, the findings suggest consistent associations between individual facial signals and social actions, at least at the level of affiliative versus disaffiliative (or preferred versus dispreferred) social actions. Moreover, the compatibility of some signals, or signal combinations, with specific social actions, as well as incompatibility with others [42], may make them a critical tool for guiding fast social action interpretation, especially in the context of predictive social action signalling as a conflict avoiding strategy [47,57,58].

Box 2. Where do facial expressions of social actions come from?

Darwin [105] hypothesized that the origin of facial expressions of emotion lies in instinctive, hard-wired connections between bodily movements and internal sensations, thus explaining the large pancultural similarities in facial expressions he described. His views have shaped facial expression research for many decades, leading to the idea of six basic, culturally universal emotion categories [106]. Recent research using large datasets of video recordings from across the globe and machine-learning methods seem to corroborate the idea of culturally universal facial emotion expressions [107]. However, corpus studies, elicitation paradigms and psychophysical experimentation with people from Western and non-Western backgrounds highlight cultural variation in facial expressions and in their perception [23,92,108]. Nose wrinkling, as one specific example, accompanies negatively valanced addressee responses in Dutch conversation [25] but expresses positive astonishment in Yélî Dnye [5], a Papuan language isolate.

Comparative research across species also casts doubt on universal, biologically hard-wired facial expressions in humans, offering a more refined picture instead. There is convincing evidence for certain similarities suggestive of continuity of facial expressions between human and non-human primates. For example, the silent bared teeth (SBT) display occurs mainly in appeasement/fear-related contexts and is seen as the precursor to the human smile; the relaxed open mouth (ROM) display (the so-called 'play-face') mainly occurs in play and affiliative contexts in primates as well as some non-primate species and is considered the precursor to the facial actions accompanying human laughter (with their blended forms in humans and some non-human primate species pointing towards possible links between more relaxed dominance hierarchies and their emergence) [56]. However, for facial expressions of anger or aggression which in humans tend to feature furrowed brows, deep evolutionary roots seem more unlikely. While non-human primates had originally been claimed to also furrow their brows in such contexts [109], systematic, fine-grained approaches using the Facial Action Coding Scheme adapted for chimpanzees revealed that the brow furrow as it occurs in modern humans (AU4) is actually not present in chimpanzees [56,110]. Regarding eyebrow movements for social action communication, intraspecies learning and ritualization therefore seem to be a more plausible explanation. One possibility is that through the process of human self-domestication, which led to flatter foreheads and highly mobile eyebrows instead of a brow ridge, eyebrow movements became salient social signals [111]. These anatomical changes correlate with the evolution of more complex social relationships and fit the idea that facial expressions are linked to social advantage [57]. Research on domesticated felines revealed morphological differences between facial signals in affiliative versus disaffiliative contexts, further corroborating the idea of an evolutionary link between domestication, facial signalling and social relationships [112].

Facial signals have the potential to speed up social action ascription

Because social action ascription has to happen quickly in conversation [4,27], additional facial clues to the social action a turn is performing may help. Recent findings reveal that multimodal language stimuli showing a speaker's body (including facial information) lead to facilitated language processing in a shadowing task [59]. There is evidence that the presence of a static face by itself facilitates language processing [60], and facial movements, too, can speed up processing,such as eyebrow movements leading to faster question identification [51,61]. This facilitation might be attributed to the additional information that facial signals provide to the receiver. However, at which point in time this additional information occurs in relation to the verbal information may also play a role. From conversational corpora, we know that facial signals often occur early on during, or even prior to, a speaking turn. This holds for eyebrow frowns and raises accompanying repair initiations in Dutch [40], frowns and smiles in Finnish conversation foreshadowing troublesome or humorous turn content, respectively [62,63], and gaze aversion preceding disaffiliative responses in English conversation (often occurring while the question is even still underway) [47]. Moreover, quantitative analyses of Dutch conversations have shown that most facial signals occur early in relation to speaking turns (mainly within the first 40 ms of the utterance)[37]. This early timing fits the idea of **front-loading** in conversational turns [4] and parallels the anticipatory production of manual and other bodily movements [64–66]. Early, anticipatory production equips movements with **predictive potential**, which may lead to faster responses [64,65] based on predictive cognitive processing [67], in this case **predictions** about the social action of the unfolding turn (but, see Box 3 for instances of late facial signalling). Data on gaze aversion in English conversation are compatible with this idea. Examples showed that questioners seemed to respond to early gaze aversions by rephrasing their question prior to the dispreferred response being issued in an attempt to increase the likelihood of obtaining a

Box 3. Facial signals conveying social action information late

Quantitative corpus studies of Dutch conversation [25,37] have shown that in the majority of cases, facial signals occur before or early on during conversational utterances, but this is not always so. Smiles occurred at any point during questions, including at utterance end, for most of the social actions investigated. Other mouth movements (pressed lips and mouth corners down) occurred mainly towards utterance end; mouth movements may of course be more difficult to incorporate mid-stream of an utterance when the lips are busy speaking, but we see late timing also for other signals. Unilateral eyebrow raises, for example, occurred early with clarification questions, but towards utterance end in questions putting forward proposals about joint plans and actions (e.g., 'how about lunch together?'). French data [113] show that, amongst others, eyebrow flashes (which occurred throughout utterances, including at the end) were associated with utterances performing ironic assessments, whereas sustained eyebrow raises were associated with nonironic, sincere assessments, and English data show similar eyebrow irony marking [114]. Analyses of assessments during story telling by Finnish speakers showed that facial signals often occurred late during utterances [115], or at the very end of stories even, such as a smile on the last word which delivered the punchline. Analyses of English conversations have also shed light on some late facial signals, such as eyebrow raises following turn completion (i.e., in the turn transition space) [116]. Detailed form analyses revealed that sustained eyebrow raises in this position were used when pressing for a response to a disaffiliative (e.g. challenging) action. However, quick eyebrow flashes that occurred in this same position were used in affiliative contexts, such as inviting a response to a joke. Thus, fine-grained kinematic differences, too, seem to be indicative of differences in social action, and their later temporal placement may help to deliver the intended interactional effect in some cases. Certain social actions may thus benefit precisely from not being recognized early.

Note that this Review covers facial signals contributing to social actions, and that the timing relationships for facial signals functioning in other ways, such as by contributing to mark information structure [117,118] or performing visual enactments [119–121] is likely to pattern in quite different ways. This may also hold for visual expressions of speaker stance (see [122] for an overview of multimodal stance expressions).

preferred response instead [47] (Figure 1). And a recent experimental study corroborated that facial signals, including gaze aversion, can be perceived as predictive of the social action to follow, at least in off-line judgements [68] (Figure 1).

This demonstrates that facial signals can and do contribute to social action ascription, over and above prosodic, lexical, and syntactic information, and that they have predictive, and thus strong interactional value in terms of their early production [see Box 4 for facial signals coordinating interactions through (re)calibration and resonance but not necessarily through early occurrence]. One next step is obtaining experimental evidence that facial signals indeed influence social action recognition online during conversation (see Outstanding questions).

Although studies often foreground stable mappings between individual or just a few facial signals and social actions, it is important to note that these often form part of more complex constructs. In addition to multiple facial signals occurring in combination [25,39,69], they may also combine with signals from other articulators, such as eye gaze [70] and head movements [71], underlining the multimodal combinatorial power of utterances. Moreover, manual gestures (e.g., the palm-up gesture), also fulfil social action-related functions (such as offering and requesting in human [19,72] and non-human primates [73]) and postural changes may do too [74,75]. How these signals combine with facial signals and gaze into multi-articulator social-action-related displays is largely uncharted terrain (see Outstanding questions). Prosodic patterns and systematicities in conversational structures add further layers [15,76]. Multimodal utterances constructed of these various articulators and signals they convey results in multilayered, complex gestalts [16,77,78]. Within these, there is no exclusive one-to-one mapping between individual facial signals and social actions, as each signal may contribute to several different gestalts. But what the studies reviewed here draw attention to is some recurring, stable connections that make certain facial signals more likely to be paired with a certain range of social actions than others. These statistical associations may be further refined through the presence of signals in the other layers, thus resulting in complex constructions that are quite unambiguously indicative of specific social actions.

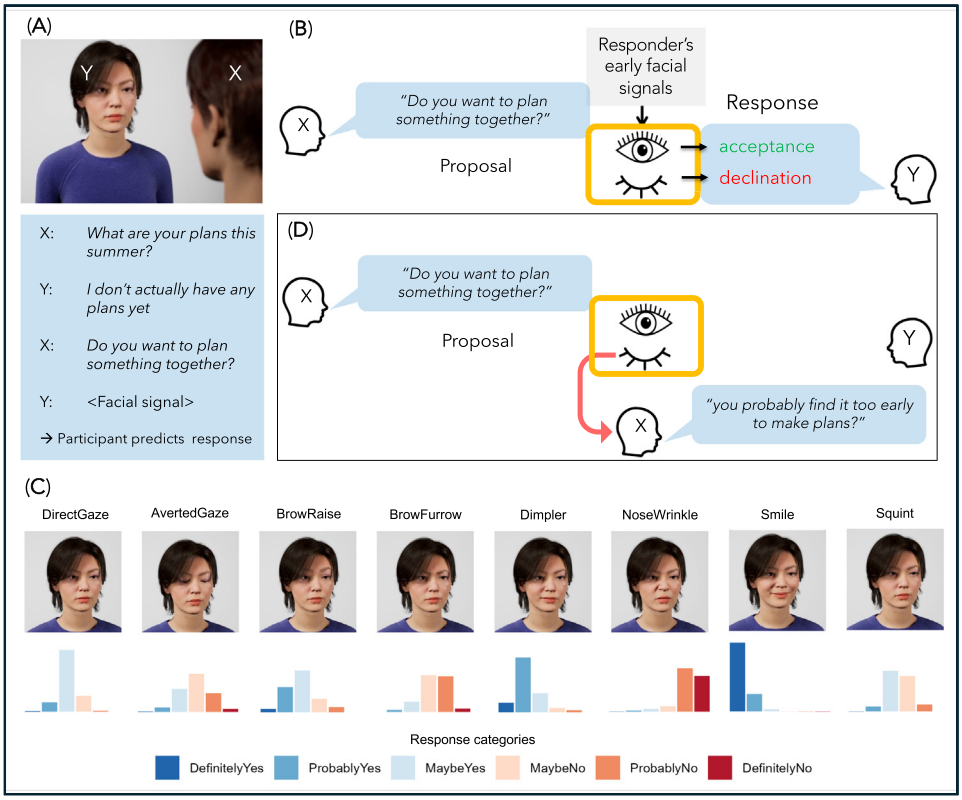

Figure 1. Responders' facial signals predictive of upcoming social actions. Panel A depicts a recent experimental paradigm that embeds facial signals in conversational sequences to measure their predictive potential. Panel B illustrates the differential effects of two of the facial signals manipulated in the study, namely responder gaze directed at the questioner and responder gaze averted from the questioner. It shows the differential predictions made when perceiving the different facial signals that occur in response to the same proposal, namely the prediction of an acceptance as the responding social action when seeing direct responder-gaze, and the prediction of a rejection as the responding social action when seeing averted responder-gaze. Panel C shows participants' response patterns (from [68]) for the entire range of facial signals tested for in their study, demonstrating the different prediction patterns that result from them. Panel D is an abstraction from the qualitative findings (from [47]), where some questioners were found to refashion the formulation of their question on the perception of the responder's averting gaze. The authors hypothesized that questioners were thus trying to maximize the likelihood of receiving a preferred instead of a dispreferred response. This hypothesis remains to be tested, as does the effect of even earlier responder signals that occur while the question is still underway (as observed in [47]). Together, the studies exemplify an advance towards more interaction-capturing and corpus-informed experimental paradigms.

A processing account of fast multimodal social action ascription in conversation

Based on the evidence in the preceding sections, it is plausible that facial signals, especially early ones, play a pivotal role in the process of fast responding in conversation, at least in terms of a crude, initial ascription of the social action, even if the further development of the utterance encourages fine-tuning. Early facial signals may help comprehenders by constraining the possibility space, narrowing down the number of social action interpretations compatible with the signals observed so far, and continuing to do so as the turn develops and further signals come in. An early eyebrow raise, for example, may select a set of social actions that frequently occur with eyebrow raises (e.g., questionhood and surprise) and reduce the likelihood of those more rarely encountered in the presence of a raise (e.g., orders and disagreements). The set may be narrowed down further if additional facial signals are present (e.g., a smile). Such learned statistical associations would come together with other visual and vocal information streams [16].

Box 4. Coordinating interaction with facial signals

Facial signalling fulfils core functions in conversational interaction beyond facilitating addressees' social action ascription or prediction, especially the coordination of interaction through calibrating on each other's understanding and creating affiliative resonance, involving what is often referred to as addressee backchannelling [123].

Interactional (re)calibration

Facial signalling is also common in interactants communicating their understanding (such as through smiles [124] and long blinks [125,126]), or trouble in understanding (e.g., with furrowed brows [40,127]) or communicating their take on information provided by the current speaker (e.g. sustained eyebrow raises indicating surprise or astonishment, or eyebrow flashes which signal the information is news in German [128], or provide an affirmative addressee response in Yélî Dnye [125]). Addressee responses can occur while the current speaker is still speaking [47,126], giving them concurrent feedback about how the utterance they are in the course of producing is understood and received. This, in turn, provides an opportunity for rerouting the course of the interaction if the feedback suggests this to be advantageous. This could entail abandoning the current turn to include a clarification that increases the chance of being understood (when addressees signal trouble in understanding [40]), or to provide justification (in case addressees signal surprise or unexpectedness [128]). It may also help to increase the likelihood that an addressee is able to respond in an affiliative manner [47] (see Figure 1 in main text). In addition to having the potential to speed up social action ascription and response planning, early facial signals may thus also influence the type of social action an ongoing turn is performing by rerouting, or (re)calibrating, the interaction 'on the fly'.

Interactional resonance

In addition to interactional (re)calibration, concurrent facial social action signalling may also serve to create explicit moments of shared understanding and experience, or shared realities [129]. Interlocutors can more actively participate as addressees (also referred to as providing specific feedback) than by simply issuing tokens of understanding, such as 'yeah' or 'uhu' [11]. In Dutch conversation, they often do this with eyebrow raises (e.g., accompanying utterances such as 'really?', 'wow, did he say that?') and smiling [25], and analyses of English dialogues found addressees responding with facial expressions of concern, fear or wincing during close-call stories [130]. In Finnish story telling, speakers produce assessments with facial signals, allowing their interlocutor to join in, mimic and resonate with the type of assessment conveyed [115]. Facial signals in these cases may facilitate early social action and sentiment ascription, which in turn facilitates coparticipation in the particular assessment conveyed and the affiliative resonance that results. In addition to psycholinguistic consequences speakers' early facial signalling may have, they may also contribute to interlocutors' social and affective coordination during the interaction.

Together with anticipations based on preceding conversational structures [15], they result in a probabilistic representation of the pragmatic interpretation of the utterance, updated with every new bit of information coming in [16,79,80].

Such a temporally unfolding social action signalling system bears resemblance to the dynamic and hierarchically organised facial expressions of emotion. It is suggested that early facial signals may facilitate the basic categorisation of emotions, to be refined as the facial expression evolves through time, and where the early signals bear a strong predictive advantage [81]. This is in line with, for example, the finding that in conversation early facial signals can distinguish between basic social actions (such as raised or furrowed eyebrows as markers of questionhood versus other social actions [37,51,61]); this may be followed by a refinement of the precise social action of the question through other visual signals that one encounters further downstream.

Critically, such a fast, low-level mechanism would have to interface closely with the processes requiring more cognitive computation and more fully-fledged modelling of other minds. Essentially, this requires a type of dual-route, interactive process combining simple heuristics and more complex, slower and costly thinking to accomplish the pragmatics of conversational interaction (Figure 2). Such a dual-level mechanism has been postulated to underpin many different domains of processing, from decision making [82] to emotion processing [83], but also language use. Understanding the perspectives and knowledge of others [84], for example, has been claimed to involve both high-level, more costly reflections, as well as leaner, low-cost processes [85,86]. The latter are conceived of as learned patterns and routines, as well as simple 'one-bit'

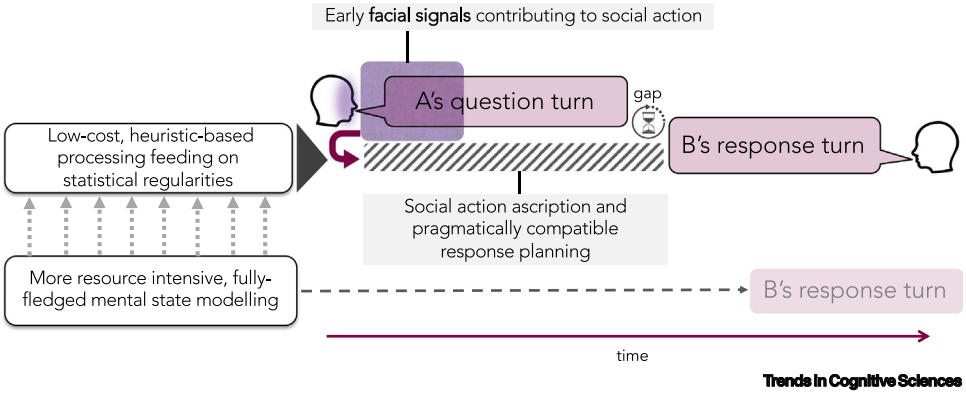

Figure 2. Multimodal social action processing account. Grey square represents facial signals occurring prior to, or early, during A's question turn, thus channelling social action ascription and pragmatically compatible response planning that follows (hatched bar) from the earliest moment (purple bent arrow). This process is based on low-cost processes fuelled by learned patterns based on statistical regularities. This leaner processing route interfaces with a more resource-intensive processing layer which comes into play when more fully-fledged mental state modelling is required. The more costly processing route would likely lead to slower responses.

partner models (e.g., child vs adult addressee [87]) and 'social microprocesses' [88] anchored in the interaction and its environment (e.g. judging someone's attention from their gaze direction). The existence of a shallower, less costly language processing level is also the very essence of the 'good enough' (GN) approach in psycholinguistics [89]. While the GN approach was primarily developed with a focus on meaning derived from syntactic analysis, the authors' conclusion [89, p.14] that 'the goal of the comprehension system might be to deliver an interpretation that is just good enough to allow the production system to generate an appropriate response' nicely aligns with the proposal made here. Facial signalling regularities may be the entryway to a shallow processing route for ascribing (at the very least coarse-grained) social actions quickly to facilitate pragmatically appropriate response planning under the temporal constraints of conversation.

As noted previously, the proposed processing account does not rest on the idea of fixed connections between specific facial signals and social actions. Rather, it capitalises on the statistical regularities that seem to exist between different facial signals, or combinations thereof, and different likelihoods of specific social actions. These associations are most probably learned and ritualised patterns and thus culture and language dependent as suggested by crosscultural and cross-species differences (see Box 2). Nevertheless, the predictive mechanism that might underpin facial social action communication, as argued in the preceding text, may have deep evolutionary roots. Predictive mechanisms allow non-human animals to quickly adapt to constantly changing environments and avoid aversive stimuli. These predictions must happen quickly to guarantee survival [90]. Moreover, recent findings show that non-human primates (e.g., macaques) can use conspecifics' facial expressions to predict next actions [91]. Evolutionarily old mechanisms that help predict the next move, combined with a capacity for using facial signals to facilitate those predictions, may thus underpin social action prediction also in face-to-face human communication. In modern human interaction, predicting next actions may not be a matter of life and death to quite the same extent as in the non-human animal kingdom; however, this mechanism may have been co-opted for navigating the delicate social bonds that our relationships rely on, and where responses are often time-sensitive. Such a perspective fits with approaches that ground human facial expression in the evolution of animal communication, considering them as highly context-dependent social tools, produced with the aim to influence the addressee and navigate social interactions [92,93], and with accounts of layered prediction mechanisms [94].

Concluding remarks

The evidence from corpus analyses, experimental production and perception studies reviewed here suggests that facial signals are an integral component of how social actions are communicated in face-to-face interaction. This conclusion demands a new conceptualisation of the notion of speech act – and terminology, for that matter – namely, one that subscribes to the visual foundations of human language. Modality-neutral concepts and terms, such as illocutionary force and social action, are clearly to be favoured in the light of the evidence presented here, but it is paramount that they be understood as referring to multimodal notions of those concepts. The putative processing mechanism sketched in the preceding text highlights the need to embed such a new notion of multimodal illocutionary force, or multimodal social action, within a framework of psycholinguistic processing in face-to-face conversation.

More fine-grained investigations of the interplay between conversational structures and facial signalling for social action communication, accompanied by a move towards experimental paradigms informed by naturalistic behaviour and able to capture more of these social embeddings, are also urgently needed. A better understanding of the interplay between facial and other visual, as well as verbal and vocal behaviours, is another important next step, and this needs to cover a much wider range of languages, cultures, and groups, as well as the rapidly rising challenge of artificial agents' conversational behaviour (see Outstanding questions). The research reviewed here provides the stepping-stone for taking these next steps by highlighting the relevance of the face for conveying and understanding intentions in face-to-face conversation, the very environment in which the capacity of our species for sharing thoughts and intentions has evolved.

Outstanding questions

How does facial signalling for social action compare in different languages and cultures (including sign languages, where nonmanuals can convey grammatical information, and cultures where positioning one another face to face, or attending to faces, may be less common)?

How do different facial signals combine with one another to form more complex displays, and to what extent may such combinations distinguish between different, perhaps more fine-grained social actions? Does the temporal structure of such complex facial expressions relate to differences in pragmatic meaning?

How do facial signals interact with other visual bodily signals, words, syntax and prosody? How flexible is this multimodal, multiarticulator system? How is it impacted by environmental circumstances (e.g., visual access and noise)?

To what extent do facial signals influence predictive online processing? Can we tease apart whether facial signals facilitate predictive processing versus integration during comprehension, as has been done for the semantic contribution of hand gestures?

How does facial signalling in social interaction compare across species? Can we glean further insights into how facial signalling for social action communication evolved?

What is the degree of interindividual variation using these different modes of communication for accomplishing social actions in conversational interaction?

To what extent can individuals with difficulty in facial communication compensate through other means? Do pragmatic difficulties influence the multimodal realisation of social actions and lead to differences or difficulties in conversation?

Do we need, and want, artificial agents producing highly naturalistic facial behaviour for social action communication? Would this be beneficial in some contexts (e.g., creating trust, enhancing the impression

Declaration of interests

The author declares no competing interests.

References

1. Austin, J.L. (1962) *How to Do Things with Words. The William James Lectures Delivered at Harvard University in 1955*, Oxford University Press
2. Searle, J.R. (1969) *Speech Acts: an Essay in the Philosophy of Language*, Cambridge University Press
3. Atkinson, J.M. and Heritage, J. (1984) *Structures of Social Action: Studies in Conversation Analysis*, Cambridge University Press
4. Levinson, S.C. (2013) Action formation and ascription. In *The Handbook of Conversation Analysis* (Stivers, T. and Sidnell, J., eds), pp. 101–130, Wiley-Blackwell
5. Levinson, S.C. (2024) *The Dark Matter of Pragmatics: Known Unknowns*, Cambridge University Press
6. Levinson, S.C. (2006) On the human "interaction engine". In *Roots of Human Sociality: Culture, Cognition and Interaction* (Enfield, N.J. and Levinson, S.C., eds), pp. 39–69, Berg
7. Ekman, P. (1979) About brows: emotional and conversational signals. In *Human Ethology* (Cranach, M. *et al.*, eds), pp. 169–249, Cambridge University Press
8. Cohn, J.F. and Ekman, P. (2005) Measuring facial action. In *New Handbook of Methods in Nonverbal Behavior Research* (Harrigan, J. *et al.*, eds), pp. 9–64, Oxford University Press
9. Rollings, J. *et al.* (2024) Facial behaviour and first impressions in computer mediated communication. *Comput. Hum. Behav.* 161, 108391
10. Ekman, P. (1993) Facial expression and emotion. *Am. Psychol.* 48, 384–392
11. Bavelas, J.B. (2022) *Face-to-Face Dialogue. Theory, Research, and Applications*, Oxford University Press
12. Holtgraves, T.M. (2002) *Language as Social Action: Social Psychology and Language Use*, Lawrence Erlbaum Associates
13. Couper-Kuhlen, E. (2014) What does grammar tell us about action? *Pragmat. Q. Publ. Int. Pragmat. Assoc. (IPrA)* 24, 623–647
14. Enfield, N.J. *et al.* (2010) Question–response sequences in conversation across ten languages: an introduction. *J. Pragmat.* 42, 2615–2619
15. Stivers, T. *et al.* (2023) Ambiguities in action ascription. *Soc. Forces* 101, 1552–1579
16. Holler, J. and Levinson, S.C. (2019) Multimodal language processing in human communication. *Trends Cogn. Sci.* 23, 639–652
17. Levinson, S.C. and Holler, J. (2014) The origin of human multi-modal communication. *Philos. Trans. R. Soc. B* 369, 20130302
18. Vigliocco, G. *et al.* (2014) Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20130292
19. Kendon, A. (2004) *Gesture: Visible Action as Utterance*, Cambridge University Press
20. Enfield, N.J. (2009) *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*, Cambridge University Press
21. Kita, S. and Özyürek, A. (2003) What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *J. Mem. Lang.* 48, 16–32
22. Hadley, L.V. *et al.* (2022) A review of theories and methods in the science of face-to-face social interaction. *Nat. Rev. Psychol.* 1, 42–54
23. Barrett, L.F. *et al.* (2019) Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. *Psychol. Sci. Public Interest* 20, 1–68
24. Ekman, P. (2006) *Darwin and Facial Expression: a Century of Research in Review*, Malor Books
25. Nota, N. *et al.* (2023) Specific facial signals associate with categories of social actions conveyed through questions. *PLoS ONE* 18, e0288104

26. Stivers, T. and Rossano, F. (2010) Mobilizing response. *Res. Lang. Soc. Interact.* 43, 3–31
27. Levinson, S.C. (2016) Turn-taking in human communication – origins and implications for language processing. *Trends Cogn. Sci.* 20, 6–14
28. Borràs-Comes, J. *et al.* (2014) Audiovisual correlates of interrogativity: a comparative analysis of Catalan and Dutch. *J. Nonverbal Behav.* 38, 53–66
29. Rossano, F. *et al.* (2009) Gaze, questioning, and culture. In *Conversation Analysis* (Sidnell, J., ed.), pp. 187–249, Cambridge University Press
30. Torreira, F. and Valtersson, E. (2015) Phonetic and visual cues to questionhood in French conversation. *Phonetica* 72, 20–42
31. Englert, C. (2010) Questions and responses in Dutch conversations. *J. Pragmat.* 42, 2666–2684
32. Stivers, T. (2010) An overview of the question–response system in American English conversation. *J. Pragmat.* 42, 2772–2781
33. Hayashi, M. (2010) An overview of the question–response system in Japanese. *J. Pragmat.* 42, 2685–2702
34. Dachkovsky, S. and Sandler, W. (2009) Visual intonation in the prosody of a sign language. *Lang. Speech* 52, 287–314
35. de Vos, C. *et al.* (2009) Mixed signals: combining linguistic and affective functions of eyebrows in questions in sign language of the Netherlands. *Lang. Speech* 52, 315–339
36. Manrique, E. (2016) Other-initiated repair in Argentine Sign Language. *Open Linguist.* 2, 1–34
37. Nota, N. *et al.* (2021) Facial signals and social actions in multimodal face-to-face interaction. *Brain Sci.* 11, 1017
38. Kendrick, K.H. (2015) Other-initiated repair in English. *Open Linguist.* 1, 164–190
39. Bavelas, J. and Chovil, N. (2018) Some pragmatic functions of conversational facial gestures. *Gesture* 17, 98–127
40. Hoemke, P. *et al.* (2025) Eyebrow movements as signals of communicative problems in human face-to-face interaction. *R. Soc. Open Sci.* 12, 241632
41. Srinivasan, R.J. and Massaro, D.W. (2003) Perceiving prosody from the face and voice: distinguishing statements from echoic questions in English. *Lang. Speech* 46, 1–22
42. Domaneschi, F. *et al.* (2017) Facial expressions and speech acts: experimental evidences on the role of the upper face as an illocutionary force indicating device in language comprehension. *Cogn. Process.* 18, 285–306
43. Borràs-Comes, J. and Prieto, P. (2011) 'Seeing tunes.' The role of visual gestures in tune interpretation. *Lab. Phonol.* 2, 355–380
44. Crespo Sendra, V. *et al.* (2013) Perceiving incredulity: the role of intonation and facial gestures. *J. Pragmat.* 47, 1–13
45. Miranda, L. *et al.* (2021) The role of the auditory and visual modalities in the perceptual identification of Brazilian Portuguese statements and echo questions. *Lang. Speech* 64, 3–23
46. Li, X. (2021) Multimodal practices for negative assessments as delicate matters: Incomplete syntax, facial expressions, and head movements. *Open Linguist.* 7, 549–568
47. Kendrick, K.H. and Holler, J. (2017) Gaze direction signals response preference in conversation. *Res. Lang. Soc. Interact.* 50, 12–32
48. Pekarek Doehler, S. *et al.* (2021) Multimodal assemblies for prefacing a dispreferred response: a cross-linguistic analysis. *Front. Psychol.* 12, 689275
49. Benitez-Quiroz, C.F. *et al.* (2016) The not face: a grammaticalization of facial expressions of emotion. *Cognition* 150, 77–84
50. Nota, N. *et al.* (2023) Conversational eyebrow frowns facilitate question identification: an online study using virtual avatars. *Cogn. Sci.* 47, e13392
51. Nota, N. *et al.* (2023) Facilitating question identification through natural intensity eyebrow movements in virtual avatars. *Sci. Rep.* 13, 21295
52. Trujillo, J.P. and Holler, J. (2024) Conversational facial signals combine into compositional meanings that change the interpretation of speaker intentions. *Sci. Rep.* 14, 2286
53. Nusseck, M. *et al.* (2008) The contribution of different facial regions to the recognition of conversational expressions. *J. Vis.* 8, 1
54. Nölle, J. *et al.* (2021) Facial expressions of emotion include iconic signals of rejection and acceptance. *J. Vis.* 21, 2932
55. Wu, Y. *et al.* (2024) Facial expressions of apology comprise complex social signals. *J. Vis.* 24, 868
56. Kavanagh, E. *et al.* (2022) Revisiting Darwin's comparisons between human and non-human primate facial signals. *Evol. Hum. Sci.* 4, e27
57. Kavanagh, E. *et al.* (2024) Being facially expressive is socially advantageous. *Sci. Rep.* 14, 12798
58. Whitehouse, J. *et al.* (2024) Facial expressivity in dominant macaques is linked to group cohesion. *Proc. R. Soc. B Biol. Sci.* 291, 20240984
59. Drijvers, L. and Holler, J. (2023) The multimodal facilitation effect in human communication. *Psychon. Bull. Rev.* 30, 792–801
60. Hernández-Gutiérrez, D. *et al.* (2021) Situating language in a minimal social context: how seeing a picture of the speaker's face affects language comprehension. *Soc. Cogn. Affect. Neurosci.* 16, 502–511
61. Nota, N. *et al.* (2022) Conversational eyebrow frowns facilitate question identification: An online VR study. *Cogn. Sci.* 47, e13392
62. Kaukomaa, T. *et al.* (2014) Foreshadowing a problem: turn-opening frowns in conversation. *J. Pragmat.* 71, 132–147
63. Kaukomaa, T. *et al.* (2013) Turn-opening smiles: facial expression constructing emotional transition in conversation. *J. Pragmat.* 55, 21–42
64. ter Bekke, M. *et al.* (2024) Hand gestures have predictive potential during conversation: an investigation of the timing of gestures in relation to speech. *Cogn. Sci.* 48, e13407
65. Mondada, L. (2021) How early can embodied responses be? Issues in time and sequentiality. *Discourse Process.* 58, 397–418
66. Deppermann, A. and Schmidt, A. (2021) Micro-sequential coordination in early responses. *Discourse Process.* 58, 372–396
67. Ter Bekke, M. *et al.* Co-speech hand gestures are used to predict upcoming meaning. Psychol. Sci. (in press)
68. Emmendorfer, A.K. and Holler, J. (2025) Facial signals shape predictions about the nature of upcoming conversational responses. *Sci. Rep.* 15, 1381
69. Chovil, N. (1991) Discourse-oriented facial displays in conversation. *Res. Lang. Soc. Interact.* 25, 163–194
70. Goodwin, M.H. and Goodwin, C. (1986) Gesture and coparticipation in the activity of searching for a word. 62, 51–76
71. Bavelas, J. *et al.* (2014) Including facial gestures in gesture-speech ensembles. In *From Gesture in Conversation to Visible Action as Utterance: Essays in Honor of Adam Kendon* (Seyfeddinipur, M. and Gullberg, M., eds), pp. 15–34, John Benjamins Publishing
72. 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* (Posner, R. and Müller, C., eds), pp. 234–256, Weidler
73. Rossano, F. and Liebal, K. (2014) "Requests" and "offers" in orangutans and human infants. In *Requesting in Social Interaction* (Drew, P. and Couper-Kuhlen, E., eds), pp. 335–364, Benjamins
74. Trujillo, J.P. and Holler, J. (2021) The kinematics of social action: visual signals provide cues for what interlocutors do in conversation. *Brain Sci.* 11, 996
75. Floyd, S. *et al.* (2016) Timing of visual bodily behavior in repair sequences: evidence from three languages. *Discourse Process.* 53, 175–204
76. Couper-Kuhlen, E. (2011) Pragmatics and prosody: prosody as social action. In *Foundations of Pragmatics* (Bublitz, W. and Norrick, N.R., eds), pp. 491–510, De Gruyter Mouton
77. Lücking, A. and Ginzburg, J. (2023) Leading voices: dialogue semantics, cognitive science and the polyphonic structure of multimodal interaction. *Lang. Cogn.* 15, 148–172
78. Mondada, L. (2024) Requesting in shop encounters: multimodal gestalts and their interactional and institutional accountability. In *New Perspectives in Interactional Linguistic Research* (Selting, M. and Barth-Weingarten, D., eds), pp. 278–309, John Benjamins Publishing
79. Trujillo, J.P. and Holler, J. (2023) Interactionally embedded Gestalt principles of multimodal human communication. *Perspect. Psychol. Sci.* 18, 1136–1159

of mutual understanding) but counterproductive in others (e.g., blurring boundaries between real and virtual agents)?

80. Rabovsky, M. *et al.* (2018) Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nat. Hum. Behav.* 2, 693–705
81. Jack, R.E. *et al.* (2014) Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Curr. Biol.* 24, 187–192
82. Kahneman, D. (2011) *Thinking, Fast and Slow*, McMillan
83. LeDoux, J.E. (1996) *The Emotional Brain: the Mysterious Underpinnings of Emotional Life*, Simon and Schuster
84. Clark, H.H. (1996) *Using Language*, Cambridge University Press
85. Ferreira, V.S. (2019) A mechanistic framework for explaining audience design in language production. *Annu. Rev. Psychol.* 70, 29–51
86. Pickering, M.J. and Garrod, S. (2004) Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* 27, 169–190
87. Galati, A. and Brennan, S.E. (2010) Attenuating information in spoken communication: for the speaker, or for the addressee? *J. Mem. Lang.* 62, 35–51
88. Rubio-Fernandez, P. *et al.* (2024) Tracking minds in communication. *Trends Cogn. Sci.* 0,
89. Ferreira, F. *et al.* (2002) Good-enough representations in language comprehension. *Curr. Dir. Psychol. Sci.* 11, 11–15
90. Quintana, D.S. and Guastella, A.J. (2020) An allostatic theory of oxytocin. *Trends Cogn. Sci.* 24, 515–528
91. Waller, B.M. *et al.* (2016) Macaques can predict social outcomes from facial expressions. *Anim. Cogn.* 19, 1031–1036
92. Crivelli, C. and Fridlund, A.J. (2018) Facial displays are tools for social influence. *Trends Cogn. Sci.* 22, 388–399
93. Fridlund, A.J. (1994) *Human Facial Expression: an Evolutionary View*, Academic Press
94. Schultz, J. and Frith, C.D. (2022) Animacy and the prediction of behaviour. *Neurosci. Biobehav. Rev.* 140, 104766
95. Hellbernd, N. and Sammler, D. (2018) Neural bases of social communicative intentions in speech. *Soc. Cogn. Affect. Neurosci.* 13, 604–615
96. Tomasello, R. *et al.* (2022) Instantaneous neural processing of communicative functions conveyed by speech prosody. *Cereb. Cortex* 32, 4885–4901
97. Ogden, R. (2006) Phonetics and social action in agreements and disagreements. *J. Pragmat.* 38, 1752–1775
98. Sicoli, M.A. *et al.* (2015) Marked initial pitch in questions signals marked communicative function. *Lang. Speech* 58, 204–223
99. Savino, M. and Grice, M. (2011) The perception of negative bias in Bari Italian questions. In *Prosodic Categories: Production, Perception and Comprehension* (Frota, S. *et al.*, eds), pp. 187–206, Springer Netherlands
100. Couper-Kuhlen, E. (2012) Some truths and untruths about final intonation in conversational questions. In *Questions* (de Ruiter, J.P., ed.), pp. 123–145, Cambridge University Press
101. Hirschberg, J. (2000) A corpus-based approach to the study of speaking style. In *Prosody: Theory and Experiment: Studies Presented to Gösta Bruce* (Horne, M., ed.), pp. 335–350, Springer Netherlands
102. Geluykens, R. (1988) On the myth of rising intonation in polar questions. *J. Pragmat.* 12, 467–485
103. Rossano, F. (2010) Questioning and responding in Italian. *J. Pragmat.* 42, 2756–2771
104. Beun, R.-J. (1990) The recognition of Dutch declarative questions. *J. Pragmat.* 14, 39–56
105. Darwin, C. (1872) *The Expression of the Emotions in Man and Animals*, Philosophical Library
106. Ekman, P. (1971) Universals and cultural differences in facial expressions of emotion. *Neb. Symp. Motiv.* 19, 207–283
107. Cowen, A.S. *et al.* (2021) Sixteen facial expressions occur in similar contexts worldwide. *Nature* 589, 251–257
108. Jack, R.E. *et al.* (2012) Facial expressions of emotion are not culturally universal. *Proc. Natl. Acad. Sci.* 109, 7241–7244
109. Andrew, R.J. (1963) Evolution of facial expression. *Science* 142, 1034–1041
110. Vick, S.-J. *et al.* (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
111. Godinho, R.M. *et al.* (2018) Supraorbital morphology and social dynamics in human evolution. *Nat. Ecol. Evol.* 2, 956–961
112. Scott, L. and Florkiewicz, B.N. (2023) Feline faces: unraveling the social function of domestic cat facial signals. *Behav. Process.* 213, 104959
113. Aguert, M. (2022) Paraverbal expression of verbal irony: vocal cues matter and facial cues even more. *J. Nonverbal Behav.* 46, 45–70
114. Tabacaru, S. and Lemmens, M. (2014) Raised eyebrows as gestural triggers in humour: The case of sarcasm and hyper-understanding. *Eur. J. Humour Res.* 2, 11–31
115. Ruusuvuori, J. and Peräkylä, A. (2009) Facial and verbal expressions in assessing stories and topics. *Res. Lang. Soc. Interact.* 42, 377–394
116. Clift, R. and Rossi, G. (2023) Speaker eyebrow raises in the transition space: pursuing a shared understanding. *Soc. Interact. Video-Based Stud. Hum. Soc.* 6, https://doi.org/10.7146/si.v6i3.142897
117. Ambrazaitis, G. and House, D. (2017) Multimodal prominences: exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings. *Speech Comm.* 95, 100–113
118. Swerts, M. and Krahmer, E. (2010) Visual prosody of newsreaders: effects of information structure, emotional content and intended audience on facial expressions. *J. Phon.* 38, 197–206
119. Clark, H.H. (2016) Depicting as a method of communication. *Psychol. Rev.* 123, 324–347
120. Sidnell, J. (2006) Coordinating gesture, talk, and gaze in reenactments. *Res. Lang. Soc. Interact.* 39, 377–409
121. Bavelas, J. *et al.* (2014) Effect of dialogue on demonstrations: direct quotations, facial portrayals, hand gestures, and figurative references. *Discourse Process.* 51, 619–655
122. Andries, F. *et al.* (2023) Multimodal stance-taking in interaction – a systematic literature review. *Front. Commun.* 8. https://doi.org/10.3389/fcomm.2023.1187977
123. Duncan, S. (1974) On the structure of speaker–auditor interaction during speaking turns. *Lang. Soc.* 3, 161–180
124. Brunner, L.J. (1979) Smiles can be back channels. *J. Pers. Soc. Psychol.* 37, 728–734
125. Levinson, S.C. (2015) Other-initiated repair in Yélî Dnye: seeing eye-to-eye in the language of Rossel Island. *Open Linguist.* 1, 386–410
126. Hömke, P. *et al.* (2017) Eye blinking as addressee feedback in face-to-face conversation. *Res. Lang. Soc. Interact.* 50, 54–70
127. Dingemanse, M. (2015) Other-initiated repair in Siwu. *Open Linguist.* 1, 232–255
128. Dix, C. and Groß, A. (2023) Surprise about news or just receiving information? Moving and holding both eyebrows in co-present interaction. *Soc. Interact. Video-Based Stud. Hum. Soc.* 6, 2446–3620
129. Echterhoff, G. and Higgins, E.T. (2017) Creating shared reality in interpersonal and intergroup communication: the role of epistemic processes and their interplay. *Eur. Rev. Soc. Psychol.* 28, 175–226
130. Bavelas, J.B. *et al.* (2000) Listeners as co-narrators. *J. Pers. Soc. Psychol.* 79, 941–952