A review of theories and methods in the science of face-to-face social interaction

Lauren V Hadley[1†], Graham Naylor[1], & Antonia F de C Hamilton[2]

[1] Hearing Sciences - Scottish Section, School of Medicine, University of Nottingham, Glasgow, UK.

[2] Institute of Cognitive Neuroscience, Division of Psychology and Language Sciences, University College London, London, UK.

[†]e-mail: lauren.hadley@cantab.net

Acknowledgements

L.V.H is supported by a UKRI Future Leaders Fellowship (MR/T041471/1). G.N. is supported by the Medical Research Council (MR/S003576/1); and the Chief Scientist Office of the Scottish Government. A.H. was supported by the Leverhulme Trust (RPG-20160-251).

Author contributions

L.V.H and A.H. contributed substantially to discussion of article content and wrote the article. All authors reviewed and/or edited the manuscript before submission.

Competing interests

The authors declare no competing interests.

Peer review information **[Au: This information will be added prior to publication. Please do not remove.]**

Nature Reviews XXX thanks [Referee#1 name], [Referee#2 name] and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Abstract

For most of human history, face-to-face interactions have been the primary and most fundamental way to build social relationships, and even in the digital era they remain the basis of our closest bonds. These interactions are built on the dynamic integration and coordination of verbal and nonverbal information between multiple people. However, the psychological processes underlying face-to-face interactions remain difficult to study. In this Review, we discuss three ways the multimodal phenomena underlying face-to-face social interactions can be organized to provide a solid basis for theory development. Next, we review three types of theories of social interaction: theories that focus on the social meaning of actions, theories that explain actions in terms of simple behaviour rules, and theories that rely on rich cognitive models of the internal states of others. Finally, we address how different methods can be used to distinguish between theories, showcasing new approaches and outlining important directions for future research. Advances in how face-to-face social interaction can be studied, combined with a renewed focus on cognitive theories, could lead to a renaissance in social interaction research and advance scientific understanding of face-to-face interaction and its underlying cognitive foundations.

[H1] Introduction

Our first and most important interactions are face to face. Examples include the playful interactions of infants and their parents, the exuberant games of groups of children, the exhilarating performances of musicians in bands, and the complex discussions between rival politicians. In each of these examples, two or more people act and interact across multiple modalities in a tightly timed and coordinated fashion to advance their social relationship or even change the world[1]. Understanding the psychology of face-to-face interactions—how they work, the factors that influence them, and the cognitive and brain mechanisms involved—remains a substantial challenge to researchers due to the complexity and interdependence of the behaviours involved.

The potential scope of social interaction research is huge, ranging from interactions between a few people[2,3] to groups of hundreds[4]; and from the detailed study of a few minutes of behaviour[5] to the long-term tracking of relationships or social networks[6]. Furthermore, interactions can be affiliative or transactional; casual or formal[7,8]; or collaborative or competitive[9]; and an individual's relationships[10,11], status[12,13], and goals[14] also impact their social behaviour. Many research traditions have a narrow focus on particular aspects of social interaction, with social psychology, cognitive neuroscience, linguistics, computing, and animal behaviour each taking a different perspective. This creates a wealth of diverse research that cannot easily be integrated. Drawing these different traditions together might yield new ways of thinking and important insights.

There are several reasons that studying social interaction is both important and timely. First, there is growing recognition that findings from lab studies of how one individual responds to one constrained form of a social behaviour (often in a single modality and without context) do not necessarily generalise to the messy, dynamic, multimodal behaviours seen in real-world interactions[15]. Thus, more research is needed on natural interactions, involving multiple types of stimuli across different modalities and including both the benefits of context and the challenges of ambiguity inherent during everyday interaction. Second, psychological research findings are often applied to real-world settings that involve face-to-face social situations, such as education[16], psychological therapy[17], or organizational management[18], without fully understanding how social interaction behaviour could impact the outcome. Finally, artificial agents (computer-controlled characters that might speak, move or otherwise interact with a person in a way that simulates a human partner, such as Apple's Siri or Amazon's Alexa), are increasingly being used to communicate with people. However, current systems can neither understand nor produce nonverbal behaviour. Because nonverbal behaviour can entirely change the meaning of words[19] (especially in the cases of irony or humour), its absence in communication with artificial agents could lead to misunderstandings and potentially reduce acceptance of technology[20]. A better understanding of real-world human interaction could therefore enable the design of better artificial agents that are more beneficial to end-users.

Fortunately, technical and experimental possibilities for studying real-world social interaction are expanding. Phenomena including joint action (engagement in a collaborative task)[21], synchronization[22], and audience effects (behaviour changes due to the belief that someone is watching)[23] can now be studied in multi-person situations within and beyond the lab by exploiting new developments in motion capture, machine learning, and even wearable brain imaging[24]. The time is therefore ripe to take a fresh look at the study of interactive face-to-face social behaviour[25] and to draw together the expansive but disparate literature.

In this Review, we take a cognitive approach to investigating social interaction, specifically focusing on the micro-level[26] of small groups interacting in person over short timescales (minutes to hours). Although verbal and non-verbal behaviours are closely integrated in face-to-face interaction, our focus here is primarily on the non-linguistic aspects of interaction, that is, those which would not be included in a simple text transcript. This includes the nonverbal behaviours that regulate turn-taking conversation structure (for example, the gaze and body orientation that indicate a turn-end)[27–29], that contribute to rapport (for example, the posture mirroring that is linked to a favourable impression)[30,31], and that convey information about relative power dynamics (for example, the voice features that indicate dominance)[32,33]. Although there is important work on social interactions in the context of mass communication[34], social media[35], and long-term social relationships[36], such areas are beyond the scope of this Review. We first consider the different levels at which researchers study social interaction behaviour. Next, we discuss the types of theories available to interpret current work and explore the methods available to test these theories. We end with a survey of promising approaches to move the field of social interaction forward.

[H1] Organization of social behaviour

In the study of the natural world, Linnaeus' work classifying plants and animals into different genera and species provided an essential framework which could be used in Darwin's theory of evolution. By analogy, obtaining a suitable classification schema for social behaviours could provide the foundation for new theories of social interaction. It is therefore worth considering the question of how social behaviour should be classified, especially given that different research traditions have taken different approaches.

Observable interaction behaviours can be described on many levels (Figure 1). How researchers choose to categorise interaction behaviour shapes the kinds of questions that they can ask, and the kinds of answers they receive. For example, to explore the relationship between smiling and rapport a researcher could focus on identifying mouth movements and count the smiles during an interaction. However, if the researcher does not account for the social meaning of the smiles (for example, a genuine smile versus a polite smile), they could miss the importance of contingency of interactors' behaviours: it is not the number of smiles, but the matching of smile type between interaction partners that determines rapport[37].

Here, we examine three ways social interaction behaviour could be organised. First, grouping behaviours according to the effectors by which they are implemented or modalities in which they occur enables researchers to address 'what' people do. Second, grouping behaviours according to their underlying cognitive processes enables researchers to address 'how' particular types of behaviour are generated. Third, grouping behaviours according to their social meaning enables researchers to address 'why' people use these behaviours. These are of course permeable divisions, and it is possible to analyse by social meaning then consider how those meanings were generated, or analyse by modality then investigate the cognitive processes involved. However, a researcher's initial decisions around behaviour organisation foreground particular possibilities and researchers using different organising principles might observe similar behaviours but interpret them differently.

[H3] Grouping by modality and effector

One obvious starting point for organising social interaction behaviour is in terms of the body parts that produce the behaviour (Figure 1, middle row). Given that these modes of production often map onto recording instruments, grouping by modality is the practical default in most research. For example, gaze[38] (captured with eye trackers), speech[39] (captured with microphones), and social touch[40] (captured with video) are often studied separately. Natural input modalities (visual[41], auditory[42], and somatosensory[43]) provide a similar implicit categorisation for interaction behaviour, and again it is common for labs to focus on only one.

However, the behaviours captured by a single recording instrument or through a single modality can be disparate. For example, a motion tracking system could record both head nods and postural shifts, but these are unrelated body movements likely elicited for different reasons. Instead of investigating these two very different forms of behaviour in one study, a nod might be better grouped with an ‘mm hmm’ sound (reliant on a different modality and recording device) due to its similar backchanelling function[44]. Furthermore, behaviour in one modality might have a different function depending on behaviour in other modalities. For example, an utterance can function as a statement or as an incredulous question depending on the talker’s facial expression[45]. Thus, because a lot of social behaviour is multimodal, recording only one effector or modality risks missing what can be learned from their combination. Although unimodal approaches have led to substantial progress in understanding face-to-face behaviour and are technologically convenient, multimodal approaches might deepen understanding of the fundamentals of face-to-face interaction[46].

[H3] Grouping by cognitive processes

A second means of organising behaviour is grouping according to the cognitive processes supporting the observed behaviour (Figure 1, top two rows). Although modality-specific cognitive systems can be studied via particular types of recording equipment, there are also more general cognitive systems cutting across different modalities that could be used to group and interpret social behaviours. For example, language systems[47] are important in many types of social interaction[48], and in face-to-face conversation the interplay between linguistic content and nonverbal cues such as tone of voice, gaze and facial expression is often critical to interpreting the meaning of a behaviour. Similarly, executive functions such as cognitive control[49] are important for regulating interactive behaviour, allowing people to avoid excessive imitation[50] and engage in social coordination[51].

Studies of the mirror neuron system illustrate the value of grouping behaviour according to cognitive processes. The mirror neuron system contains neurons which respond when an individual performs an action but also when they perceive that same action performed by someone else, thus providing a mapping between visual and motor action representations for self and other[52,53]. This mapping is believed to provide the basis for imitation and social learning[54] [55]. Grouping the study of behaviour in terms of ‘things that probably engage the mirror system’ allows mimicry of body postures[56] to be categorized with imitation of simple hand movements[57] and alignment of speech forms[58,59]. This grouping is consistent with the parallels between the theory of alignment[60] which developed from studies of speech forms, and the associative sequence learning theory[61] which developed from studies of hand imitation; both theories build on the general principles of matching the actions of self and other as instantiated in mirror systems. Based on this grouping, one could, for example, predict that emotional valence should influence alignment of speech forms in the same way that it influences mimicry of hand actions; a proposal that has elicited some evidence[62,63]. Thus, the mirror neuron system example illustrates how a neurocognitive theory can influence organization and interpretation of behavioural data and generate new testable hypotheses.

[H3] Grouping by social meaning

A final means of organising social interaction behaviour is to group behaviours according to their social meaning (Figure 1, bottom two rows). Attributing social meaning to behaviours implies that behaviours are meaningful, understandable signals[64]. Although some behaviours are ambiguous and people can manipulate their behaviour to deceive (for example, faking a smile), in many cases interpreting these social meaning signals is straightforward[65]. For example, if someone ostentatiously yawns, looks at their watch, or asks ‘is that the time?’ while their friend is detailing their latest work issues, the listener might in all cases be sending (and be interpreted as conveying) the message of boredom (Figure 2). Thus, a specific social meaning (‘I’m bored’) can be applied to several different behaviours. However, meanings may be flexible and change with context, such as the relationship between the interactors, the environment in which they are found, or concurrent behaviours in other modalities. For example, if someone yawns, looks at their watch, or asks ‘is that the time’ during a late night out, they might instead be sending the message that they are tired.

Grouping behaviours from different modalities that convey similar social meanings provides a way to conceptualise behaviours without making claims about their underlying cognitive systems. This approach is similar to that taken to understand communication behaviour in animals. For example, researchers might categorise the calls of vervet monkeys according to their use in the context of a snake or an eagle[66] without making claims about the cognitive mechanisms involved. This approach was dominant in many early social interaction studies[29,67], which catalogued different types of movement and assigned likely meanings to them[28,68]. For example, studies defined different facial behaviours and related them to particular emotions or intentions[69], linked different postures to interpersonal attitudes[68], or identified behavioural cues expressing power[70,71]. This approach continues to offer a valuable level of description for understanding how different behaviours relate to each other. However, it does not address how those behaviours are generated.

Given these three different schemas for classifying interaction behaviour, we can then ask how they relate to each other. A critical open question is how social meanings relate to cognitive processing. For example, if Susan wants to get John to pay attention to her, she could wave, look directly at him, or call out ‘Hey John’ (Figure 2). Although these behaviours (waving, looking, and calling out) occur in different modalities, the latter two both activate medial prefrontal cortex[72], a brain region linked to theory of mind (the ability to understand the mental states of others[73]) and to the sense of self[74]. Thus, behaviours with the same social meaning might map onto the same neurocognitive system. The field of neuropragmatics is relevant to addressing this possibility, as it investigates the neural systems involved in mapping between the words people say and what they mean by taking account of context and theory of mind[75,76]. This approach could include the neural mechanisms involved in interpreting and producing nonverbal signals, but to date there is little work in this area. If processing the social meanings of behaviours across different modalities activates the same cognitive systems, this would provide a powerful way to make sense of multimodal context-dependent interaction behaviours.

[H1] Theories of social interaction

The classifications set out in the previous section define different approaches to organising the study of social interaction, but such organisation must then relate to theories which can explain and predict patterns of behaviour. Many different theories of social interaction have been proposed in different research traditions. Here, we group them into three broad categories of explanation: Social meaning models, behaviour rules, and rich cognition theories.

[H3] Social Meaning Models

One theoretical approach in social interaction research is to focus on the function of behaviour, or ‘why’ people present the behaviours that they do. Animal research provides an example. Because the cognitive states of wild animals cannot be assessed, animal researchers tease apart the potential meanings of behaviours by considering the context in which a behaviour is produced and the way other animals respond to it[77]. For example, tracking the contexts in which different types of vervet monkey vocalisations naturally occurred revealed that ‘snake’, ‘leopard’, and ‘eagle’ were signalled with different alarm calls[78]. Finding that other monkeys responded appropriately to the communicated threat when recordings of the different types of alarm calls were played (looking down for snake, running up trees for leopard, and hiding in dense bush for eagle) confirmed the communicative function of these calls. Another study using machine learning revealed that bats have distinct vocalisations for aggression depending on the situation (for example, when squabbling over food versus resisting a mating attempt), and depending on the bat being addressed[79]. Social meaning models therefore map social behaviour functionally.

Human social behaviour has been studied in a similar way[80]. Building on work which catalogued different types of movement and assigned likely meanings to them[28,68], researchers developed methods to categorise and automatically detect signals in body posture, facial expressions and interpersonal coordination[81]. These studies fall within a framework of signalling: one person encodes a particular meaning in an action, while another person decodes the meaning of the observed signal. Thus, many studies examine either encoding, for example by instructing participants to produce actions with particular meanings[82], or decoding, for example asking participants to judge the meaning of the action in a photo or video clip[67], but not both. A consensus on the meaning of social behaviour is inferred when non-interacting observers of an interaction perceive the same social meaning as the interactors themselves[83]. However, this is not always the case. People may agree on what they perceive as the behaviours of dominant actors, but these behaviours are not always evident in real interactions[84]. This might challenge ideas about how well social meanings can be identified.

The utility of classifying behaviour by social meaning is evident in theories of gaze behaviour. Kendon[29] suggested that gaze has three dissociable functions: receiving information, regulating conversation, and controlling the intensity of the interaction. Although many studies support the use of these distinct functions[85], there is evidence that additional processes modulate gaze behaviour. For example, participants show more averted gaze when they make a response that the listener would not like compared to a response that they would like[86]. This suggests that gaze can signal more than Kendon’s model implies, and demonstrates that there is scope to refine understanding of the social meaning of gaze cues in real-world contexts.

Research on the meaning of signals is further complicated by the dependence of many perceived social meanings on the interactors’ expectations and emotions. For example, touch might be perceived as more appropriate coming from an attractive interaction partner[87], and eye contact may be perceived more positively from a strong versus weak interview applicant[88]. These findings are consistent with expectancy violations theory[89], according to which people’s communication behaviour is ambiguous and interpreted according to the observer’s positive or negative evaluation of the producer, and with discrepancy arousal theory[90], according to which the discrepancy between expected and perceived behaviours drives the observer's affective response. All these findings suggest that social meaning approaches can provide a starting point for describing and understanding social interaction but are not yet comprehensive.

[H3] Behaviour Rules

An alternative theoretical approach to social interaction starts from the idea that a series of simple behaviour rules that guide how behaviours are generated are sufficient to explain complex interactions. This approach has been used to understand group coordination in the animal behaviour literature. For example, the coordinated movement of flocks of starlings or schools of fish appears very sophisticated, but has a very simple basis: The movement of large groups of birds can be explained by combining the rules 'avoid crowding your neighbour', 'keep the same heading as your neighbour', and 'steer towards the group average'[91]; the movement of large groups of fish can be explained by the rules 'avoid those too nearby', 'align with those at an intermediate distance', and 'move towards those further away'[92]. It is striking that such simple rules at the individual level result in such apparently complex collective coordination at the group level[93]. This approach has also been applied to identify behaviour rules for how people walk in crowds[94], a situation akin to the flocking of birds.

Simple behaviour rules might be feasible explanations for social interactions in which the behaviour of one individual is closely linked to the behaviour of their partner within a relatively narrow time window. For example, people tend to mimic head movements with a delay of 600 ms. Thus, a simple rule of 'copy his head with 600 ms delay' might be enough to create naturalistic head mimicry behaviour[95]. In a slightly more complex example, the timing of turn-taking in speech could be explained if both the speaker and listener become entrained to the syllable rate of the speaker, and the listener employs an oscillator in counter-phase to the speaker so that they are ready to take their turn when the speaker finishes[96]. In fact, a study analysing the intercall intervals of marmoset pairs found significant coupling between each individual's vocalisations; calls were produced in antiphase with a period of approximately 12s, providing evidence for such a mechanism[97]. These results suggest that apparently complex interaction behaviours need not have complex bases.

Characterising human social interaction in terms of behaviour rules is appealing for several reasons. First, behaviour rules provide a very simple mechanism that need not be specific to social interaction, but could build on more general principles of sensorimotor control and motor learning. For example, simple mechanisms that link performed actions with observed actions could enable action alignment by preparing an observer's motor system to produce the same action that they see[52]. These mechanisms have been generalised in the interactive alignment model[98] which suggests that the fact that people can be primed to speak or act alike is fundamental to effective communication. For example, aligning on their use of gestures might help people build common ground and create a shared understanding[99]. Thus, simple mechanisms can have wide-ranging impacts.

Second, behaviour rules could be acquired via statistical learning[100], which is in keeping with increasing evidence for the role of learning in a wide range of social behaviours[101]. Statistical learning mechanisms could account for the origins of complex social behaviour without needing innate specifications. Claims for innateness can be hard to sustain for social behaviours which are not universal across humans or do not have a clear evolutionary purpose. By contrast, learning mechanisms are simple and highly flexible to different contexts.

Third, rules do not need to be absolute, but could be implemented in a probabilistic fashion. Probabilistic rule implementation allows for more flexible behaviours without requiring abstract representations of the interaction or partner's state. For example, Communication Accommodation Theory[102] suggests that the 'rule' for aligning with a partner is modified according to an individual's goals or motivation, such that people converge to gain approval, but diverge to differentiate themselves from others.[103] Thus, there can be flexibility in the use of simple rules which gives them potential to account for the variety of human behaviour. Finally, behaviour rules are relatively easy to implement via artificial agents[104], making them easily testable.

However, behaviour rules are limited as a general theory of interaction because they might be too simple to account for the richness of human social behaviour. Artificial agents governed only by simple behaviour rules will at some point begin to diverge from human behaviour. Thus, a critical question is when a behaviour-rule explanation alone would fail. For example, the implementation of an oscillator may be insufficient to convincingly mimic nonverbal behaviour in conversation; real conversations sometimes have much longer pauses before a person says something that their partner does not want to hear[105], which cannot be accounted for with simple behaviour rules.

[H3] Rich Cognition Theories

A different perspective on how people generate their behaviours can be drawn from theories about communication based on rich models that, implicitly or explicitly, require theory of mind and the representation of other people's mental states. Many of these ideas originate in studies of language. In particular, Sperber and Wilson's relevance theory[106] suggests that speakers select words to tailor utterances for their partner. This process requires theory of mind because speakers must infer their partner's knowledge states and needs. Rich cognitive processing is also implicit in Clark's theory of language use[48]. This theory suggests that conversations can be understood as joint projects in which people carefully structure their interaction at a basic motor level and at several more abstract levels of shared understanding. Clark's theory therefore covers a broad array of linguistic and extra-linguistic communication behaviours and describes processing at several different levels (including actual speech sounds, speech content, and meta-collaboration). Rich cognition models assume that people use high-level representations of their partner's mental state and knowledge of their partner to communicate socially, and implicate sophisticated cognitive systems in nonverbal communication. Note that these models do not imply that language itself must be invoked to explain nonverbal communication. Rather, the core ability to consider the state of another person's mind to communicate with that person is the basis of both nonverbal face-to-face interaction behaviours and linguistic communication.

The importance of rich theory-of-mind processes for nonverbal behaviour is evident in actions that have referential meaning, such as pointing. Infants learn to both produce[107] and understand[108] pointing actions between 12 and 18 months. There is debate over why infants point[109] but most accounts agree that their pointing is not just a behaviour rule or a response to a particular cue. Infant pointing in real-world contexts often seems to be about communicating to another person[110] or asking a question[111]. For example, infants point more in the presence of a knowledgeable adult than an ignorant adult, suggesting that infants are sensitive to what adults know and whether an adult can answer their question. This finding implicates referential communication abilities in this nonverbal behaviour. Social coordination provides another example: if children have the opportunity to coordinate behaviour with an adult to win a prize, they do so more when the adult engages in nonverbal communications such as eye contact and smiles[112]. This finding implies that nonverbal signals are sufficient to kick-start social coordination.

In studies of adult behaviour, the clearest evidence for rich cognitive theories is in cases where there is audience design, that is, when a behaviour is modulated to suit the person receiving the communicative signal. Research on audience design examines how social behaviour varies according to context[113]. For example, nonverbal behaviours such as gestures vary according to background noise level[114], such that people produce more gestures in noisy conditions where words are hard to hear, implying that gesture production is modulated to increase the efficacy of communication. In addition to adjusting to the environment, people adjust their behaviour according to their partner’s capabilities. For example, when talking to someone with hearing loss, people adjust their speech volume and the relative levels of different frequencies according to the profile of the listener’s hearing loss[115]. In each of these cases, adjustments are used to tailor behaviour to improve the partner’s understanding, implying the engagement of perspective-taking processes.

The same logic applies to studies that test how being observed influences social behaviour. There is increasing evidence that nonverbal behaviours such as mimicry[116] and smiles[117] are produced more often when participants are being watched by another person, that is, when participants have the potential to communicate. This suggests that these behaviours are not only driven by simple response rules (copy her action or smile when feeling positive) but are modulated by an understanding of what the observers can and will perceive. Producing a social behaviour for another person to perceive indicates that the sender is considering the communicative relevance of their action, which requires rich and sophisticated cognitive processes. However, the involvement of theory of mind in controlling simple nonverbal behaviours has not been comprehensively tested.

Finally, studying social interaction behaviour in populations with known cognitive difficulties can be used to inform rich cognition theories. For example, many autistic people show reduced understanding of other people’s mental states compared to age and IQ matched controls on classic false-belief tests of theory of mind[118]. Autistic participants also give less efficient descriptions of potentially ambiguous objects in a referential communication task[119] compared to age and IQ matched controls, suggesting difficulties in verbal tasks requiring audience design. Autistic participants also display less interpersonal synchrony of head and body movement[120] and, according to some studies, less eye contact with a conversation partner[121] compared to neurotypical controls. Together these results suggest a link between theory of mind and audience design or social coordination, though such a link has not yet been directly tested.

[H3] Theory Summary

We have described three broad categories of theories that attempt to account for social interaction behaviour. Social meaning models broadly address the functions of a behaviour (‘why’ people produce a behaviour), which is a holistic approach that places actions in context (Figure 3, bottom panel). Both behaviour rules and rich cognition theories provide a more mechanistic explanation of ‘how’ a behaviour might arise in terms of information processing mechanisms and associated neural systems (Figure 3, top panel). These mechanistic theories are differentiated by the involvement of theory of mind processes. A pure behaviour rules theory would claim that only simple input-output rules are necessary to explain social interactions. The use of these rules might be modulated by motivation, but does not require any assessment of another person’s internal state. These rules might, for example, be implemented in mirror neuron systems and general perception-action matching systems[122] . By contrast, a pure rich-cognition theory would allow complex calculations of an interaction partner’s mental states to govern even simple actions, and recruit theory of mind brain regions.

There could be hybrid theories between these extremes, where some behaviours can be explained by rules and others cannot, or where simple rules can be modulated by richer processes in some cases. The Social Top-down Response Modulation (STORM) model of mimicry[123] is an example of the latter. According to the STORM model perceptual-motor mappings (behaviour rules) are adjusted via top-down control according to the social context and interactors’ beliefs. Separating out when these different mechanisms apply and how they can be used to understand social interaction is an important goal for future studies.

[H1] Methodological advances

Robust and targeted methods are needed to test the theories described in the previous section. Recent advances in technology for data capture[124,125], and innovations in analysis based on machine learning[126] and statistical models for multiperson data[127,128,129] together with progress in modelling artificial agents[130] make it possible to study social interaction at a vastly higher resolution than past decades. These methodological advances allow today’s researchers to explore how multimodal information is integrated, quantify subtle interaction behaviours, and test hypotheses with high experimental control. However, to move the field forward it is important that these new capabilities are used in the service of theoretical questions.

A variety of assumptions influence how researchers choose to capture and analyse social interaction behaviour. Two key considerations are how to define the aspects of social interaction behaviour that are ‘relevant’, and how to determine the aspects of social interaction that need to be replicated in the lab to ensure generalisability. The methodological choices favoured in different research traditions often reflect their response to, and prioritisation of, these issues. In terms of defining relevant behaviour, researchers interested in acts identified as salient by the interactors themselves have no need to capture the imperceptible fine-grained behaviours detectable only via motion tracking systems. Thus, video recording methods are dominant in conversation analysis work[131]. Alternatively, researchers interested in non-conscious behaviours that are predictive of specific qualities of the interaction might find it necessary to use precise motion tracking to measure movement and gestures, particularly in studies of mimicry[95,132] and action coordination[133].

In terms of generalisability, researchers need to ensure that the aspects of real-world social interaction they are interested in are retained in lab settings. Whereas some researchers consider the context of the interaction inextricable from the behaviour elicited[134,135], others study impoverished forms of interaction in the lab under the assumption that the critical elements are preserved[136,137]. For example, researchers might investigate turn-taking in the lab by having participants play an artificial game with a virtual partner under the assumption that this is a stripped-down model of turn-taking that occurs in free conversation. However, it is not clear that such assumptions hold[15,138], as situational and environmental context might be critical to the expression of the behaviour. For example, turn-taking in free conversation with a friend may be influenced by prior shared knowledge or affective signals which are missing in an artificial game. Thus, the definition of the interaction behaviour of interest, and its necessary context, are critical decisions that reflect the researcher’s assumptions and can impact the results. Next, we illustrate how these methodological choices can play out, and highlight some methods available to researchers interested in face-to-face social interaction.

[H3] Observation of behaviour

Many papers in the social interaction literature are dedicated to identifying the social meaning behind particular behaviours. These studies often rely on observing and analysing how people act in the real world[139]. For example, in their study on communicative blinking, Hömke and colleagues[140] first coded hours of video data and found that ‘long blinks’ occurred disproportionately in conjunction with a change of conversation topic. This led to the hypothesis that a long blink could be a conversational signal, conveying 'I've heard enough'. To further probe this hypothesis, the researchers created a virtual agent who asked participants a question (for example, 'what did you do at the weekend?'), then listened (and blinked) during an extended answer[141]. The key question was how participants would react when the virtual agent gave a long blink. To insert long blinks at appropriate moments, an experimenter listened to each live conversation and manually pressed a key to provide input to the virtual agent when they perceived the end of a 'conversation unit'. On some trials, those key presses were used to make the virtual agent produce a short blink (~200ms); on other trials, they were used to produce a long blink (~600ms) (the experimenter was blind to the manipulation). Participants spoke less following a long blink versus a short blink, consistent with the theory that participants interpret long blinks as a communicative signal meaning 'I've heard enough' (and that a change of topic is welcome). This clever combination of manual analysis of video recording, which led to observation-based hypothesis generation, followed by testing with strong experimental control of the behaviour of interest (using a virtual agent), was critical to determining the communicative function of blinking behaviour.

More subtle facial movements can also be analysed for social meaning when high resolution recordings are available. Chen and colleagues[142] assigned pairs of participants to the roles of 'doctor' or 'patient' and used camera-based automated facial tracking to capture their facial movements and expressions. Participants in the role of doctor believed that one of two inert creams was able to reduce thermal pain. They applied each cream followed by a pain stimulus to their patient's arm following a clearly defined protocol and with minimal verbal communication. On trials where doctors believed the cream was effective, patients reported less heat pain, and both doctors and patients showed fewer facial expressions associated with pain compared to trials where doctors believed the cream was ineffective. These findings suggest that the reduced pain expressions from doctors (together with other nonverbal cues) might have induced a placebo effect in patients. Thus, by using sophisticated face-tracking technology as part of a complex but well-defined social interaction, it was possible to explore the nonverbal communication behaviours that underlie the transmission of pain information and beliefs in a placebo between two people.

Together, these two studies show how the detailed study of movements during an interaction enable researchers to label specific social behaviours as signals that communicate specific messages ('I'm bored' or 'this will hurt') in a manner that is effectively received by the interaction partner. However, such studies do not delve into the cognitive mechanisms underlying these behaviours. To understand mechanisms, researchers need to address how behaviours are generated, and also distinguish between simple behaviour rules and richer cognitive mechanisms.

[H3] Artificial agents

Artificial agents are commonly used to investigate the value of simple rules as potential explanations of behaviour because these rules can be programmed into the agents. It is then possible to test how participants respond to agents with or without the behaviour rule. Furthermore, identifying where the simple rules implemented in artificial agents break down suggests behaviours where richer cognitive models are likely to be required. Studies using artificial agents typically have two phases: first, observation of natural behaviour enables researchers to identify a likely behaviour rule; second, an artificial agent is built to enact the rule so researchers can test how people engage in a dynamic interaction that includes this rule. Note that this is slightly different to the method used by Hömke and colleagues[140] described above: In that study the behaviour rule 'blink if bored' was implemented by an experimenter rather than being fully programmed into the agent.

Van der Steen & Keller[143] demonstrated how very simple behaviour rules can be tested using computer models. They modelled how people perform a synchronised tapping task in which participants needed to flexibly react to errors if they tapped at a different time to their partner (adapt), and prepare to coordinate their next tap (anticipate)[143]. Implementing these rules computationally in the Adaptation and Anticipation Model (ADAM), which combines reactive error correction with predictive temporal extrapolation, enabled the researchers to build a responsive virtual partner approximating human synchronisation behaviour[144]. This suggests that for a joint tapping sensorimotor synchronisation task, simple rules of adaptation and anticipation are sufficient to mimic real interaction behaviour, and that more complex cognition is not necessary.

Simple behaviour rules can also be tested by modifying artificial agents to exhibit interaction behaviour that is more (or less) similar to humans. For example, one study manipulated whether an artificial agent's nods were timed to the appropriate points of a human's speech or presented randomly during a conversation between a participant and an artificial agent[145]. Participants reported greater feelings of rapport in the former case, which demonstrates sensitivity to contingent nod timing and the importance of temporally contingent behaviour more broadly for developing rapport[145].

Overall, using artificial agents to implement particular behaviour rules shows how close these agents can get to real interactive behaviours without any deep understanding of the human partner. The success of artificial agents as communication partners in specific contexts can be taken as an argument against rich models, and in favour of minimal rules that can be combined to generate apparently sophisticated behaviour.

[H3] Manipulation of face-to-face communication

Experiments in which aspects of face-to-face communication are artificially manipulated can be used to test whether and how participants take a partner's beliefs and mental state into account during communication. Research on audience design is one important approach to examining these issues.

In an innovative experiment, Hazan and colleagues[39] manipulated a conversational interaction so that each interlocutor experienced a different type of noisy environment. In a spot-the-difference task pairs of friends were given similar pictures with 12 differences that they needed to locate by describing the pictures to each other. In some trials one of the two participants heard their partner's voice vocoded or masked by babble noise. Different vocal adjustments are needed to be heard clearly in these two conditions. Importantly, interlocutors adjusted the pitch of their speech in ways that took their partner's needs and environment into account: they increased their pitch and pitch range more in the babble condition in which these adjustments would benefit the partner than in the vocoded condition in which they would not. This suggests a role for theory of mind or perspective taking in the communicative interaction: participants inferred what their partner was experiencing and adjusted behaviour accordingly. Research on speech-related adjustments based on a partner's hearing loss, cognitive capacity, or knowledge, imply similar high-level processes[146] whereby interactors adjust their vocal signal to meet the needs of their audience.

Manipulating the communicative goal in an interaction can also provide insight into the necessity of high level cognitive models. For example, in one study participants either performed a xylophone tune alone, with another participant, or with a learner watching them[147]. Motion trackers captured participants' precise hand movements. The results showed that the performer modulated the velocity of their actions according to whether or not the observing partner knew the sequence. This careful control of action kinematics according to the needs of an observer suggests the involvement of theory of mind processes. This study also illustrates how the use of precision motion tracking in well-controlled interactions can reveal nuances in people's behaviour that have implications for theories of social interaction.

[H3] Combining multiple methods

The studies reviewed above highlight how new technologies and innovative experimental designs can be used to address core theories of human face-to-face interaction. However, a deeper understanding will arise as we bring together multiple methods in conjunction with theories. Figure 4 is a representation of how these different approaches can be combined (inspired by prior representations of the scientific method[148–150]) and can build on one another to advance the study of human social interaction.

For example, as described above Hömke et al.[140] used observations to develop a theory for why long blinks occur in conversation (Figure 4, arrow A), and proposed that they are used as a specific communicative signal. By manipulating the communicative partner's blinks using a virtual agent[141], they then progressed to hypothesis-testing of that theory (Figure 4, arrow B). A valuable future step in this research could be to formulate an artificial agent able to detect communicative blinks and adjust their own behaviour (Figure 4, arrow C), to identify how closely this would approximate human interaction. By contrast, Keller and colleagues[143] moved straight to developing the ADAM computational model of how different adjustment processes interrelate to support interpersonal synchrony based on observations of human behaviour (Figure 4, arrows A and C). The computational model has since been used to address the importance of a human partner's goal via model-driven experimentation (Figure 4, arrow D)[151], and extended to include theories about the role of different neural regions on the basis of results from patient studies (Figure 4, arrow E)[152].

By considering the different ways that behaviours of interest can be studied, and using a variety of methods in combination, researchers can map a path between theorising about the social meaning of a behaviour and understanding its underlying cognitive basis. This could involve starting with a rich cognition theory and then testing hypotheses about how the context of the interaction affects people's behaviour, or starting with a simple rule computer model and then analysing its impact on the perceived social meaning of the rule implemented. Starting from computational models, and moving through model-driven experimentation to theory reformulation could provide a framework for differentiating behaviour rules and rich cognition models.

[H1] Summary and future directions

In this Review, we described how the study of social interaction behaviour can be organized. Next, we outlined three broad types of theory that focus on social meaning, behaviour rules, and rich cognition. Each of these approaches derive from different research traditions and emphasise different facets of behaviour; they also relate to different levels of description. Theories based on social meaning primarily focus on ‘why’ people exhibit behaviour, while behaviour rules and rich cognition approaches consider ‘how’ the underlying mechanisms support that behaviour. The way that these levels of description map onto each other, and whether social meanings relate to specific neurocognitive systems or processes, remain open questions that could be addressed using the new technologies and methodological approaches described above.

Another question is how the theories we laid out to describe face-to-face interaction extend to non-face-to-face interaction, given social activities are increasingly being conducted remotely. Fundamental social interaction skills are based on the face-to-face social experiences people have in infancy and childhood. Moreover, technology-mediated communication still requires processes like turn-taking, rapport building and information sharing. Thus, it is likely that the same cognitive mechanisms are involved in both live and technology-mediated communication. However, it will be interesting to quantify exactly how behaviour changes[153] or stays the same[154] in online versus face-to-face communication. In particular, manipulating the technology used for communication could allow researchers to disentangle whether specific behavioural adjustments are made for the benefit of the communication producer or receiver (for example, whether the speaker adjusts their voice according to their own environment, or that of their partner[39]).

To move the field forward, research on social interaction must expand in at least three ways. First, a concerted effort should be taken to distinguish between different theories of face-to-face interaction behaviour, using robust methods combined with new experimental designs. This differentiation between potential theories could involve exploring how far behaviour rules can go in accounting for interaction behaviour by looking to current animal models[155] or implementing potential rules for particular facets of an interaction in artificial agents. Another way to differentiate between theories would be to test which specific contexts and manipulations require a rich understanding of other people’s internal states for successful task performance, that is, when rich cognition is required. For example, it may be that certain behaviours (such as mimicry) can be modelled using simple behaviour rules in some situations, but require rich cognition in others (such as when the participant has an explicit affiliative goal). Audience design studies are an excellent starting point here, but other manipulations of context could allow researchers to identify whether behaviour rules or rich cognition dominate in nonverbal interactions.

Second, the basic work of describing behaviour that will enable theories to be built and tested is far from complete. In particular, it is important to continue cataloguing social interaction behaviours in different contexts and participant groups. Many older studies relied on small samples from WEIRD (Western, educated, industrialized, rich and democratic) populations[156]. Understanding which social interaction behaviours vary across cultures, development, and clinical conditions will shed light on how social behaviour is learnt, and may also elucidate neurocognitive processes underlying these behaviours (see Box 1).

Finally, theories of social interaction should not be seen as a unique domain, detached from the rest of cognitive processing. Most (or all) interaction behaviours rely on perceptual, motor, motivational and cognitive systems, in conjunction with language and memory, all of which have been extensively studied. Investigating how mechanisms of social interaction integrate with general motor and cognitive theories will enable researchers to build on existing models of brain function and cognition. Furthermore, explicitly comparing verbal and non-verbal aspects of social interaction will enable researchers to identify whether and how these forms of communication differ, and, perhaps more importantly, how they interact.

We believe that the next decade will be an exciting time for research on face-to-face interaction. New technologies and methods are enabling more detailed behavioural research, and there is increasing recognition that understanding real-world social behaviour is critical for applying psychological findings to important real-world settings and for developing the next generation of artificial agents. By building on current theories and exploring cutting-edge research methods, a new generation of researchers will be poised to uncover the fundamental cognitive architecture of the interactions that make us human.

References

1. Sparks, A. *Tomorrow is Another Country: the inside story of South Africa's road to change*. (University of Chicago Press, 1996).
2. Argyle, M. *Social interaction: process and products.* (Routledge, 2017).
3. Hartley, P. *Group communication.* (Routledge, 2006).
4. Whittaker, S., Terveen, L., Hill, W. & Cherny, L. The Dynamics of Mass Interaction. in *From Usenet to CoWebs. Computer Supported Cooperative Work* (eds. Lueg, C. & Fisher, D.) 79–91 (Springer, 2003).
5. Mondada, L. The multimodal interactional organization of tasting: Practices of tasting cheese in gourmet shops. *Discourse Stud.* **20**, 743–769 (2018).
6. Stepanyan, K., Borau, K. & Ullrich, C. A social network analysis perspective on student interaction within the twitter microblogging environment. in *10th IEEE international conference on advanced learning technologies* 70–72 (2010).
7. Eggins, S. & Slade, D. *Analysing casual conversation.* (Equinox Publishing Ltd, 2004).
8. Heydon, G. *The language of police interviewing.* (Palgrave Macmillan, 2005).
9. Tschacher, W., Rees, G. & Ramseyer, F. Nonverbal synchrony and affect in dyadic interactions. *Front. Psychol.* **5**, 1323 (2014).
10. Remland, M. S. *Nonverbal communication in everyday life.* (Sage Publications, 2016).
11. Guerrero, L. K. & Floyd, K. *Nonverbal communication in close relationships.* (Routledge, 2006).
12. Keltner, D., Gruenfeld, D. H. & Anderson, C. Power, approach, and inhibition. *Psychol. Rev.* **110**, 265–284 (2003).
13. Hall, J. A. Nonverbal behavior, status, and gender: How do we understand their relations? *Psychol. Women Q.* **30**, 384–391 (2006).
14. Thibaut, J. W. & Kelley, H. H. *The social psychology of groups.* (Routledge, 2017).
15. Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A. & Poeppel, D. Neuroscience needs behavior: correcting a reductionist bias. *Neuron* **93**, 480–490 (2017).
16. Diamond, A. & Lee, K. Interventions shown to aid executive function development in children 4 to 12 years old. *Science.* **333**, 959–964 (2011).
17. Ramseyer, F. & Tschacher, W. Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome. *J. Consult. Clin. Psychol.*

79, 284–295 (2011).

18. Cuddy, A. *Presence: Bringing your boldest self to your biggest challenges*. (Hachette UK, 2015).
19. Friedman, H. S. The modification of word meaning by nonverbal cues. in *Nonverbal Communication Today* 57–68 (De Gruyter Mouton, 1982).
20. Ritschel, H., Aslan, I., Sedlbauer, D. & André, E. Irony man: augmenting a social robot with the ability to use irony in multimodal communication with humans. in *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems* 86–94 (2019).
21. Sebanz, N. & Knoblich, G. Progress in Joint-Action Research. *Curr. Dir. Psychol. Sci.* **30**, 138–143 (2021).
22. Konvalinka, I., Vuust, P., Roepstorff, A. & Frith, C. Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Q. J. Exp. Psychol.* **63**, 2220–2230 (2010).
23. Hamilton, A. F. de C. & Lind, F. Audience effects: what can they tell us about social neuroscience, theory of mind and autism? *Cult. Brain* **4**, 159–177 (2016).
24. Pinti, P. *et al.* The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Ann. N. Y. Acad. Sci.* **40**, 1–25 (2018).
25. Hecht, M. A. & Ambady, N. Nonverbal communication and psychology: Past and future. *Atl. J. Commun.* **7**, 156–170 (1999).
26. Babbie, E. R. *The practice of social research*. (Cengage learning, 2020).
27. Argyle, M. & Kendon, A. The experimental analysis of social performance. in *Advances in experimental social psychology* (ed. Berkowitz, L.) 55–98 (Academic Press, 1967).
28. Duncan, S. Some signals and rules for taking speaking turns in conversations. *J. Pers. Soc. Psychol.* **23**, 283–292 (1972).
29. Kendon, A. Some functions of gaze-direction in social interaction. *Acta Psychol. (Amst).* **26**, 22–63 (1967).
30. Tickle-Degnen, L. & Rosenthal, R. The nature of rapport and its nonverbal correlates. *Psychol. Inq.* **1**, 285–293 (1990).
31. LaFrance, M. & Broadbent, M. Group rapport: Posture sharing as a nonverbal indicator. *Gr. Organ. Stud.* **1**, 328–333 (1976).
32. Witkower, Z., Tracy, J. L., Cheng, J. T. & Henrich, J. Two signals of social rank: Prestige and dominance are associated with distinct nonverbal displays. *J. Pers. Soc.*

Psychol. **118**, 89–120 (2020).

33. Carney, D. R. The nonverbal expression of power, status, and dominance. *Curr. Opin. Psychol.* **33**, 256–264 (2020).

34. Sanborn, F. W. & Harris, R. J. *A cognitive psychology of mass communication.* (Routledge, 2019).

35. Meikle, G. *Social media: Communication, sharing and visibility.* (Routledge, 2016).

36. Fitzpatrick, M. A. & Noller, P. Marital communication in the eighties. *J. Marriage Fam.* **52**, 832–843 (1990).

37. Heerey, E. A. Decoding the dyad: Challenges in the study of individual differences in social behavior. *Curr. Dir. Psychol. Sci.* **24**, 285–291 (2015).

38. Hirvenkari, L. *et al.* Influence of Turn-Taking in a Two-Person Conversation on the Gaze of a Viewer. *PLoS One* **8**, e71569 (2013).

39. Hazan, V. & Baker, R. Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *J. Acoust. Soc. Am.* **130**, 2139–2152 (2011).

This innovative study manipulates the acoustic environment of talker and listener separately, demonstrating the importance of communicative intent on speech adjustments.

40. Lee Masson, H. & Op de Beeck, H. Socio-affective touch expression database. *PLoS One* **13**, e0190921 (2018).

41. Tsao, D. Y. & Livingstone, M. S. Mechanisms of face perception. *Annu. Rev. Neurosci.* **31**, 411–437 (2008).

42. Sauter, D. A., Eisner, F., Ekman, P. & Scott, S. K. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proc. Natl. Acad. Sci.* **107**, 2408–2412 (2010).

43. Cascio, C. J., Moore, D. & McGlone, F. Social touch and human development. *Dev. Cogn. Neurosci.* **35**, 5–11 (2019).

44. Poppe, R., Truong, K. P. & Heylen, D. Backchannels: Quantity, type and timing matters. in *International Workshop on Intelligent Virtual Agents* 228–239 (2011).

45. Kessous, L., Castellano, G. & Caridakis, G. Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. *J. Multimodal User Interfaces* **3**, 33–48 (2010).

46. Patterson, M. L. *Nonverbal behavior: A functional perspective*. (Springer Science & Business Media, 2012).

47. Friederici, A. D. The brain basis of language processing: from structure to function. *Physiol. Rev.* **91**, 1357–1392 (2011).

48. Clark, H. *Using language*. (Cambridge University Press, 1996).

49. Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. & Cohen, J. D. Conflict monitoring and cognitive control. *Psychol. Rev.* **108**, 624–652 (2001).

50. Darda, K. M. & Ramsey, R. The inhibition of automatic imitation: A meta-analysis and synthesis of fMRI studies. *Neuroimage* **197**, 320–329 (2019).

51. Volman, I., Roelofs, K., Koch, S., Verhagen, L. & Toni, I. Anterior prefrontal cortex inhibition impairs control over social emotional actions. *Curr. Biol.* **21**, 1766–1770 (2011).

52. Rizzolatti, G. & Craighero, L. The mirror-neuron system. *Annu. Rev.iew Neurosci.* **27**, 169–192 (2004).

53. Hamilton, A. F. D. C. The social function of the human mirror system: a motor chauvinist view. in *Shared representations: Sensorimotor foundations of social life* (eds. Obhi, S. S. & Cross, E. S.) (Cambridge University Press, 2015).

54. Rizzolatti, G. & Sinigaglia, C. The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat Rev Neurosci* **11**, 264–274 (2010).

55. Caspers, S., Zilles, K., Laird, A. R. & Eickhoff, S. B. *ALE meta-analysis of action observation and imitation in the human brain. NeuroImage* **50**, (Elsevier B.V., 2010).

56. Chartrand, T. L. & Bargh, J. A. The chameleon effect: the perception–behavior link and social interaction. *J. Pers. Soc. Psychol.* **76**, 893–910 (1999).

57. Heyes, C., Bird, G., Johnson, H. & Haggard, P. Experience modulates automatic imitation. *Cogn. Brain Res.* **22**, 233–240 (2005).

58. Adank, P., Nuttall, H., Bekkering, H. & Maegherman, G. Effects of stimulus response compatibility on covert imitation of vowels. *Attention, Perception, Psychophys.* **80**, 1290–1299 (2018).

59. Polyanskaya, L., Samuel, A. G. & Ordin, M. Speech Rhythm Convergence as a Social Coalition Signal. *Evol. Psychol.* **17**, https://doi.org/10.1177/1474704919879335 (2019).

60. Menenti, L., Pickering, M. J. & Garrod, S. Toward a neural basis of interactive alignment in conversation. *Front. Hum. Neurosci.* **6**, 185 (2012).

61. Heyes, C. What's social about social learning? *J. Comp. Psychol.* **126**, 193–202 (2011).

62. Virhia, J., Kotz, S. A. & Adank, P. Emotional state dependence facilitates automatic

imitation of visual speech. *Q. J. Exp. Psychol.* **72**, 2833–2847 (2019).

63. Kuhbandner, C., Pekrun, R. & Maier, M. A. The role of positive and negative affect in the "mirroring" of other persons' actions. *Cogn. Emot.* **24**, 1182–1190 (2010).

64. Smith, J. M. & Harper, D. *Animal signals*. (Oxford University Press., 2003).

65. Csibra, G. Action mirroring and action interpretation: An alternative account. in *Sensorimotor foundations of higher cognition: Attention and performance XX* (eds. Haggard, P., Rosetti, Y. & Kawato., M.) 461–479 (Oxford University Press, 2008).

66. Seyfarth, R. M., Cheney, D. L. & Marler, P. Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Anim. Behav.* **28**, 1070–1094 (1980).

67. Burgoon, J. K. Relational message interpretations of touch, conversational distance, and posture. *J. Nonverbal Behav.* **15**, 233–259 (1991).

68. Argyle, M. *Bodily communication.* (International Universities Press, 1975).

69. Ekman, P., Sorenson, E. R. & Friesen, W. V. Pan-cultural elements in facial displays of emotion. *Science (80-.).* **164**, 86–88 (1969).

70. Henley, N. *Body politics: Power, sex, and nonverbal communication.* (Prentice Hall, 1977).

71. Henley, N. M. Body politics revisited: What do we know today. in *Gender, power, and communication in human relationships* (eds. Kalbfleisch, P. J. & Cody, M. J.) 27–61 (Psychology press, 1995).

72. Kampe, K. K., Frith, C. D. & Frith, U. "Hey John": signals conveying communicative intention toward the self activate brain regions associated with "mentalizing," regardless of modality. *J. Neurosci.* **23**, 5258–5263 (2003).

This is one of the few fMRI studies to examine social signals across different modalities, showing common activation in prefrontal cortex for different communicative signals.

73. Frith, C. & Frith, U. Theory of mind. *Curr. Biol.* **15**, R644–R645 (2005).

74. Amodio, D. M. & Frith, C. D. Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* **7**, 268–77 (2006).

75. Hagoort, P. & Levinson, S. C. Neuropragmatics. in *The cognitive neurosciences* 667–674 (MIT Press, 2014).

76. Van Berkum, J. J. The neuropragmatics of 'simple' utterance comprehension: An ERP review. in *Semantics and pragmatics: From experiment to theory* 276–316 (Palgrave Macmillan, 2009).

77. Stegmann, U. *Animal communication theory: information and influence*. (Cambridge University Press, 2013).

78. Seyfarth, R. M., Cheney, D. L. & Marler, P. Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science.* **210**, 801–803 (1980).

79. Prat, Y., Taub, M. & Yovel, Y. Everyday bat vocalizations contain information about emitter, addressee, context, and behavior. *Sci. Rep.* **6**, 1–10 (2016).

This paper shows how large datasets and machine learning approaches can help us understand the social meanings of animal communications.

80. Mehu, M. & Scherer, K. R. A psycho-ethological approach to social signal processing. *Cogn. Process.* **13**, 397–414 (2012).

81. Burgoon, J. K., Magnenat-Thalmann, N., Pantic, M. & Vinciarelli, A. *Social signal processing.* (Cambridge University Press, 2017).

82. Remland, M. S. Leadership impressions and nonverbal communication in a superior-subordinate interaction. *Commun. Q.* **32**, 41–48 (1984).

83. Burgoon, J. K. & Newton, D. A. Applying a social meaning model to relational message interpretations of conversational involvement: Comparing observer and participant perspectives. *South. J. Commun.* **56**, 96–113 (1991).

84. Hall, J. A., Coats, E. J. & LeBeau, L. S. Nonverbal behavior and the vertical dimension of social relations: a meta-analysis. *Psychol. Bull.* **131**, 898–924 (2005).

85. Hessels, R. S. How does gaze to faces support face-to-face interaction? A review and perspective. *Psychon. Bull. Rev.* **27**, 856–881 (2020).

86. Kendrick, K. H. & Holler, J. Gaze direction signals response preference in conversation. *Res. Lang. Soc. Interact.* **50**, 12–32 (2017).

87. Burgoon, J. K. & Walther, J. B. Nonverbal expectancies and the evaluative consequences of violations. *Hum. Commun. Res.* **17**, 232–265 (1990).

88. Burgoon, J. K., Coker, D. A. & Coker, R. A. Communicative effects of gaze behavior: A test of two contrasting explanations. *Hum. Commun. Res.* **12**, 495–524 (1986).

89. Burgoon, J. K. A communication model of personal space violations: Explication and an initial test. *Hum. Commun. Res.* **4**, 129–142 (1978).

90. Cappella, J. N. & Greene, J. O. A discrepancy-arousal explanation of mutual influence in expressive behavior for adult and infant-adult interaction. *Commun. Monogr.* **49**, 89–114 (1982).

91. Hildenbrandt, H., Carere, C. & Hemelrijk, C. Self-organized aerial displays of thousands of starlings: a model. *Behav. Ecol.* **21**, 1349–1359 (2010).

92. Huth, A. & Wissel, C. The simulation of the movement of fish schools. *J. Theor. Biol.*

156, 365–385 (1992).

93. Couzin, I. D. Collective cognition in animal groups. *Trends Cogn. Sci.* **13**, 36–43 (2009).

94. Moussaïd, M. *et al.* Experimental study of the behavioural mechanisms underlying self-organization in human crowds. *Proc. R. Soc. B Biol. Sci.* **276**, 2755–2762 (2009).

95. Hale, J., Ward, J. A., Buccheri, F., Oliver, D. & Hamilton, A. F. d. C. Are You on My Wavelength? Interpersonal Coordination in Dyadic Conversations. *J. Nonverbal Behav.* **44**, 63–83 (2020).

This study of head nodding behaviour in conversation identifies two different types of nods using motion capture

96. Wilson, M. & Wilson, T. An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.* **12**, 957–968 (2005).

97. Takahashi, D., Narayanan, D. & Ghazanfar, A. A. Coupled oscillator dynamics of vocal turn-taking in monkeys. *Curr. Biol.* **23**, 2162–2168 (2013).

98. Pickering, M. J. & Garrod, S. The interactive-alignment model: Developments and refinements. *Behav. Brain Sci.* **27**, 212–225 (2004).

99. Holler, J. & Wilkin, K. Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *J. Nonverbal Behav.* **35**, 133–153 (2011).

100. Heyes, C. Where do mirror neurons come from? *Neurosci. Biobehav. Rev.* **34**, 575–583 (2010).

101. Heyes, C. *Cognitive gadgets.* (Harvard University Press, 2018).

102. Giles, H. Accent mobility: A model and some data. *Anthropol. Linguist.* **15**, 87–109 (1973).

103. Dragojevic, M., Gasiorek, J. & Giles, H. Communication accommodation theory. in *The international encyclopedia of interpersonal communication* (eds. Berger, R. C. & Roloff, E. M.) 1–21 (Wiley Blackwell, 2015).

104. Bailenson, J. N. & Yee, N. Digital Chameleons: Automatic Assimilation of Nonverbal Gestures in Immersive Virtual Environments. *Psychol. Sci.* **16**, 814–819 (2005).

105. Pomerantz, A. Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. in *Structures of Social Action: Studies in Conversation Analysis* (eds. Atkinson, J. & Heritage, J.) 57–101 (Cambridge University Press, 1984).

106. Sperber, D. & Wilson, D. *Relevance: Communication and cognition*. (Harvard University Press, 1986).

107. Butterworth, G. & Morissette, P. Onset of pointing and the acquisition of language in infancy. *J. Reprod. Infant Psychol.* **14**, 219–231 (1996).

108. Morissette, P., Ricard, M. & Décarie, T. G. Joint visual attention and pointing in infancy: A longitudinal study of comprehension. *Br. J. Dev. Psychol.* **13**, 163–175 (1995).

109. Southgate, V., Van Maanen, C. & Csibra, G. Infant pointing: Communication to cooperate or communication to learn? *Child Dev.* **78**, 735–740 (2007).

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

111. Begus, K. & Southgate, V. Infant pointing serves an interrogative function. *Dev. Sci.* **15**, 611–617 (2012).

112. Wyman, E., Rakoczy, H. & Tomasello, M. Non-verbal communication enables children's coordination in a "Stag Hunt" game. *Eur. J. Dev. Psychol.* **10**, 597–610 (2013).

113. Clark, H. H. & Murphy, G. L. Audience design in meaning and reference. *Adv. Psychol.* **9**, 287–299 (1982).

114. Trujillo, J., Özyürek, A., Holler, J. & Drijvers, L. Evidence for a Multimodal Lombard Effect: Speakers modulate not only speech but also gesture to overcome noise. (2020).

115. Beechey, T., Buchholz, J. M. & Keidser, G. Hearing impairment increases communication effort during conversations in noise. *J. Speech, Lang. Hear. Res.* **63**, 305–320 (2020).

This study shows how talkers spontaneously modify their speech according to the acoustic environment and their partner's hearing ability.

116. Krishnan-Barman, S. & Hamilton, A. F. D. C. Adults imitate to send a social signal. *Cognition* **187**, 150–155 (2019).

117. Fridlund, A. Sociality of Solitary Smiling : Potentiation by an Implicit Audience. *J. Pers. Soc. Psychol.* **60**, 229–240 (1991).

118. Senju, A., Southgate, V., White, S. & Frith, U. Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science.* **325**, 883–885 (2009).

119. Nadig, A., Vivanti, G. & Ozonoff, S. Adaptation of object descriptions to a partner under increasing communicative demands: A comparison of children with and without autism. *Autism Res.* **2**, 334–347 (2009).

120. Georgescu, A. L. *et al.* Reduced nonverbal interpersonal synchrony in autism spectrum disorder independent of partner diagnosis: a motion energy study. *Mol. Autism* **11**, 1–

14 (2020).

121. Freeth, M. & Bugembe, P. Social partner gaze direction and conversational phase; factors affecting social attention during face-to-face conversations in autistic adults? *Autism* **23**, 503–513 (2019).

122. Cisek, P. & Kalaska, J. F. Neural mechanisms for interacting with a world full of action choices. *Annu. Rev. Neurosci.* **33**, 269–298 (2010).

123. Wang, Y. & Hamilton, A. F. de C. Social top-down response modulation (STORM): A model of the control of mimicry in social interaction. *Front. Hum. Neurosci.* **6**, 153 (2012).

124. Cornejo, C., Cuadros, Z., Morales, R. & Paredes, J. Interpersonal coordination: methods, achievements, and challenges. *Front. Psychol.* **8**, 1685 (2017).

125. Onnela, J. P., Waber, B. N., Pentland, A., Schnorf, S. & Lazer, D. Using sociometers to quantify social interaction patterns. *Sci. Rep.* **4**, 1–9 (2014).

126. Baltrušaitis, T., Robinson, P. & Morency, L. P. Openface: an open source facial behavior analysis toolkit. in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* 1–10 (2016).

127. Issartel, J., Bardainne, T., Gaillot, P. & Marin, L. The relevance of the cross-wavelet transform in the analysis of human interaction–a tutorial. *Front. Psychol.* **5**, 1566 (2015).

128. Gatica-Perez, D. Automatic nonverbal analysis of social interaction in small groups: A review. *Image Vis. Comput.* **27**, 1775–1787 (2009).

129. Richardson, M. J., Dale, R. & Marsh, K. L. Complex dynamical systems in social and personality psychology: Theory, modeling, and analysis. in *Handbook of research methods in social and personality psychology* (eds. H. T. Reis & Judd, C. M.) 253–282 (Cambridge University Press, 2014).

130. Wykowska, A., Chaminade, T. & Cheng, G. Embodied artificial agents for understanding human social cognition. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20150375 (2016).

131. Parry, R. Video-based conversation analysis. in *Sage handbook of qualitative methods in health research* 373–96 (Sage, 2010).

132. Wild, K. S., Poliakoff, E., Jerrison, A. & Gowen, E. The influence of goals on movement kinematics during imitation. *Exp. brain Res.* **204**, 353–360 (2010).

133. McEllin, L., Sebanz, N. & Knoblich, G. Identifying others' informative intentions from movement kinematics. *Cognition* **180**, 246–258 (2018).

134. Schegloff, E. A. Reflections on talk and social structure. in *Talk and social structure* (eds. Boden, D. & Zimmerman, D.) 44–70 (University of California Press, 1991).

135. Schegloff, E. A. Reflections on quantification in the study of conversation. *Res. Lang. Soc. Interact.* **26**, 99–128 (1993).

136. Kendrick, K. H. Using conversation analysis in the lab. *Res. Lang. Soc. Interact.* **50**, 1–11 (2017).

137. de Ruiter, J. P. & Albert, S. An appeal for a methodological fusion of conversation analysis and experimental psychology. *Res. Lang. Soc. Interact.* **50**, 90–107 (2017).

A helpful review that draws together ideas from the conversation analysis tradition and cognitive science.

138. Gomez-Marin, A. & Ghazanfar, A. A. The Life of Behavior. *Neuron* **104**, 25–36 (2019).

This paper explores the importance of context in studying and understanding behaviour.

139. Baxter, L. A. & Babbie, E. R. Qualitative Communication Research. in *The basics of communication research* 296–380 (Cengage Learning, 2003).

140. Hömke, P., Holler, J. & Levinson, S. C. Eye Blinking as Addressee Feedback in Face-To-Face Conversation. *Res. Lang. Soc. Interact.* **50**, 54–70 (2017).

141. Hömke, P., Holler, J. & Levinson, S. C. Eye blinks are perceived as communicative signals in human face-to-face interaction. *PLoS One* **13**, e0208030 (2018).

This study elucidates the role of blinks in conversation and communication.

142. Chen, P. H. A. *et al.* Socially transmitted placebo effects. *Nat. Hum. Behav.* **3**, 1295–1305 (2019).

This high-resolution face-tracking study demonstrates the power of detailed analyses of structured social interactions.

143. van der Steen, M. C. & Keller, P. E. The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. *Front. Hum. Neurosci.* **7**, 253 (2013).

144. van der Steen, M. C., Jacoby, N., Fairhurst, M. T. & Keller, P. E. Sensorimotor synchronization with tempo-changing auditory sequences: Modeling temporal adaptation and anticipation. *Brain Res.* **1626**, 66–87 (2015).

145. Gratch, J., Ning, W., Gerten, J., Fast, E. & Duffy, R. Creating rapport with virtual agents. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **4722 LNCS**, 125–138 (Springer, Berlin, Heidelberg, 2007).

This study investigates how artificial agents can create rapport, demonstrating the possibilities and limitations of this technology.

146. Cooke, M., King, S., Garnier, M. & Aubanel, V. The listening talker: A review of human and algorithmic context-induced modifications of speech. *Comput. Speech Lang.* **28**, 543–571 (2014).
147. McEllin, L., Knoblich, G. & Sebanz, N. Distinct kinematic markers of demonstration and joint action coordination? Evidence from virtual xylophone playing. *J. Exp. Psychol. Hum. Percept. Perform.* **44**, 885–897 (2018).

This study shows how people use subtle variations in action kinematics to communicate to a partner in different contexts.

148. Nunamaker Jr, J. F., Chen, M. & Purdin, T. D. Systems development in information systems research. *J. Manag. Inf. Syst.* **7**, 89–106 (1990).
149. Wallace, W. L. *The logic of science in sociology*. (Routledge, 2017).
150. Marsella, S. & Gratch, J. Computational models of emotion as psychological tools. in *Handbook of emotions* 113–132 (The Guilford Press, 2016).
151. Mills, P. F., Harry, B., Stevens, C. J., Knoblich, G. & Keller, P. E. Intentionality of a co-actor influences sensorimotor synchronisation with a virtual partner. *Q. J. Exp. Psychol.* **72**, 1478–1492 (2019).
152. van der Steen, M. C., Schwartze, M., Kotz, S. A. & Keller, P. E. Modeling effects of cerebellar and basal ganglia lesions on adaptation and anticipation during sensorimotor synchronization. *Ann. New York Acad. Sci.* . **1337**, 101–110 (2015).
153. Tracy, L. F., Segina, R. K., Cadiz, M. D. & Stepp, C. E. The Impact of Communication Modality on Voice Production. *J. Speech, Lang. Hear. Res.* **63**, 2913–2920 (2020).
154. Cañigueral, R., Ward, J. A. & Hamilton, A. F. D. C. Effects of being watched on eye gaze and facial displays of typical and autistic individuals during conversation. *Autism* **25**, 210–226 (2021).
155. Pika, S., Wilkinson, R., Kendrick, K. H. & Vernes, S. C. Taking turns: bridging the gap between human and animal communication. *Proc. R. Soc. B* **285**, 20180598 (2018).
156. Henrich, J., Heine, S. J. & Norenzayan, A. Most people are not WEIRD. *Nature* **466**, 29 (2010).
157. Stivers, T. *et al.* Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 10587–92 (2009).

This cross-linguistic study of question-answer pairs in spontaneous conversation shows the speed and accuracy with which people take turns.

158. Dingemanse, M., Torreira, F. & Enfield, N. J. Is “Huh?” a universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PLoS One* **8**, e78273 (2013).
159. Bandura, A. & Walters, R. H. *Social learning theory*. (Prentice Hall: Englewood cliffs, 1977).
160. Haensel, J. X., Smith, T. J. & Senju, A. Cultural differences in mutual gaze during face-to-face interactions: A dual head-mounted eye-tracking study. *Vis. cogn.* https://doi.org/10.1080/13506285.2021.1928354 (2021).
161. De Lillo, M. *et al.* Tracking developmental differences in real-world social attention across adolescence, young adulthood and older adulthood. *Nat. Hum. Behav.* **5**, 1381–1390 (2021).
162. Eaton, L. G. & Funder, D. C. The creation and consequences of the social world: An interactional analysis of extraversion. *Eur. J. Pers.* **17**, 375–395 (2003).
163. Back, M. D., Schmukle, S. C. & Egloff, B. Predicting actual behavior from the explicit and implicit self-concept of personality. *J. Pers. Soc. Psychol.* **97**, 533–548 (2009).
164. Uekermann, J. *et al.* Social cognition in attention-deficit hyperactivity disorder (ADHD). *Neurosci. Biobehav. Rev.* **34**, 734–743 (2010).
165. Heerey, E. A. & Kring, A. M. Interpersonal consequences of social anxiety. *J. Abnorm. Psychol.* **116**, 125–134 (2007).
166. McNaughton, K. A. & Redcay, E. Interpersonal synchrony in Autism. *Curr. Psychiatry Rep.* **22**, 1–11 (2020).
167. Barzy, M., Ferguson, H. J. & Williams, D. M. Perspective influences eye movements during real-life conversation: Mentalising about self versus others in autism. *Autism* **24**, 2153–2165 (2020).
168. Zhao, Z. *et al.* Random and Short-Term Excessive Eye Movement in Children with Autism During Face-to-Face Conversation. *J. Autism Dev. Disord.* https://doi.org/10.1007/s10803-021-05255-7 (2021).
169. Cañigueral, R. & Hamilton, A. F. D. C. The role of eye gaze during natural social interactions in typical and autistic people. *Front. Psychol.* **10**, 560 (2019).

Figure 1: Different ways to organise the study of interaction. Data is captured at the level of behaviour (middle row), including both motor systems (face and body) and perceptual systems (auditory, visual, and somatosensory). These can be mapped in a relatively straightforward fashion to domain-specific cognitive systems, although multi-modal mappings are still important. How different behaviours draw on general cognitive processes (top row) or can be understood in terms of specific or general social meanings (bottom rows) remains a topic of investigation.

Figure 2. Organising behaviour by modality or social meaning. Behaviours are often studied in terms of modality, with one lab investing hand actions while another studies speech or gaze. However, in a face-to-face interaction, people have many modalities available for communication, and may switch rapidly between them. Thus, it may be helpful to group behaviours by their social meaning, not their modality. Here, we give examples of two social meanings ('I'm bored' and 'attend to me') that may be signalled across different modalities (movement, gaze, and voice).

Figure 3. Summary of theories. Theories of face-to-face interaction in terms of social meaning describe an interaction as a whole without specifying cognitive processes (bottom panel). Theories of behavior rules and rich cognition can be specified in cognitive terms (top panel, left) and possibly in terms of brain systems (top panel, right). Here, we suggest that behavior rules map directly from perception to action (for example via the mirror neuron system) without the need for additional processing. Rich cognition theories, on the other hand, require more elaboration and recruitment of Theory of Mind (ToM) neural systems. Both behaviour rules and rich cognition theories might be modulated by motivation.

Figure 4. Linking different approaches to interaction behaviour. To advance the science of social interaction it is necessary to bring together a range of methods including observing behaviour, computational modelling, and experimentation. This integrative approach involves building theories (pink arrows), developing hypothesis-driven experiments (green arrows) and generating computational models (grey arrows).

Box 1: Diversity of social interaction behaviour

Understanding which features of social interaction are universal and which are variable is important for the study of cognitive mechanisms and the application of research across diverse contexts. If an interaction behaviour is universal across situations, cultures and groups, we can infer that it might depend on a specific cognitive mechanism[157,158]. For example, it has been found that 'huh' is a universal word used to indicate a failure of communication, where the speaker needs to repeat or re-explain what they just said[158]. This implies that repairing communication breakdowns is a fundamental and universal process, and might motivate researchers to search for a specific underlying cognitive mechanism. By contrast, identifying features of interaction that vary by population can reveal how context and learning influence behaviour[159]. For example, East Asian participants engage in more eye contact than British participants[160], and older adults in the UK look at faces less than younger adults during face-to-face conversation[161]. This implies that culture and social context can change gaze behaviour.

Social behaviours can also vary substantially between people. Individual differences in personality traits such as extraversion[162] and neuroticism[163] predict social behaviour in real interactions, but understanding how individual-level factors contribute to dyad interactions remains challenging[37]. These individual factors might be even more pronounced in neurodiverse populations. For example, people with ADHD show poor recognition of facial expressions compared to age-matched controls without ADHD diagnoses (although co-morbidities such as depression are often not assessed)[164], and undergraduates with social anxiety are less likely to match the type of smile given by a partner (rather than defaulting to the polite smile type) compared to those without diagnoses of social anxiety[165]. Differences in social behaviour in autistic people have also been extensively studied, and, consistent with the heterogeneity of this population, results have been mixed. Autistic people might show reduced interpersonal synchrony compared to age- and IQ-matched controls[120,166]. However, differences in gaze behaviour are more varied, with some studies reporting that autistic people look less at their partner's face during conversation than age- and IQ-matched controls[167,168] and others reporting no difference[121,169]. Further research with more participants and a variety of contexts and conversation types will be needed to precisely quantify social behaviour differences and similarities in autism. Overall, however, studying the differences in neural and cognitive systems that underlie differences in social behaviour can contribute to theories of face-to-face interaction and support neurodiverse people in their daily lives.