Trends in
Cognitive Sciences

CellPress

Review

Tracking minds in communication

Paula Rubio-Fernandez[1,*], Marlene D. Berke[2], and Julian Jara-Ettinger[2,3,*]

How does social cognition help us communicate through language? At what levels does this interaction occur? In classical views, social cognition is independent of language, and integrating the two can be slow, effortful, and error-prone. But new research into word level processes reveals that communication is brimming with social micro-processes that happen in real time, guiding even the simplest choices like how we use adjectives, articles, and demonstratives. We interpret these findings in the context of advances in theoretical models of social cognition and propose a communicative mind-tracking framework, where social micro-processes are not a secondary process in how we use language – they are fundamental to how communication works.

Highlights

Past research suggesting that social cognition and language have limited interaction has focused on complex mental state inferences, often derived from complete utterances.

Analyzing how people produce and process language at the word level, in real time, reveals rich social micro-processes that permeate linguistic communication.

These social inferences often reflect representations of other people's cognitive processes (such as their memory and attention) rather than only classic notions of mental state attribution, such as beliefs and desires.

These new findings suggest that social cognition and language might have deep, pervasive, and real-time interactions that make linguistic communication possible.

Based on this, we propose a framework where the construction and interpretation of utterances at the word level is supported by social micro-processes that instantiate dynamic representations of other minds.

The connection between social cognition and language

Consider how you might joke around with a friend, hint at a problem to a coworker, or have a candid conversation with a loved one. To do this successfully, you would need to think about what your interlocutor is thinking, and they would need to do the same for you. This intuition is at the heart of many theories of how we use language for everyday communication [1–7], proposing that language use is intrinsically social and requires that we constantly represent and track each other's minds.

Yet, many experiments have failed to find this connection. In simple referential communication tasks – where a listener has to identify what object a speaker is talking about (Figure 1A)– people first interpret the speaker's words using their own egocentric perspective. That is, listeners first look at, and sometimes even reach for, the object that matches the description according to their own knowledge, rather than according to the speaker's knowledge [8–11]. This seemingly slow interaction between language and social cognition is consistent with evidence that the language and Theory of Mind networks are fully dissociated, functionally and anatomically [12–14].

Even frameworks that embrace the importance of social cognition in communication often treat it a secondary process [15–19]. According to these two-stage models, people first decode a sentence's literal meaning and then apply social cognition to infer its enriched meaning. For example, if someone says 'it's getting late,', the listener first interprets it as a factual comment about time and then uses domain-general social cognition to figure out why the speaker made this remark in this particular context – perhaps signaling it's time to leave. This process where social cognition operates over literal representations of meaning, a form of **pragmatic reasoning** (see Glossary), successfully explains many complex phenomena in communication, such as irony, metaphor, and indirect speech [19]. While powerful, this process is computationally costly [20], which could further explain why language cannot pervasively rely on social cognition.

Nonetheless, advances in theoretical and empirical work have converged on a set of views – which we broadly call 'mind-tracking' approaches – that reveal a pervasive contribution of social cognition to linguistic communication. These approaches offer three key explanations for why the connection has historically appeared impoverished. We review each in detail in the following section, but in summary: first, social cognition's contribution to language use is specialized for

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]Department of Psychology, Yale University, New Haven, CT, USA
[3]Wu Tsai Institute, Yale University, New Haven, CT, USA

*Correspondence:
Paula.RubioFernandez@mpi.nl (P. Rubio-Fernandez) and
Julian.jara-ettinger@yale.edu (J. Jara-Ettinger).

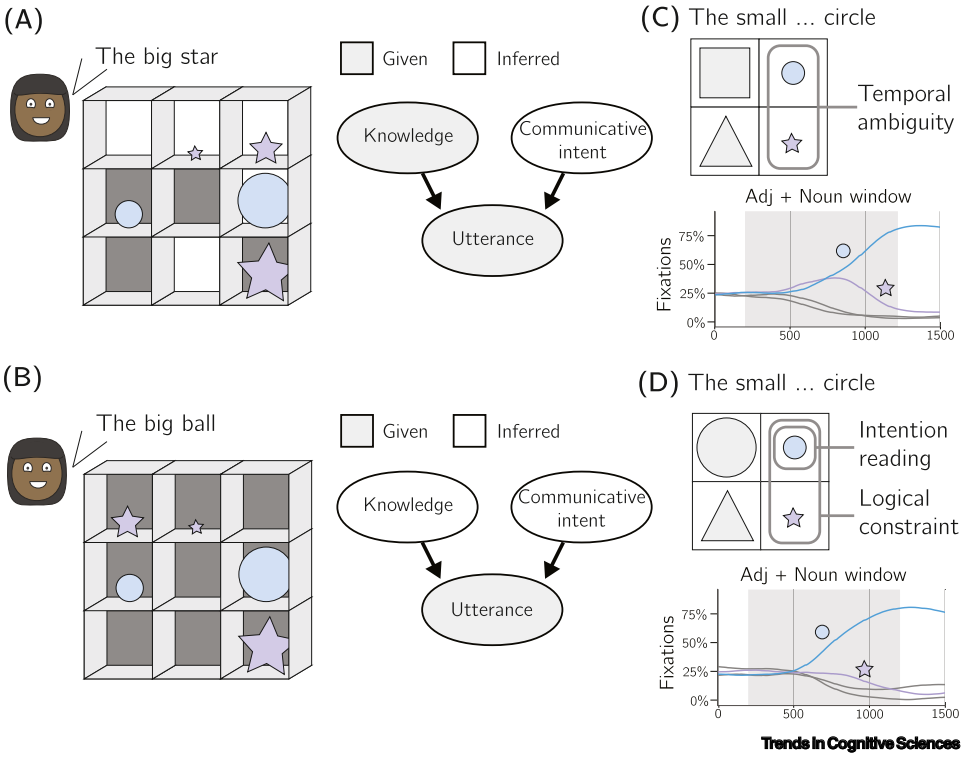

Figure 1. Listener's capacity to infer referential intent. (A) Paradigm suggesting people struggle to integrate mental states in language understanding. A grid of objects stands between a participant with full visual access and a confederate that only sees the unoccluded objects. When the confederate requests an object, participants first interpret the expression egocentrically [8–10]. In this example, participants might first look at the bottom-right corner (the big star according to their own knowledge) rather than at the top-right corner (the big star according to the confederate's knowledge). (B) Modified version where participants are unsure of what the speaker sees and must simultaneously infer the speaker's knowledge and communicative intent. This task shows radically different results, with participants now providing fine-grained joint inferences about what the speaker is talking about and what they know. This is arguably because this task is closer to real-world situations, where we do not necessarily know what others know, and occlusion does not imply ignorance. (C) Real-time language processing under temporal ambiguity. Upon hearing 'the small …', participants fixate on the two small objects and then fixate on the circle upon hearing '… circle'. (D) Real-time pragmatic inference. In a nearly identical display, listeners immediately fixate on the small circle upon hearing 'the small …', because they understand the speaker intends to contrast the referent with a larger alternative. Data in panels (C) and (D) are from Hindi speakers in [21], but this pattern appears across multiple languages.

problems faced in everyday communication. While the connection is swift and dynamic in tasks that mirror real-world interactions, it becomes effortful and prone to errors in classical tasks that impose unusual assumptions. Second, social cognition operates in real time to support language use. These computations shape online construction and interpretation of sentences, often reaching completion before a first interpretation is even available for secondary analysis. Therefore, examining how social cognition modifies an existing interpretation of a complete sentence overlooks the local processes that shape meaning as sentences are being understood. Third, many social computations are not adequately captured in frameworks that exclusively focus on propositional attributions of mental states like beliefs, desires, and intentions. Instead, the social computations supporting real-time linguistic communication frequently involve tracking cognitive processes, like attention, decision making, and recall.

Our article is organized around three core sections. First, we review new research supporting each of these three arguments, explaining how and when social cognition supports linguistic

Glossary

Mentalistic representations: agent representations that cannot be expressed by or reduced to physical representations alone, requiring us to posit a subjective observer (i.e., a mind). For instance, representations of an agent's desires, attention, or reasoning are mentalistic. Representations of their body position and face orientation are not. Mentalistic representations do not need to be propositional and can involve minimal or partial representations of a mind.

Pragmatic reasoning, processes, or inference: computations through which speakers tailor their messages to their addressees, and listeners interpret the messages in context (e.g., a listener successfully understanding that 'she' refers to Jane). Pragmatics encompass any contribution of context to meaning, but here we focus on social contexts and the mentalistic representations needed to represent them. Pragmatic reasoning spans from sentence-level inferences to word-level ones.

Social micro-processes: temporally-bound mentalistic representations that allow us to reason about others without fully modeling their minds. Examples include estimating someone's attention from their gaze, or inferring that a hesitant pause reflects internal reasoning. These lightweight representations can be quickly instantiated and discarded for real-time social behavior, contrasting with more traditional forms of social cognition where we represent complex configurations of others' beliefs and desires to explain behavior over extended periods.

Word-level social computations: process through which mentalistic representations support word level production and interpretation. These local computations differ from sentence-level ones in that they typically use social micro-processes (e.g., saying 'that one' to guide the listener's attention to a referent) rather than over propositional mental states (e.g., processing a complete sentence to derive an indirect meaning).

communication. Together, this research reveals a pervasive connection between social cognition and language. We then introduce mind-tracking approaches, including a specific instantiation we call the Communicative Mind-Tracking framework (CMT), which grounds these ideas using the language of planning frameworks and posits that language use is supported by **social microprocesses**. Finally, we argue the CMT provides a unified account that explains how people communicate, track the interaction's success as it unfolds in real time, and repair it when it breaks down.

When and how social cognition supports linguistic communication

Supporting communication in tasks that mirror everyday conversational pressures

Is it possible that past experiments found failures in social cognition because they used unnatural forms of communicative interaction? Consider again situations like the one in Figure 1A, where people show an egocentric bias. This is a simple task because the listener knows what the speaker knows and what they said. All they have to do is deduce the speaker's communicative intent. But this situation is unusual in several ways [22]. To start, creating situations where the meaning of an expression changes depending on whose knowledge you consider (the speaker's or your own) requires careful and clever experimental set ups, with scripted speakers, that people might not usually encounter in everyday life. When regular people are placed in the role of the speaker, they spontaneously offer additional information to prevent ambiguities from arising in the first place (e.g., such as saying 'the big star at the top' in Figure 1A [23]).

These paradigms make another unusual assumption: that people only know about what they can currently see. In everyday life, people know much more than what's in their visual field, and it is unremarkable when someone references something that is not in their view. Restricting our interpretation of what someone says to only what they can currently see would be a serious mistake in real-life conversation.

Finally, this type of task does not capture how social cognition needs to support everyday conversation. In everyday life, we rarely enter a conversation with a new person having a complete description of their relevant knowledge. This is something we have to infer based on how our interlocutor speaks (Figure 1B). Even though this is a more challenging inference (requiring people to simultaneously infer the speaker's communicative intent and knowledge from the utterance), adults are surprisingly skilled at it, making fine-grained inferences about others' mental states based on their exact choice of words (and even appropriately discounting a speaker's propensity to use adjectives redundantly [24]). In some cases, these inferences are even spontaneous and automatic [25].

Supporting communication at the world level

A second reason why the connection between social cognition and language use can appear impoverished is because of a traditional focus on how the two systems interact only after a full sentence has been produced or decoded (e.g., [15–18]). Departing from a sentence-level analysis and zooming into how words are produced and processed in real time [26] uncovers social cognition's pervasive support. At this level, adults' production and processing of words reveal rich representations of what they expect each other to understand (avoiding ambiguity), what knowledge they share (tracking common ground), and how to direct attention efficiently (manipulating attention). We review the key findings behind each of these conclusions next.

Avoiding ambiguity. Adults strategically adjust how specific their words are to avoid ambiguity (e.g., referring to a pet as a 'dog' when no other dogs are present, but switching to 'Dalmatian' otherwise [27]). Conversely, listeners can infer a speaker's knowledge based on these choices

(e.g., referring to a 'Dalmatian' instead of a 'dog' suggests the speaker knows there are other dogs in the scene [24]). While seemingly simple, this flexible change in specificity requires speakers and listeners to track each others' knowledge and expectations about potential interpretations that could create ambiguity in real-time language use.

Social cognition not only shapes how we choose and interpret nouns, but also how we modify them with adjectives. Consider the simple situations shown in Figure 1C,D. When a listener hears 'look at the small…', a literal interpretation of these words would narrow down the reference to two possibilities: the small circle or the small star (as in Figure 1C). However, if this real-time interpretation goes beyond literal semantics, the adjective 'small' can sometimes reveal the speaker's referential intention before the noun is even mentioned. In the event in Figure 1D, the speaker's decision to say "small" suggests that they wanted to preempt an ambiguity between the two circles (since the adjective would be unnecessary to refer to the star). This type of pragmatic inference operates in real time [28], also with other adjective types (such as material and scalar adjectives [29,30]), and it appears across cultures [21,31]. At the same time, the inferences available to listeners vary across languages depending on whether they position adjectives before or after nouns, further showing that these inferences do indeed occur at the word rather than the sentence level [21].

Tracking common ground. Adults' word choices also reveal that they track what knowledge they share with their interlocutors. In classic coordination games with abstract shapes, participants quickly invent new labels (e.g., 'the skater' or 'the Viking ship') that they use exclusively with the label's co-inventor, but not with others, revealing that we track how we use different words with different people [32,33].

Representations of shared knowledge are so important that they can be part of grammar itself. Consider articles such as 'a' and 'the' in English. These articles encode two main types of meaning [34]. The first are non-social meanings that signal whether the speaker has a specific referent in mind (e.g., 'I want to talk to the manager' versus 'I want to talk to a manager'). The second are social meanings that convey whether the referent is new or familiar to the listener (e.g., 'We bought a house' versus 'We bought the house'). To use the social meanings appropriately, speakers must track what their interlocutor knows, and listeners must integrate this information to identify the intended referent. For example, if your dog runs away in a park and you walk up to a stranger to ask if they have seen it, having a specific dog in mind would warrant the use of a definite article ('Have you seen the dog?'). However, because the stranger does not know your dog, you should use an indefinite article instead ('Have you seen a dog?'). And indeed, the listener would understand that you are looking for a specific dog, not just any dog, therefore understanding the pragmatic meaning of the indefinite description [35].

Definite markers can signal common ground so strongly that they can even override the literal description of the referent (e.g., if there are two cars and a speaker mentions one of them, adults will then interpret 'that car' as referring to the same object, even if it has been transformed into a toy duck and is no longer a car at all [36]). The way we seamlessly select and interpret indefinite and definite descriptions to distinguish what information is new and familiar to our interlocutors shows how social cognition supports real-time communication.

Manipulating attention. In some situations, adjectives are technically unnecessary (e.g., referring to a cup as 'my blue cup' when no other cups are in sight). However, adults often use color words anyway to help listeners because searching for colors is easier than searching for objects, revealing a deep integration of social cognition in referential communication. The more objects there are

in a scene, the more likely adults are to add color adjectives to help the listener (Figure 2A, left [38]). This strategy becomes more pronounced when adults are talking to children, recognizing that children need additional support [41]. However, adults do not use this strategy when all objects are the same color (and so the color adjective is no longer helpful, Figure 2A, right [42]), when the listener already knows the intended referent [37], or when the listener cannot see the objects [39].

The use of adjectives to guide attention also successfully predicts cross-linguistic variation. For instance, American Sign Language (ASL) has flexible adjective placement. Nonetheless, ASL signers often place color and material adjectives before the noun to help the addressee's visual search and they place scalar adjectives after the noun because they require the noun to be interpreted first (e.g., small has a different scale on 'small car' versus 'small pencil' [43]). And, in languages where adjectives follow nouns, like Spanish, speakers use fewer redundant color words, because they are less helpful for visual search, particularly when there are few objects [38]. This is because, by the time the speaker finishes uttering the noun, the listener has probably already identified the referent.

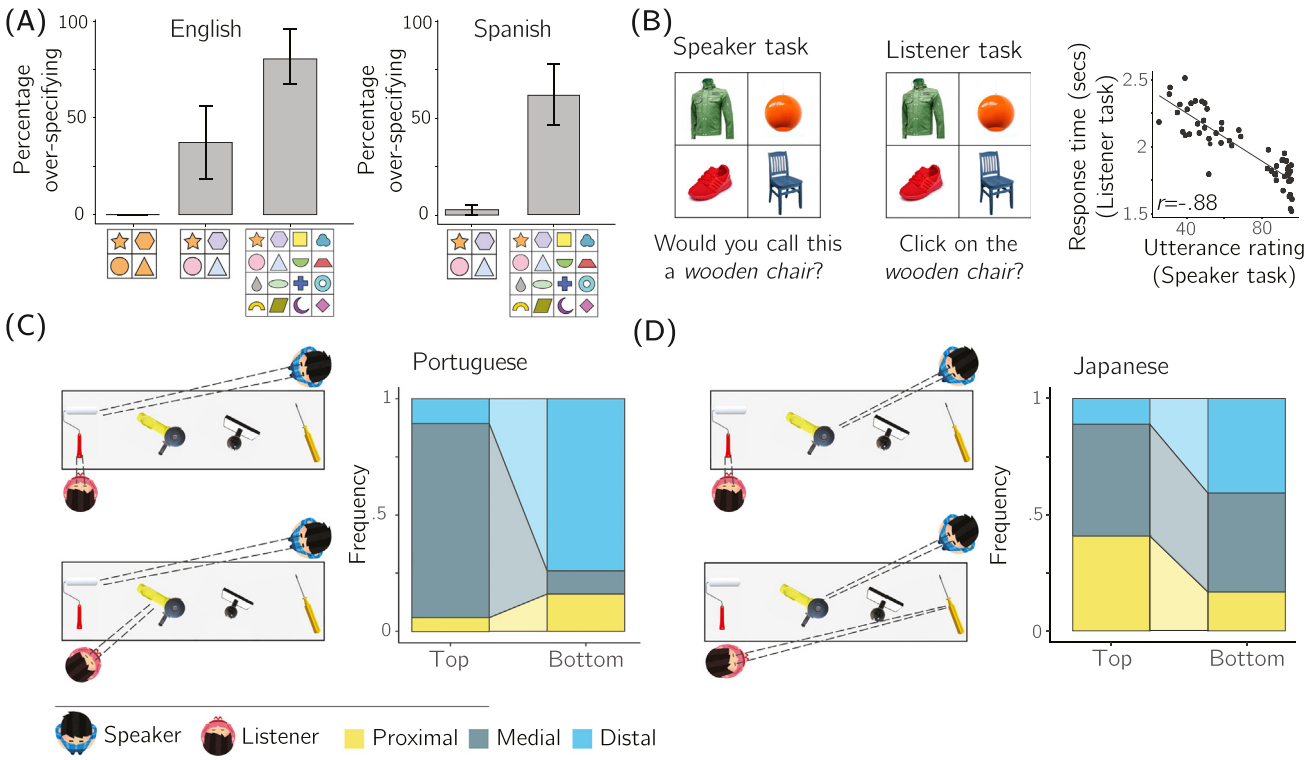

Figure 2. Speaker word choices guide listener attention. (A) Propensity to use redundant color adjectives in English (left plot) as a function of color variability (monochrome vs. polychrome objects, from [37]) and set size (4 versus 16 objects), also in Spanish (right plot, from [38]). This shows how speakers strategically use color adjectives to facilitate listener visual search, showing sensitivity to color variability, set size, and adjective position. (B) English speakers' ratings of how likely they would be to use a particular referential expression in a visual display (speaker task) are negatively correlated with the response time needed for people to identify the object given the same expression (listener task). This suggests that, in visual contexts, adults have a natural preference for referential expressions that reduce search time for listeners (both for color and material adjectives [39]). (C,D) Demonstrative choices by Portuguese (D) and Japanese (D) speakers in pairs of events where the speaker (top of scene; blue) has the same referential intent (the second object), but the listener (bottom of scene; pink) is looking in different places (from [40]). Panel (C) contrasts joint attention versus the need for attention correction and (D) contrasts direction in which attention needs correcting. These results show how demonstratives are not only used to mark spatial locations but also show sensitivity to listener attention.

These phenomena are not specific to color adjectives, nor do they require adults to be in a situation where they are thinking about visual search. When simply asked how likely they would be to use different referential expressions (e.g., 'the leather couch' or 'the red couch'), without any mention of visual search or a listener, adults naturally prefer expressions that minimize the time it would take for a listener to locate the object (Figure 2B [39]), revealing that a sensitivity to other people's attention shapes how appropriate different expressions sound to us.

Interestingly, attention manipulation is also a core feature of a linguistic universal: demonstratives. Traditionally, demonstratives – words like 'this' and 'that' in English – are thought to encode spatial locations [44,45]. For instance, the proximal 'this' indicates something close to the speaker, while the distal 'that' indicates something farther away. In languages like Spanish, which have an additional medial demonstrative, meanings sometimes encode location relative to both participants in the interaction (e.g., the medial 'ese' means far from the speaker but close to the listener). Traditionally, speakers are thought to use demonstratives to mark spatial regions that the listener can use to identify the referent.

However, recent work has shown that demonstratives are intrinsic attention tools [46,47]. We use them to signal something in joint attention (using the proximal in two-way systems and the medial in three-way systems), to pull the listener's attention towards us (through the proximal), and to push their attention away (through the distal), controlling for the referent's location (Figure 2C, D) [40]. This suggests that demonstratives are more than spatial markers: they are also devices to establish joint attention.

Demonstratives are also often accompanied by pointing and even the way we point reveals we are tracking the listener's perceptual abilities. Speakers tend to use pointing with proximal demonstratives in near space, where pointing is more accurate, but not with distal demonstratives in far space, where the gesture would be more ambiguous for the listener – a consistent effect across different languages; [48–50].

Representing cognitive processes for communication

Beyond the phenomena we reviewed, many other **word-level social computations**have been identified in pragmatics (e.g., [3,51–55]). But these word-level phenomena are rarely considered when discussing whether **mentalistic representations** support language use. This is partly because the standard model of social cognition (Figure 3A) does not offer the right vocabulary to capture these processes, even though they are social. Consider the finding that adults use

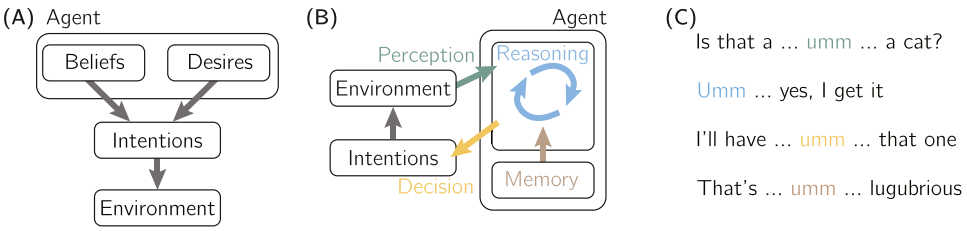

Figure 3. Moving from representations of mental states to representations of cognitive processes. (A) Classical belief-desire-intention model. (B) Model of primary cognitive processes we represent in other minds: real-time perception, reasoning, memory, and decision making. (C) Example of four disfluencies that we readily map onto different cognitive processes happening in the speaker's mind. In the first case, the disfluency suggests the person was pausing to ensure their perception was accurate; in the second case, the person was thinking further to ensure they were confident in their opinion; in the third case, they were making a decision; and in the final case, they were retrieving a low-frequency word from memory.

redundant adjectives to speed up the listener's referent identification, or that demonstratives redirect the listener's attention. Cases like these cannot be reduced to the belief or desire we intend to ultimately impart to the interlocutor. Instead, they require appealing to the underlying processes by which interlocutors form these mental states.

When we zoom into the word-level processes behind how people communicate, capturing those social computations requires us to also zoom into how we represent the real-time processes in other minds: their cognitive processes (Figure 3B). Focusing on dynamic cognitive processes in other minds offers a direct vocabulary to capture the phenomena reviewed in the previous section, such as cueing a listener to retrieve a referent from memory (by using the definite article 'the'), or efficiently guiding their attention towards an object in their visual field (by using a color adjective).

This idea converges with recent advances in models of social cognition that emphasize the importance of representing the dynamic cognitive processes that construct and manipulate mental states: People's real-time perception, recall, reasoning, and decision making (Figure 3B). Recent work has shown that people are surprisingly skilled at representing these cognitive processes. For instance, from the length of a single pause, people can infer what the speaker might be thinking about [56–58], how strong their preferences might be [59,60], whether they got distracted [61], or whether they are thinking about something for the first time [62]. As the next section shows, combining these two sets of parallel advances offers a clear way to capture how social computations support real-time communication.

The Communicative Mind-Tracking framework

The arguments reviewed earlier converge on a set of approaches that emphasize the pervasive support that social computations provide to linguistic communication, which we broadly refer to as 'mind-tracking' approaches (e.g., [3,4,33,53,63–70]), and posit that these representations are needed to leap from non-human primate communication to human language [71–74]. Here we present a concrete instantiation that offers a unified way to understand these phenomena and their computational substrate (formalized in Box 1), grounded in the language of mentalistic representations and planning. Our starting point is the long tradition of conceptualizing language production as a planning problem [75–78], but we further argue this planning is over an interlocutor's cognitive processes. That is, people approach linguistic communication as a planning problem, where the goal of a sentence is to guide a listener's real-time cognitive processes – including their attention, memory, and reasoning – toward constructing a target mental state. This approach highlights that analyzing communication should go beyond focusing on the final mental states a sentence induces in the listener and also consider the underlying trajectory of cognitive processes that the listener undergoes to reach these mental states.

Under a planning approach, we can formalize how actions affect the world through a transition function. Analogously, CMT posits a transition function that captures how words (actions) affect our interlocutor's mind (states)– both their mental states and ongoing cognitive processes (Box 1). Thus, all communication is grounded in expectations about how it guides the interlocutors' mind, making it inherently mentalistic. However, these are not standard full-blown propositional representations of a mind. Consider three examples: using 'the' to create an expectation for a known referent, a color adjective to trigger the listener's visual search, and 'this' to redirect their attention. None of these cases require a full-fledged, complete representation of a mind. Instead, they are time-bounded expectations about the listener's mind, helping us to reason about their dynamic processing of the sentence even before any propositional content can be derived. We call these social micro-processes.

Box 1. The computational basis of the CMT

CMT casts communication as the problem of constructing a sequence of words $\vec{w} = \{w_1, , w_n\}$ that, when processed incrementally, induce state s, which can be a mental state, a cognitive process, or a combination of both. This can be conceptualized as a planning problem where actions are words and states are representations of the interlocutors' mind. The expected effect words have on a listener's mind is defined by a transition function T where $T(s_{i-1}, f, s_i)$ is the probability that a mind transitions onto state s_i, upon hearing fragment f, given a previous state s_{i-1}. f is a subset of $\vec{w}$, allowing T to encode transitions at multiple levels of resolution, but frequently operating at the word level.

T models how minds change as a communicative interaction unfolds, with social micro-processes playing two supporting roles: computing the effect of a word on a listener mind, and representing the ending state. In the first role, social micro-processes determine which state a mind will enter after hearing a word. These transitions can then become hard-coded into T, making them implicit and automatic (e.g., 'the' prepares a listener to search in common ground), but they can also be computed online in communication (e.g., determining whether your interlocutor will recall the referent given a referential pact). In addition, social micro-processes can also represent the final state s itself. For instance, the outcome effect of a demonstrative is itself mentalistic as it is an instruction to redirect attention.

CMT is compatible with Gricean approaches, positing that the planning goal is to maximize utility function

$$U(\vec{w}s) = \overbrace{p_L(s|\vec{w})}^{\text{say as much as necessary...}} - \underbrace{C(\vec{w})}_{\text{...but no more}}, \quad (1)$$

but it emphasizes the collaborative costs, where $C(\vec{w})$ is not only a production cost for the speaker, but the decoding and reasoning costs it imposes on the listener [33]. CMT therefore allows mental states to be further refined via recursive social reasoning.

Two recent computational models have instantiated CMT. First, adults' nuanced patterns of adjective use can be predicted with quantitative accuracy through a model that constructs sentences by modeling how the words, processed incrementally, will sequentially guide the listener's visual search in real time [39]. Second, modeling demonstratives (e.g., 'this' versus 'that') as attention tools where their use redirects the listener's attention produces more human-like behavior compared with models where demonstratives mark regions in space and guide attention through a secondary pragmatic process [40,108].

Critically, the integration of social cognition into language use is specialized for the types of problems we confront in everyday communication and it becomes slower, more effortful, or prone to errors when people engage in unusual communicative interactions (including some laboratory tasks). This suggests that the transition function capturing how words guide cognitive processes is largely an implicit model built with experience [75], rather than an explicit transition function that people generate and use on the spot by drawing on domain-general social cognition, pointing to a form of model-free, but mentalistic, communication.

CMT rejects the two-stage model notion of a literal non-social message that is initially computed and becomes social only through enrichment processes (based on [15–18]), but it does not conflict with other, complementary, frameworks of social language use. While an exhaustive comparison is beyond the scope of this paper, two relevant ones are worth mentioning. First, two-stage models of pragmatic reasoning can still be applied in this framework, allowing for additional recursive pragmatic reasoning at the sentence level (e.g., about the communicative intention behind an entire message). Second, our approach can be compatible with the relevance theoretic account of word-level pragmatics and the idea that interlocutors constantly synthesize *ad hoc* concepts in conversation [53], but extends it by offering a planning approach over targeted trajectories in mental states.

Achieving communicative success and repairing it when it breaks down

According to the CMT, speakers are guiding listeners' cognitive processes in real time when they communicate, but they cannot directly see whether the listener is processing the message as

intended because other minds are unobservable. Therefore, verbal communication should be complemented by mechanisms that help both the speaker and the listener track that the message is unfolding successfully so that any miscommunication can be repaired. As expected by CMT, these processes are not only present but pervasive in everyday conversation.

In face-to-face interactions, listeners provide real-time non-verbal feedback to help the speaker ensure the message is being understood [68,79–81]. This is done through feedback that includes eye contact, nodding [82], and brief verbal responses such as 'mhm' or 'yeah' [83]. Crucially, speakers adapt their contributions in real time by responding to these subtle visual cues (e.g., giving shorter answers if the listener signals comprehension [84]).

Conversely, listeners also rely on non-linguistic signals from speakers – like pauses and disfluencies – to refine their understanding. Even toddlers are sensitive to disfluencies and use them to infer that the speaker is about to refer to something new [85]. As adults, inferences about pauses are context sensitive [86] and can lead us to infer that the speaker is trying to deceive [87], is uncomfortable with the topic [88], or is about to introduce something complex or novel [89–91]. Figure 3C shows how disfluencies can trigger a variety of inferences about the speaker's cognitive processes, which in turn, help us plan how to respond.

Listeners are also sensitive to multimodal signals that guide language interpretation. The speaker's gaze and pointing gestures are often key to the listener's reference resolution [92], and facial signals can influence how listeners interpret the speaker's intention during conversation [93] and help identify speech acts (e.g., eyebrow movements can help listeners recognize the speaker is asking a question [94–96]).

These signals not only increase successful face-to-face communication but also play a critical role in detecting when communication is breaking down. In these cases, adults engage in interactive repair [66]: when a listener realizes they are not understanding the message, they quickly signal this to the speaker with simple cues like raised eyebrows or an interjection like 'huh?'. This allows the speaker to decide whether to repeat, rephrase, or reassure the listener, depending on the situation [97,98]. The speed and flexibility with which these social processes unfold in conversation further show that social micro-processes are central to real-time communication [99].

Concluding remarks

The recent advances we reviewed highlight a set of mind-tracking approaches (e.g., [3,4,33,53,63–70]) that view language use as an intrinsically social activity, with social computations happening in real time and at the level of word choice and processing. We also offered a concrete instantiation called the CMT, which casts communication as a planning problem and integrates advances in models of social cognition to clarify which social representations support real-time communication. In the remainder, we touch on important implications of mind-tracking approaches (see Outstanding questions).

First, mind-tracking approaches generally view all word-level processes as pragmatic from the onset rather than operating at a secondary stage after the message has been interpreted literally. These word-level pragmatics sometimes mirror the logic from sentence-level pragmatics (e.g., inferring that the word 'some' implies 'not all' by extracting a literal meaning that is enriched by further social reasoning [100,101]), but the meaning can also be intrinsically mentalistic from the onset (e.g., demonstratives directly encoding how to redirect listener attention [40]). This view is consistent with evidence that many languages directly encode mentalistic representations in grammar itself [64,65,102,103]. For example, Japanese uses evidentials to mark the source of

Outstanding questions

To what extent are word-level social computations performed online vs precomputed and stored for quick access? For example, experience with referential communication may lead to the automatic use of adjectives to distinguish between two objects of the same kind, without having to reason about how the listener would interpret alternatives (i.e., computing that they would find a bare noun ambiguous). If so, what are the mechanisms that determine when and how mentalistic inferences are stored for future retrieval?

Are some aspects of social cognition specialized for supporting communicative interactions rather than more general forms of action understanding? And if so, would those specialized forms of social cognition be internal to the brain's language network?

How does sentence-level pragmatic reasoning work in orchestration with word-level inferences? One possibility is that, in many important cases, word-level processes produce the same result than would be obtained by computing a literal message and then refining it through sentence level pragmatics. If so, then sentence-level pragmatic reasoning could be a computational-level description of processes that are often algorithmically implemented at the word level.

How do social micro-processes emerge, develop, and vary across species? To what extent are they unique to humans and shaped by pressures to communicate and navigate the social world?

How do we make sense of recent successes in Large Language Models? On the one hand, these models exhibit proficient language use, and this presumably requires mastering complex use of nouns, adjectives, demonstratives, and articles, and in our account, this requires social micro-processes. On the other hand, these models are not grounded in reference to the real world, where mentalistic representations appear to be essential.

one's knowledge (e.g., perception versus hearsay); Kogi, a Chibchan language of Colombia, uses verb prefixes to signal whether the speaker has exclusive knowledge or shares it with the listener; and Kakataibo, a Panoan language of Peru, differentiates the accessibility of information in recounted events (narrative genre) and the here-and-now (conversational genre).

Integrating CMT with standard models of pragmatic reasoning [15–19] might suggest a two-system theory. One system uses specialized social micro-processes at the word level, while the other applies domain-general social cognition over complete linguistic representations (see Box 2 for hypotheses about neural implementations that might help answer this question). This would also be consistent with dual models of Theory of Mind where one is fast and automatic and the other is slow and effortful [104]. However, our approach does not necessarily predict two systems. Mentalistic representations might be multi-granular, allowing people to flexibly model other minds with different degrees of complexity [105]. Under this view, social micro-processes are the most minimal instantiations of a mind that allow us to reason about others (e.g., representing an agent's focus of attention and the place where we wish to redirect it) and can be scaled up to to richer models of another mind, all the way up to full-fledged propositional models.

Alternatively, social micro-processes might instantiate a type of mentalistic representation that is qualitatively different from the ones used in non-linguistic propositional Theory of Mind. This could be particularly true if these social micro-processes are implemented as a fast forward-model that is specific to language (Box 2).

What this literature does show, is that using our Theory of Mind is not inevitably costly and effortful [8–10,104,106]. Many mentalistic representations are fast, cheap, and readily accessed in social interactions (see [107] for review). The mentalistic representations that people find easy to compute might be precisely the ones that language recruits. This suggests a fruitful bi-directional research program: identifying the mentalistic representations that support real-time communication can uncover the foundational representations of social cognition. Conversely, identifying rapid or

Box 2. Neural basis of social micro-processes

How are the types of social micro-processes we proposed instantiated in the brain? Although the Theory of Mind network and the language network are largely thought to be functionally and anatomically separable [12], the current literature points to two ways in which this interaction might occur. First, the language and Theory of Mind networks are synchronized at rest and in comprehension, more so than with other areas like the multiple demand network [109], which might reflect a pervasive interaction that supports all language use. Second, some evidence suggests that the two networks might overlap in the superior temporal sulcus (STS) [110] – an ideal place given the role of the STS in many social computations [111], including encoding eye gaze and attention [112] – although the evidence is contested [13].

It is also possible that the types of mentalistic representations supporting language are not part of what is standardly thought to be the Theory of Mind network. This is because this network is usually localized through tasks that target explicit representational attribution of mental states via false beliefs [113], which would fail to identify regions specialized for social micro-processes. Non-verbal localizers might be better suited for this task [114], provided that social micro-processes are not language specific.

Another intriguing possibility comes from recent research suggesting that the cerebellum is engaged in a large variety of cognitive functions [115], including social computations [116]. One hypothesis is that the cerebellum is particularly well-suited for encoding forward models that support planning and prediction [117–119], which would make it a strong candidate for the transition function that the CMT posits.

Finally, social cognition might be so essential to linguistic communication that it may be fully specialized and part of the language network itself, such that it is always active in any form of linguistic communication, but not in non-verbal social tasks. The challenge, ultimately, is that relatively little is known about the neural basis of social micro-processes. Answers to these questions might also ultimately shed light on whether language use is supported by a single domain-general Theory of Mind system, or by two separate systems that work in orchestration to support communication.

automatic processes within social cognition can reveal which social computations are readily available in language use.

The challenge is that we ultimately do not know as much about the nature of social micro-processes relative to richer propositional representations. As such, our ability to develop a full-fledged framework of linguistic communication is fundamentally constrained by our models of social cognition. A complete characterization of the social micro-processes that structure language down to the level of words and grammar will ultimately reveal what makes humans exceptional at understanding each other and making ourselves understood.

Acknowledgments

Thanks to Aaron Baker, Daniel Harris, Alexander Paunov, Hilary Richardson, Amanda Royka, and Urvi Suwal for comments. This work was supported by National Science Foundation (NSF) award BCS-2045778 to J.J.E. and Norwegian Research Council (NRC) award FRIPRO-275505 to P.R.F.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used GPT-4o in order to create a customized grammar checker that identified sentences that were too long or did not use a right-branching structure, and to identify unnecessary adjectives and adverbs. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of interests

No interests are declared.

References

1. Tomasello, M. (2005) *Constructing a language: A usage-based theory of language acquisition*, Harvard university press
2. Christiansen, M.H. and Chater, N. (2016) *Creating language: Integrating evolution, acquisition, and processing*, MIT Press
3. Clark, H.H. (1996) *Using language*, Cambridge university press
4. Sperber, D. and Wilson, D. (1986) *Relevance: Communication and cognition*, vol. 142. Harvard University Press Cambridge, MA
5. Jackendoff, R.S. (2002) *Foundations of language: Brain, meaning, grammar, evolution*, Oxford University Press
6. Scott-Phillips, T. (2014) *Speaking our minds: Why human communication is different, and how language evolved to make it special*, Bloomsbury Publishing
7. Levinson, S.C. (2020) On the human "interaction engine". In *Roots of human sociality, Routledge*, pp. 39–69
8. Keysar, B. *et al.* (2000) Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychol. Sci.* 11, 32–38
9. Lin, S. *et al.* (2010) Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *J. Exp. Soc. Psychol.* 46, 551–556
10. Horton, W.S. and Keysar, B. (1996) When do speakers take into account common ground? *Cognition* 59, 91–117
11. Epley, N. *et al.* (2004) Perspective taking in children and adults: Equivalent egocentrism but differential correction. *J. Exp. Soc. Psychol.* 40, 760–768
12. Fedorenko, E. *et al.* (2024) The language network as a natural kind within the broader landscape of the human brain. *Nat. Rev. Neurosci.* 1–24
13. Shain, C. *et al.* (2023) No evidence of theory of mind reasoning in the human language network. *Cereb. Cortex* 33, 6299–6319
14. Paunov, A.M. *et al.* (2022) Differential tracking of linguistic vs. mental state content in naturalistic stimuli by language and theory of mind (tom) brain networks. *Neurobiol. Lang.* 3, 413–440
15. Grice, H.P. (1975) *Logic and conversation*, Speech acts, Brill, in, pp. 41–58
16. Searle, J. (1993) Metaphor. In *Metaphor and Thought*, pp. 83–111, Cambridge University Press
17. Dascal, M. (1987) Defending literal meaning. *Cogn. Sci.*,
18. Dascal, M. (1989) On the roles of context and literal meaning in understanding. *Cogn. Sci.*,
19. Goodman, N.D. and Frank, M.C. (2016) Pragmatic language interpretation as probabilistic inference. *Trends Cogn. Sci.* 20, 818–829
20. Van De Pol, I. *et al.* (2018) Parameterized complexity of theory of mind reasoning in dynamic epistemic logic. *J. Log. Lang. Inf.* 27, 255–294
21. Rubio-Fernandez, P. and Jara-Ettinger, J. (2020) Incrementality and efficiency shape pragmatics across languages. *Proc. Natl. Acad. Sci.* 117, 13399–13404
22. Rubio-Fernández, P. (2017) The director task: A test of theory-of-mind use or selective attention? *Psychon. Bull. Rev.* 24, 1121–1128
23. Hawkins, R.D. *et al.* (2021) The division of labor in communication: Speakers help listeners account for asymmetries in visual perspective. *Cogn. Sci.* 45, e12926
24. Jara-Ettinger, J. and Rubio-Fernandez, P. (2021) Quantitative mental state attributions in language understanding. *Sci. Adv.* 7, eabj0970
25. Rubio-Fernández, P. *et al.* (2019) How do you know that? automatic belief inferences in passing conversation. *Cognition* 193, 104011
26. Hagoort, P. and van Berkum, J. (2007) Beyond the sentence given. *Philos. Trans. R. Soc. B Biol. Sci.* 362, 801–811
27. Degen, J. *et al.* (2020) When redundancy is useful: A bayesian approach to "overinformative" referring expressions. *Psychol. Rev.* 127, 591
28. Sedivy, J.C. *et al.* (1999) Achieving incremental semantic interpretation through contextual representation. *Cognition* 71, 109–147
29. Sedivy, J.C. (2005) Evaluating explanations for referential context effects: Evidence for cricean mechanisms in online language interpretation. In *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*, pp. 345–364, MIT Press
30. Ronderos, C. *et al.* (2024) Perceptual, semantic and pragmatic factors affect the derivation of contrastive inferences. *Open Mind*,

31. Ryskin, R. *et al.* (2023) Real-time inference in communication across cultures: Evidence from a nonindustrialized society. *J. Exp. Psychol. Gen.* 152, 1245
32. Horton, W.S. and Gerrig, R.J. (2002) Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *J. Mem. Lang.* 47, 589–606
33. Clark, H.H. and Wilkes-Gibbs, D. (1986) Referring as a collaborative process. *Cognition* 22, 1–39
34. Abbott, B. (2006) *Definiteness and indefiniteness*. The handbook of pragmatics pp. 122–149
35. Rubio-Fernandez, P. (2025) First acquiring articles in a second language: A new approach to the study of language and social cognition. *Lingua* 313, 103851
36. Brody, G. and Feiman, R. (2024) Mapping words to the world: Adults, but not children, understand how mismatching descriptions refer. *J. Exp. Psychol. Gen.*,
37. Rubio-Fernandez, P. (2019) Overinformative speakers are cooperative: Revisiting the gricean maxim of quantity. *Cogn. Sci.* 43, e12797
38. Rubio-Fernandez, P. *et al.* (2021) Speakers and listeners exploit word order for communicative efficiency: A cross-linguistic investigation. *J. Exp. Psychol. Gen.* 150, 583
39. Jara-Ettinger, J. and Rubio-Fernandez, P. (2022) The social basis of referential communication: Speakers construct physical reference based on listeners' expected visual search. *Psychol. Rev.* 129, 1394
40. Jara-Ettinger, J. and Rubio-Fernandez, P. (2024) Demonstratives as attention tools: Evidence of mentalistic representations within language. *Proc. Natl. Acad. Sci.* 121, e2402068121
41. Taliaferro, M. and Schulz, L. (2024) Brown bear, brown bear, what do you see? speakers use more redundant color adjectives when speaking to children than adults. In *Proceedings of the Annual Meeting of the Cognitive Science Society*
42. Rubio-Fernández, P. (2016) How redundant are redundant color adjectives? an efficiency-based analysis of color overspecification. *Front. Psychol.* 7, 153
43. Rubio-Fernandez, P. *et al.* (2022) Adjective position and referential efficiency in american sign language: Effects of adjective semantics, sign type and age of sign exposure. *J. Mem. Lang.* 126, 104348
44. Coventry, K.R. *et al.* (2023) Spatial communication systems across languages reflect universal action constraints. *Nat. Hum. Behav.* 7, 2099–2110
45. Diessel, H. (2014) Demonstratives, frames of reference, and semantic universals of space. *Lang. Ling. Compass* 8, 116–132
46. Levinson, S.C. (2018) Introduction: demonstratives: patterns in diversity. In *Demonstratives in cross-linguistic perspective*, pp. 1–42, Cambridge University Press
47. Diessel, H. (2006) *Demonstratives, joint attention, and the emergence of grammar*.
48. Bangerter, A. (2004) Using pointing and describing to achieve joint focus of attention in dialogue. *Psychol. Sci.* 15, 415–419
49. Cooperrider, K. (2016) The co-organization of demonstratives and pointing gestures. *Discourse Process.* 53, 632–656
50. Piwek, P. *et al.* (2008) 'proximal'and 'distal'in language and cognition: evidence from deictic demonstratives in dutch. *J. Pragmat.* 40, 694–718
51. Çokal, D. *et al.* (2014) Deixis: This and that in written narrative discourse. *Discourse Process.* 51, 201–229
52. Arnold, J.E. (2010) How speakers refer: The role of accessibility. *Lang. Ling. Compass* 4, 187–203
53. Carston, R. (2008) *Thoughts and utterances: The pragmatics of explicit communication*, John Wiley & Sons
54. Rubio Fernandez, P. (2007) Suppression in metaphor interpretation: differences between meaning selection and meaning construction. *J. Semant.* 24, 345–371
55. Long, M. *et al.* (2023) The role of cognitive control and referential complexity on adults' choice of referring expressions: Testing and expanding the referential complexity scale. *J. Exp. Psychol.*, Learning, Memory, and Cognition
56. Zhang, C. *et al.* (2023) Goal recognition with timing information. *Proc. Int. Conf. Automat. Plan. Sched.* 33, 443–451
57. Zhang, C. *et al.* (2024) Human goal recognition as bayesian inference: Investigating the impact of actions, timing, and goal solvability. *arXiv*, preprint arXiv:2402.10510
58. Richardson, E. and Keil, F.C. (2022) Thinking takes time: Children use agents' response times to infer the source, quality, and complexity of their knowledge. *Cognition* 224, 105073
59. S. Bavard, E. Stuchlý, A. Konovalov, S. Gluth, Beyond choices: humans can infer social preferences from response times alone, 2023. URL: osf.io/preprints/psyarxiv/38yrw. doi: 10.31234/osf.io/38yrw.
60. Gates, V. *et al.* (2021) A rational model of people's inferences about others' preferences based on response times. *Cognition* 217, 104885
61. Berke, M. and Jara-Ettinger, J. () Thinking about thinking through inverse reasoning, in: Proceedings of the Annual Meeting of the Cognitive Science Society, 2021 https://doi.org/10.31234/osf.io/r25qn
62. M. Berke et al., Thinking about thinking as rational computation, in: Proceedings of the Annual Meeting of the Cognitive Science Society, 2023. URL: https://doi.org/10.31234/osf.io/e65p3.
63. Levinson, S.C. (2022) The interaction engine: cuteness selection and the evolution of the interactional base for language. *Philos. Trans. R. Soc. B* 377, 20210108
64. Evans, N. *et al.* (2018) The grammar of engagement i: Framework and initial exemplification. *Lang. Cogn.* 10, 110–140
65. Evans, N. *et al.* (2018) The grammar of engagement ii: Typology and diachrony. *Lang. Cogn.* 10, 141–170
66. Dingemanse, M. and Enfield, N. (2023) Interactive repair and the foundations of language. *Trends Cogn. Sci.*,
67. Holler, J. and Levinson, S.C. (2019) Multimodal language processing in human communication. *Trends Cogn. Sci.* 23, 639–652
68. Holler, J. (2022) Visual bodily signals as core devices for coordinating minds in interaction. *Philos. Trans. R. Soc. B* 377, 20210094
69. Glucksberg, S. (2003) The psycholinguistics of metaphor. *Trends Cogn. Sci.* 7, 92–96
70. Sperber, D. and Wilson, D. (2002) Pragmatics, modularity and mind-reading. *Mind Lang.* 17, 3–23
71. Moore, R. (2021) The cultural evolution of mind-modelling. *Synthese* 199, 1751–1776
72. Harris, D.W. (2025) Gricean communication, natural language, and human evolution. In *Evolutionary Pragmatics*, Oxford University Press, Oxford
73. Scott-Phillips, T. and Heintz, C. (2023) Animal communication in linguistic and cognitive perspective. *Annu. Rev. Linguist.* 9, 93–111
74. Gómez, J.C. (1994) Mutual awareness in primate communication: A gricean approach. *Self Aw. Animals Hum.* 61–80
75. Ferreira, V.S. (2019) A mechanistic framework for explaining audience design in language production. *Annu. Rev. Psychol.* 70, 29–51
76. Levelt, W.J. (1993) *Speaking: From intention to articulation*, MIT press
77. Bock, K. *et al.* (2002) *Language production: Grammatical encoding, Psycholinguistics: Critical concepts in psychology*. 5 pp. 405–452
78. Pickering, M.J. and Garrod, S. (2004) Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* 27, 169–190
79. Bavelas, J. *et al.* (2017) Doing mutual understanding. calibrating with micro-sequences in face-to-face dialogue. *J. Pragmat.* 121, 91–112
80. Dideriksen, C. *et al.* (2023) Language-specific constraints on conversation: Evidence from danish and norwegian. *Cogn. Sci.* 47, e13387
81. Dideriksen, C. *et al.* (2023) Quantifying the interplay of conversational devices in building mutual understanding. *J. Exp. Psychol. Gen.* 152, 864
82. De Stefani, E. (2021) Embodied responses to questions-in-progress: silent nods as affirmative answers. *Discourse Process.* 58, 353–371
83. Knudsen, B. *et al.* (2020) Forgotten little words: How backchannels and particles may facilitate speech planning in conversation? *Front. Psychol.* 11, 593671
84. Hömke, P. *et al.* (2018) Eye blinks are perceived as communicative signals in human face-to-face interaction. *PLoS One* 13, e0208030

85. Kidd, C. *et al.* (2011) Toddlers use speech disfluencies to predict speakers' referential intentions. *Dev. Sci.* 14, 925–934
86. Heller, D. *et al.* (2015) Inferring difficulty: Flexibility in the real-time processing of disfluency. *Lang. Speech* 58, 190–203
87. Loy, J.E. *et al.* (2017) Effects of disfluency in online interpretation of deception. *Cogn. Sci.* 41, 1434–1456
88. Fox Tree, J.E. (2002) Interpreting pauses and ums at turn exchanges. *Discourse Process.* 34, 37–55
89. Arnold, J.E. *et al.* (2004) The old and thee, uh, new: Disfluency and reference resolution. *Psychol. Sci.* 15, 578–582
90. Arnold, J.E. *et al.* (2007) If you say thee uh you are describing something hard: the on-line attribution of disfluency during reference comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 914
91. Watanabe, M. *et al.* (2008) Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Comm.* 50, 81–94
92. Auer, P. and Stukenbrock, A. (2022) Deictic reference in space. *Pragmat. Space* 23–62
93. 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
94. Nota, N. *et al.* (2022) *Conversational eyebrow frowns facilitate question identification: An online vr study*.
95. Nota, N. *et al.* (2023) Specific facial signals associate with categories of social actions conveyed through questions. *PLoS One* 18, e0288104
96. 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
97. Dingemanse, M. (2024) Interjections at the heart of language. *Annu. Rev. Linguist.* 10, 257–277
98. Hömke, P. *et al.* (2022) Eyebrow movements as signals of communicative problems in human face-to-face interaction. *PsyArxiv*,
99. Albert, S. and De Ruiter, J.P. (2018) Repair: the interface between interaction and cognition. *Top. Cogn. Sci.* 10, 279–313
100. Huang, Y.T. and Snedeker, J. (2009) Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cogn. Psychol.* 58, 376–415
101. Huang, Y.T. and Snedeker, J. (2011) Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Lang. Cogn. Process.* 26, 1161–1172
102. Bergqvist, H. and Knuchel, D. (2019) Explorations of engagement: Introduction. *Open Linguist.* 5, 650–665
103. Evans, N. (2022) *Words of wonder: Endangered languages and what they tell us*, John Wiley & Sons
104. Apperly, I.A. and Butterfill, S.A. (2009) Do humans have two systems to track beliefs and belief-like states? *Psychol. Rev.* 116, 953
105. Burger, L. and Jara-Ettinger, J. (2020) *Mental inference: Mind perception as bayesian model selection*, CogSci, in
106. Apperly, I. (2010) *Mindreaders: the cognitive basis of "theory of mind"*, Psychology Press
107. Kampis, D. and Southgate, V. (2020) Altercentric cognition: How others influence our cognitive processing. *Trends Cogn. Sci.* 24, 945–959
108. Woensdregt, M.S. *et al.* (2022) Language universals rely on social cognition: Computational models of the use of this and that to redirect the receiver's attention. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (vol. 44)
109. Paunov, A.M. *et al.* (2019) Functionally distinct language and theory of mind networks are synchronized at rest and during language comprehension. *J. Neurophysiol.* 121, 1244–1265
110. Deen, B. *et al.* (2015) Functional organization of social perception and cognition in the superior temporal sulcus. *Cereb. Cortex* 25, 4596–4609
111. Allison, T. *et al.* (2000) Social perception from visual cues: role of the sts region. *Trends Cogn. Sci.* 4, 267–278
112. Materna, S. *et al.* (2008) Dissociable roles of the superior temporal sulcus and the intraparietal sulcus in joint attention: a functional magnetic resonance imaging study. *J. Cogn. Neurosci.* 20, 108–119
113. Saxe, R. and Kanwisher, N. (2013) *People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind"*, Social neuroscience, Psychology Press, in, pp. 171–182
114. Jacoby, N. *et al.* (2016) Localizing pain matrix and theory of mind networks with both verbal and non-verbal stimuli. *Neuroimage* 126, 39–48
115. King, M. *et al.* (2019) Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nat. Neurosci.* 22, 1371–1378
116. Van Overwalle, F. *et al.* (2020) Consensus paper: cerebellum and social cognition. *Cerebellum* 19, 833–868
117. Leggio, M. and Molinari, M. (2015) Cerebellar sequencing: a trick for predicting the future. *Cerebellum* 14, 35–38
118. Wolpert, D.M. *et al.* (1998) Internal models in the cerebellum. *Trends Cogn. Sci.* 2, 338–347
119. Ito, M. (2008) Control of mental activities by internal models in the cerebellum. *Nat. Rev. Neurosci.* 9, 304–313