

Layers of Social Meaning: Innate, Acquired, and Deliberative Levels of Referential Signal Processing

Fang Yang,^{1,2*} Ningfeng Liu,^{3,4,5*} Yi Jiang,^{1,2} and Lusha Zhu^{3,4,5}

¹State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China, ²Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China, ³School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China, ⁴IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China, and ⁵Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

Referential signaling is a fundamental component of social interaction, enabling individuals to direct attention, convey intention, and establish shared understanding. The processing of referential signals operates across multiple levels of neurocognitive systems. First, innate mechanisms rapidly detect socially relevant cues, such as gaze and biological motion, prioritizing social information from infancy. Second, social learning enables the establishment of symbolic conventions in service of referential interactions, implicating the alignment of abstract representations across brains. Third, a deliberative system supports the flexible production and interpretation of referential signals in context by anticipating others' beliefs and motivation. Despite the diversity and complexity of these processes, recent research has made significant strides in identifying their neural and cognitive bases. Studies indicate that evolutionarily conserved subcortical pathways support early-emerging sensitivity, while cortical networks involved in social reasoning and decision-making underlie acquired and deliberative levels of processing. Future research is needed to clarify how these mechanisms interact across development and how their disruption contributes to neurodevelopmental conditions such as autism spectrum disorder.

Key words: communication; developmental; human; neuroimaging; perception; social decision-making; social neuroscience; social reasoning

Significance Statement

Referential signaling is essential to social interaction, allowing individuals to direct attention to specific objects, events, or concepts with minimal communicative effort. Understanding how the brain produces and interprets these signals provides key insights into social information processing. While research on this topic has often focused separately on either lower-level perception or higher-level decision-making, emerging evidence suggests multilevel processing. Here, we propose a framework categorizing referential signal processing into three broad levels: innate, acquired, and deliberative. This framework helps understand how referential signals emerge, evolve, and function across contexts, integrating perspectives from perception to communication and strategic decision-making.

Introduction

Effective social interaction relies on the ability to perceive, interpret, and respond to a diverse array of social signals, including facial expressions, gaze direction, bodily gestures, and vocal expressions. These signals convey essential information about others' emotions, beliefs, and intentions, shaping adaptive social behaviors.

Understanding mechanisms that support social signal processing is a central theme in social neuroscience (Gangopadhyay et al., 2021). Much of the research has approached this question from disparate perspectives, focusing either on lower-level perceptual processes or higher-level functions such as learning and decision-making (Schurz et al., 2014; Schilbach, 2015; Wang et al., 2024; Shen et al., 2025). However, emerging evidence suggests that social signal processing is not a unitary process but operates across multiple levels of cognition and neural implementation. At the one end of the spectrum, it involves automatic, subcortical mechanisms that prioritize socially salient stimuli. At the other end, it involves cortical processing that enables flexible, goal-directed signal production and interpretation. The interplay between bottom-up, sensory-driven computations and top-down, predictive processes is thought to support interpersonal interactions by facilitating the

Received March 17, 2025; revised July 1, 2025; accepted July 10, 2025.

Author contributions: F.Y., N.L., Y.J., and L.Z. wrote the paper.

This research was supported by STI2030-Major Projects (2022ZD0205100 to L.Z., 2021ZD0203800 to Y.J.) and National Natural Science Foundation of China (32325023, T2421004, 32441104 to L.Z., 32430043 to Y.J.).

*F.Y. and N.L. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Yi Jiang at yijiang@psych.ac.cn or Lusha Zhu at lushazhu@pku.edu.cn.

<https://doi.org/10.1523/JNEUROSCI.0535-25.2025>

Copyright © 2025 the authors

selective filtering and prioritizing of socially relevant cues (Friston and Frith, 2015b; Saxe and Houlihan, 2017). Bridging insights across these levels of processing is therefore important for understanding the neurocognitive architecture underlying social signal processing.

Referential signals provide a valuable probe into this broader question. As a cornerstone of social interaction, referential signals direct others' attention to specific entities, events, or concepts in the environment, often at reduced communicative costs (Palazzolo, 2024). These signals—ranging from biologically rooted cues such as eye gaze and body movements to culturally established symbols and expressions—convey intention, evoke shared representations, and facilitate information exchange. From early infancy, humans exhibit remarkable sensitivity to signals with referential function, responding to the caregiver's gaze and gestures in ways that suggest an innate predisposition for detecting referential intent (Farroni et al., 2004; Simion et al., 2008). This early-emerging sensitivity is believed to scaffold the development of more sophisticated communicative abilities and broader social functions (Astor and Gredebäck, 2022; Yang et al., 2024). Beyond the early-emerging mechanisms, referential signals can be acquired through social learning or established via interactive negotiation: For example, symbolic gestures gain their meanings through processes involving coordination, imitation, and shared context-building (Brignani et al., 2009; Stolk et al., 2013; Chacón-Candia et al., 2022; Hawkins et al., 2023). Beyond learning-guided processing, referential signals can also be flexibly generated through deliberative reasoning and perspective-taking on

the fly, as seen in strategic communicative scenarios where individuals adjust their signal use by anticipating partners' encoding and decoding processes (Mi et al., 2021; Vélez et al., 2023).

Given the functional diversity of referential signals, their processing can be broadly categorized as the innate, acquired, and deliberative levels of processing (Fig. 1). While these categories do not represent strict divisions, as social cognition is inherently complex and dynamic, they provide a framework for understanding how referential signals emerge, evolve, and are flexibly deployed, in intentional and unintentional interacting contexts. This perspective helps integrate ideas and findings from research on perception, communication, and strategic decision-making, drawing from developmental, computational, and neuroscientific approaches. Rather than to provide a comprehensive survey, the goal of this review is to give an illustrative overview of key topics, methods, and findings related to referential signal processing, to identify open questions in the field, and to outline potential directions for future research.

In the remainder of this paper, we first review evidence about the innate responses to referential signals, focusing on the processing of evolutionarily significant cues such as eye gaze and biological motion (BM) particularly in infants. Next, we compare these processes with responses acquired through social learning and adaptation, focusing on the neurocognitive processing of culturally established symbols and interaction-driven conventions. Finally, we explore recent findings on the deliberative processing of referential signals, focusing on how the brain encodes intention into referential signals and decodes others' intentions from received signals.

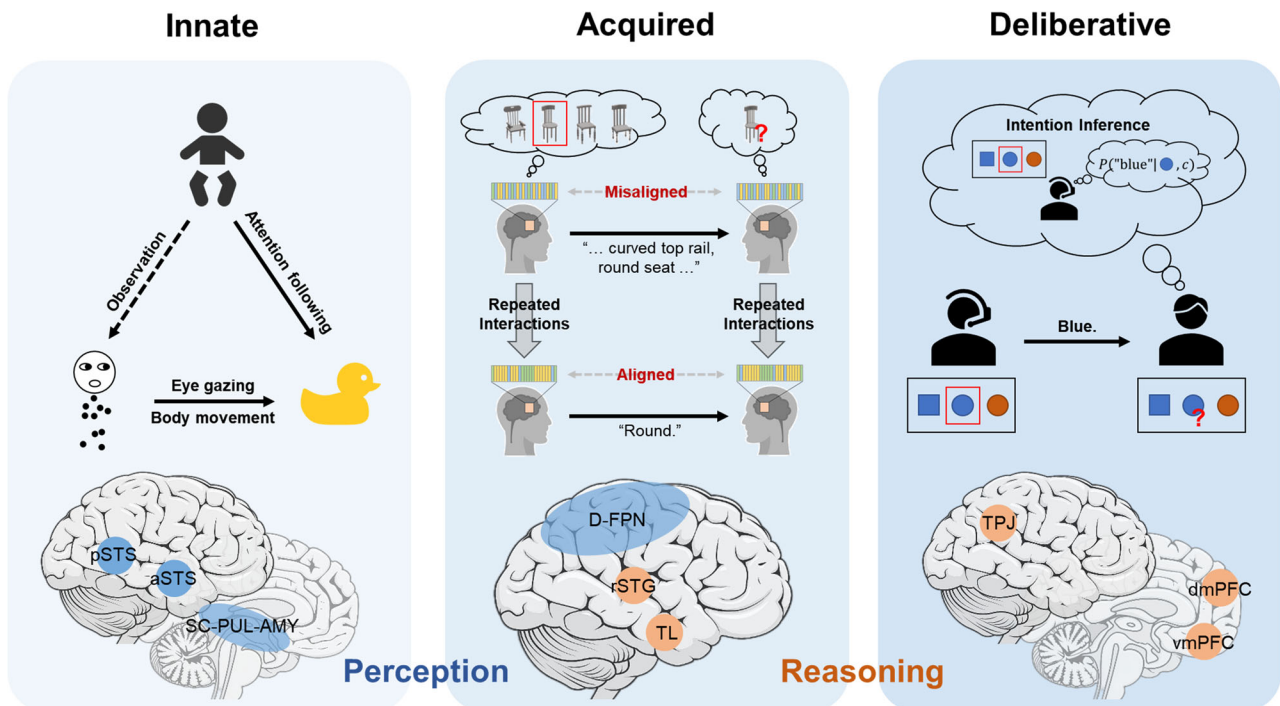


Figure 1. Multilevel processing in referential interactions. Referential signal processing engages distinct behavioral and neural processing depending on whether signals are interpreted through innate sensitivity, acquired association, or deliberative reasoning. Left, A subcortical pathway from SC through pulvinar (PUL) to amygdala (AMG) supports the automatic detection of biologically salient cues, such as eye gaze and BM. The extraction of directional information from these cues and the corresponding orienting of attention involve both the anterior and posterior portions of the STS. Middle, Referential meanings acquired from social learning and adaptation. In the illustrated repeated referential game, a sender describes a target (e.g., a chair) with a referential message, based on which the receiver needs to select the target from several alternatives. Without preestablished labels, partners coconstruct shared names over time, improving communicative efficiency. This coordinated abstraction may involve increasingly aligned activation of abstract representations across brains (color bars) in regions including the dorsal frontoparietal network (D-FPN), rSTG, and TL. Right, Deliberative referential reasoning supports flexible, context-sensitive processing when preestablished meanings are absent or insufficient. In the illustrated one-shot referential communication, a sender selects one of several constrained signals to indicate a target, and the receiver must infer the intended meaning by reasoning about the sender's likely choice, without any mutual adaptation. This requires explicit cognitive efforts to anticipate others' perspectives and model their inferential process, which recruit regions such as the TPJ, dorsomedial prefrontal cortex (dmPFC), and vmPFC.

Innate responses: biological scaffold of referential behavior

Human infants are born with a remarkable ability to detect and respond to biological signals, forming the foundation of early referential processing. From birth, they are immersed in a continuous stream of social cues—such as a caregiver’s gaze, stride, or body orientation—that convey referential intent and guide attention to socially relevant information. Remarkably, newborns can detect and respond to these signals within hours of birth, reflecting an innate sensitivity to social referential cues. These early responses enable infants to identify conspecifics in complex environments and facilitate the automatic interpretation of communicative intent. Such abilities develop along canalized trajectories (Waddington, 1942), stabilized by genetic constraints and supported by neural circuits shaped by evolutionary pressures (Johnson et al., 2015). Acting as cognitive bootstraps, these innate responses channel attention toward ecologically salient signals, fostering the emergence of sophisticated referential behaviors.

Prewired sensitivity to biological referential signals

The origins of humans’ capacity to process biological signals, whether rooted in innate neural mechanisms or shaped by post-natal experience, remain a topic of ongoing debate (Grossmann, 2015). Nonetheless, accumulating evidence suggests that the human brain is prewired with neural modules specialized for

efficiently detecting and processing biologically relevant cues, such as eye gaze and BM (Baron-Cohen, 1995; Hirai and Senju, 2020; Babinet et al., 2022). For example, newborns exhibit a spontaneous preference for faces with direct gaze and stimuli displaying BM patterns, both signaling potential social agents in the environment (Farroni et al., 2002; Simion et al., 2008; Fig. 2A). These preferences rely on specific physical features, such as the low spatial frequency configuration of facial features (typically arranged in an inverted triangular pattern; Kobylkov and Vallortigara, 2024), the high contrast between the sclera and iris (Farroni et al., 2005), and the natural acceleration dynamics of BM adhering to gravitational constraints (Troje and Westhoff, 2006). Disrupting these features eliminates preferential responses, suggesting that these perceptual biases are finely tuned to the statistical regularities of natural environments (Slater et al., 2000; Simion et al., 2008; Wang et al., 2022). Such tuning likely aids in forming prototypical templates for detecting and interpreting biological referential signals in noisy sensory environments.

Behavioral genetics research further supports the genetic basis underlying the preference for biological referential cues. Monozygotic twins demonstrate greater concordance in attention to the eye region during social interactions than dizygotic twins, even after accounting for shared environmental factors (Constantino et al., 2017; Portugal et al., 2023). Similarly, monozygotic twins exhibit stronger similarities in processing local BM patterns, a perceptual ability that is also correlated with

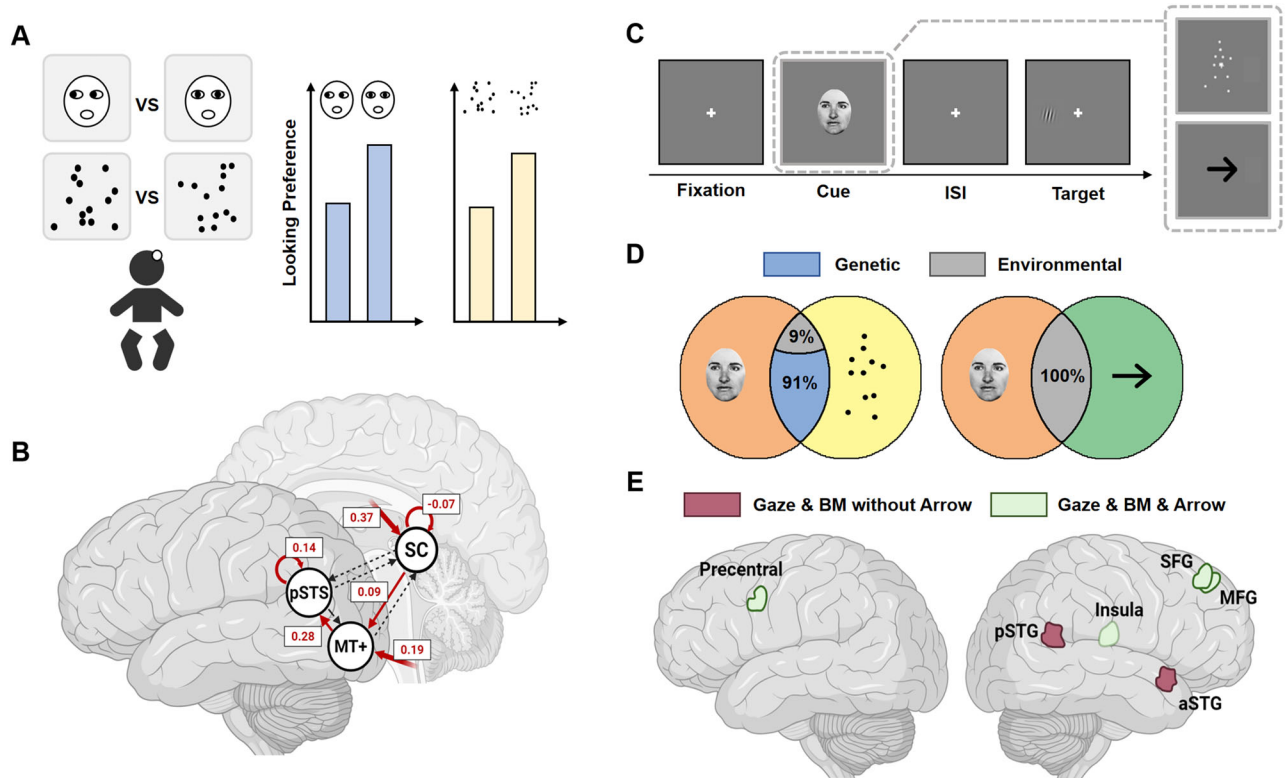


Figure 2. Experimental paradigms and key findings on innate referential responses. **A**, Human neonates exhibit a preference for faces with direct gaze (Farroni et al., 2002) and for upright, intact BM displays (Simion et al., 2008), indicating innate sensitivity to socially relevant cues. **B**, Recent evidence reveals a subcortical–cortical pathway involved in processing BM, specifically from SC through area MT+ to the pSTS; red arrows indicate significant inversion effect (i.e., stronger response to upright vs inverted BM stimuli) on modulatory and driving input parameters in dDCM; dashed black arrows indicate no significance (Lu et al., 2024). **C**, In a classic gaze-cueing task, participants view a face with averted gaze followed by a target appearing on either the cued or uncued side; faster reaction times and higher accuracy for cued targets indicate attention is guided by gaze—a pattern also observed with BM and arrow cues. **D**, A twin study shows that both genetic and environmental factors contribute to cue-elicited attention effects: shared genetics account for 91% of the covariance between gaze and BM induced orienting, whereas the association between gaze and arrow cues is entirely explained by environmental factors (Wang et al., 2020). **E**, A recent neuroimaging study shows that the posterior and anterior STG (pSTG, aSTG) can decode the direction of social attention from both gaze and BM cues, but not from arrow cues (Wang et al., 2024). SFG, superior frontal gyrus; MFG, middle frontal gyrus.

individual autistic traits (Wang et al., 2018). Together, these findings suggest that visual engagement with and perceptual processing of biological referential cues may constitute a neurodevelopmental endophenotype, potentially shaping postnatal social interactions and the interpretation of social information.

Cross-species studies reveal that the preference for biological signals is mediated by conserved mechanisms. Visually naive chicks exhibit spontaneous preferences for face-like configurations and BM (Salva et al., 2015). Similar early preferences are observed in nonhuman primates: adult macaques fixate faster and longer on faces with direct gaze (Leonard et al., 2012), and infant chimpanzees and macaques preferentially attend to faces with open eyes and direct gaze over those with closed or averted gaze (Myowa-Yamakoshi et al., 2003; Hirata et al., 2010; Muschinski et al., 2016). Interestingly, nursery-reared macaques, despite limited exposure to conspecific faces, retain a baseline attraction to direct gaze but lack the gaze specialization seen in socially reared peers (Simpson et al., 2019). This suggests an early-developing sensitivity to mutual gaze, shaped by both genetic predispositions and experience-dependent processes. Recent studies have further demonstrated that preferences for biological cues extend to aquatic species, such as zebrafish (Ma et al., 2022) and oviparous reptiles, like tortoises reared without parental care (Versace et al., 2020). These findings highlight the evolutionary conservation of these mechanisms.

Recent years have seen significant progress in understanding the neural basis of innate biological preferences. Studies on visually naive chicks have demonstrated that the optic tectum—homologous to the mammalian superior colliculus (SC)—acts as a critical node in the innate processing of biological signals, such as faces and BM (Salva et al., 2015; Kobylkov et al., 2024; Lorenzi et al., 2024). Complementing these findings, a recent cross-species study using ultrahigh-field functional magnetic resonance imaging (fMRI) revealed that the SC, particularly its superficial layers, selectively responds to local BM signals (Lu et al., 2024). Dynamic causal modeling (DCM) analysis further identified a subcortical–cortical pathway transmitting these BM signals from the SC to higher-order areas (Fig. 2B), including the middle temporal visual complex (MT+) and the posterior superior temporal sulcus (pSTS).

Collectively, these findings suggest that innate sensitivity to biological signals is supported by evolutionarily conserved subcortical structures, particularly the SC, which functions as a key node in the subcortical visual processing pathway. Retinal fibers project via two primary routes: the geniculostriate pathway, where axons reach the lateral geniculate nucleus (LGN) and subsequently the primary visual cortex (Ling et al., 2015), and the extrageniculate pathway, where axons first connect to the SC, thalamic pulvinar, and amygdala before reaching extrastriate parietal and temporal regions (Knudsen, 2018). While both pathways operate in parallel, the extrageniculate route enables faster processing of dynamic biological cues (Mizzi and Michael, 2016). In this pathway, retinal inputs are relayed to the SC, which integrates spatial and salient features such as prototypical face templates (Le et al., 2020), before transmitting this information to the pulvinar and amygdala (Lyon et al., 2010; Mizzi and Michael, 2016). This integration of biological signals through evolutionarily conserved subcortical mechanisms provides the neural substrate for the rapid processing of information essential for adaptive interactions (Hirai and Senju, 2020; Babinet et al., 2022).

Automatic attentional orienting directed by referential cues

Merely detecting biologically salient signals, such as fixating on another's eyes, is insufficient to elicit referential behaviors

unless the communicative intent embedded in gaze dynamics (e.g., subtle shifts in gaze direction) is extracted. Even when biological cues are perceptually registered, failing to interpret their referential significance hinders their effective use in social interactions. Converging evidence suggests that humans exhibit a remarkably automatic tendency to align their attention with the direction of biological cues from early infancy, a phenomenon known as social attention (Birmingham and Kingstone, 2009). For instance, primitive gaze-following behavior, observed in infants as young as a few days old, reflects an innate capacity to shift visual attention in response to others' gaze direction (Farroni et al., 2004). This early-developing ability facilitates the use of biological referential signals, enabling attention orientation and enhancing the interpretation of others' intentions. Therefore, we propose that this automatic mechanism serves as a foundational scaffold supporting the development of more complex shared referential systems.

Studies employing the gaze-cueing paradigm have investigated the characteristics of social attention by presenting centrally positioned cues indicating gaze or walking direction (Fig. 2C). Extensive research shows that attention shifts triggered by biological cues are highly reflexive, emerging as rapidly as 100 ms after cue onset and persisting even when the cue is nonpredictive or counterpredictive of the target's location (Friesen and Kingstone, 1998; Driver et al., 1999; Friesen et al., 2004; Shi et al., 2010; Wang et al., 2014; Liu et al., 2021). Furthermore, studies employing unconscious paradigms reveal that gaze cues trigger attention shifts even in the absence of visual awareness (Sato et al., 2007; Xu et al., 2011; Yang et al., 2024), suggesting that social attention operates automatically and independently of top-down cognitive control. These findings, together with evidence from neonatal studies, suggest that social attention is underpinned by innate mechanisms. Baron-Cohen (1995) proposed the existence of a "shared attention module" (SAM), which enables individuals to align their attention with that of social partners—a fundamental ability for the development of theory of mind. However, social attention is also modulated by social factors (Dalmaso, 2014) and experiential learning (Senju et al., 2015). Consequently, more integrative perspectives propose that, akin to the Conspic and Conlern theory of face perception (Morton and Johnson, 1991), social attention emerges from the interplay between innate predispositions and experience-dependent learning (Astor and Gredebäck, 2022).

Despite differences in physical characteristics, attentional shifts elicited by gaze and BM share key properties of automaticity and reflexivity. Recent studies suggest that attention guided by these biological cues may rely on shared neural and genetic mechanisms. Using an adaptation paradigm, one study demonstrates that prolonged exposure to averted gaze modulates attentional effects triggered by BM and vice versa, revealing a robust cross-category adaptation effect (Ji et al., 2020). This attenuation indicates a shared neural coding mechanism for attentional orienting triggered by both gaze and BM cues. Furthermore, genetic-behavioral studies show that social attention effects induced by eye gaze and BM exhibit higher concordance in monozygotic twins (Fig. 2D), suggesting a common genetic basis for these processes (Wang et al., 2020). Importantly, control experiments using nonbiological directional cues, such as arrows, confirm that these adaptation effects and genetic influences are specific to social stimuli. These findings suggest the existence of a specialized attentional orienting mechanism for biological signals.

Cross-species studies on social attention highlight its deep evolutionary origins. Nonhuman primates, such as chimpanzees and macaques, exhibit gaze-following behaviors, relying more on head

orientation than eye gaze due to morphological differences (Tomasello et al., 2007; Rosati et al., 2016). Beyond mammals, newly hatched chicks reared in darkness instinctively align their movements with observed BM trajectories, indicating an innate social attention ability (Vallortigara et al., 2005). Rudimentary forms of social attention behaviors have also been observed in reptiles (Zeiträg et al., 2023) and archerfish (Leadner et al., 2021). However, more advanced social attention abilities appear to emerge at specific evolutionary stages, with only certain avian species demonstrating complex attention-following behaviors (Zeiträg et al., 2023). These findings suggest that social attention is a widely conserved ability across taxa, with deep evolutionary roots in sociocognitive processing (Zeiträg et al., 2022).

Neuroimaging studies have yielded mixed findings regarding the neural mechanisms underlying social attention. Some studies suggest that biological cues (e.g., gaze direction) engage specialized attention mechanisms (Kingstone et al., 2004; Engell et al., 2010), while others argue that these cues rely on mechanisms similar to those activated by general, nonsocial directional signals, like arrows (Tipper et al., 2008; Sato et al., 2009; Uono et al., 2014). A recent meta-analysis systematically compared neural activation patterns associated with gaze- and arrow-triggered attention (Salera et al., 2024). The results indicated that gaze-triggered attention more strongly recruits brain regions involved in reflexive attention (medial frontal gyrus), biological processing (STS), and mental state attribution (temporoparietal junction, TPJ). Supporting evidence comes from an fMRI study applying multivariate analysis to examine whether attention guided by BM, gaze, and arrows shares common neural substrates (Wang et al., 2024). The findings identified the right pSTS and anterior STS (aSTS) as key hubs encoding attentional direction from both gaze and BM cues, but not from nonsocial cues, providing robust evidence for a specialized social attention module (Fig. 2E).

The pSTS, functionally connected to the TPJ and parietal lobe, plays a crucial role in aligning self and others' attention, particularly in response to dynamic gaze shifts in humans (Nummenmaa et al., 2010; Ramsey et al., 2011) and macaques (Shepherd et al., 2009). Neuroimaging studies in monkeys have identified a specialized subregion within the pSTS, known as the gaze-following patch (GFP), that supports gaze-following behavior (Marciniak et al., 2014). Neurons in the GFP flexibly link gaze direction to target objects according to the social context (Ramezanpour and Thier, 2020). Further evidence, through electrical and pharmacological perturbation, demonstrates that disrupting this region selectively impairs gaze-following performance, highlighting the essential role of the GFP in social attention (Chong et al., 2023). Conversely, the aSTS is specialized for encoding eye gaze direction independently of facial orientation (Carlin et al., 2011, 2012). Notably, the connectivity between the aSTS and emotion-related regions such as the amygdala (Pitcher et al., 2017) suggests its potential role in integrating social attention with emotional regulation. Collectively, these findings highlight a specialized cortical network underlying social attention, with the right STS and superior temporal gyrus (STG) serving as key neural loci. Further research is needed to clarify how these regions interact with subcortical structures, general attention networks, and mentalizing systems to support complex referential behaviors.

Acquired meaning: socially established signals and adaptive referential interactions

Unlike eye gaze or BM, which elicit responses likely shaped by evolutionarily conserved neural mechanisms, a significant class

of referential signals are developed or acquired through social interactions. These signals can be broadly categorized into two forms: (1) Stable signal-referent associations, such as postnatally established symbols, whose meanings are typically acquired by individuals through learning, and (2) flexible signal-referent mappings, which emerge dynamically within social groups and evolve through ongoing interactions. While behavioral and neural research has examined these referential signals largely in separation, the extent to which they rely on separable or overlapping neural systems remains unclear. Further investigation is needed to delineate the neurocognitive mechanisms underlying referential adaptation and their role in shaping social behavior.

Postnatally established meaning and learned automaticity

While biological cues such as gaze and BM play a fundamental role in guiding social attention, referential responses are not restricted to these processes. Humans have also developed culturally established symbolic systems, such as language and traffic signals, to convey referential information. Among these, arrows exemplify abstract symbols that effectively guide attention. Although arrows lack evolutionary origins, their clear spatial meanings are widely recognized across linguistic and cultural contexts. From traffic navigation to digital interfaces, their extensive use demonstrates how learned symbolic associations can consistently achieve functional equivalence to biological cues. This capacity highlights the flexibility of the human cognitive system in attributing intentionality to both evolutionarily ancient signals and culturally constructed symbols.

Initially, arrow-induced attentional orienting was thought to be a purely voluntary process requiring explicit task instructions. Jonides (1981) demonstrated that centrally presented predictive arrows could direct attention endogenously, with effects disappearing when the predictive validity was removed. This finding aligned with the classical view of arrows as symbolic cues requiring top-down referential processing. However, accumulating behavioral evidence challenges this notion. Studies now show that even nonpredictive arrows can orient attention, resembling the automatic effects observed with gaze and BM cues (Hommel et al., 2001; Ristic et al., 2002, 2007; Tipples, 2002). These findings led to the proposal that arrows and biological cues may share a general, indistinguishable mechanism, whereby directional signals become associated with intentionality through repeated exposure and automatic learning (Tipples, 2002).

Despite behavioral similarities, psychophysical studies utilizing cross-category adaptation and behavioral genetic approaches challenge the idea that arrows and biological cues share an indistinguishable underlying mechanism. Arrow-induced attentional effects remain unaffected by adaptation to directional information from gaze or BM (Ji et al., 2020). Furthermore, while arrows and gaze cues exhibit overlapping environmental influences, their genetic heritability appears largely distinct (Wang et al., 2020). These findings suggest that, although arrows can elicit attention shifts akin to biological cues, their underlying mechanisms differ fundamentally. Developmental studies further support this distinction: younger children (<5 years) tend to rely more on perceptual features of arrows, whereas older children process their symbolic meaning (Jakobsen et al., 2013). This developmental trajectory contrasts sharply with gaze cues, which reflexively orient attention even in newborns (Farroni et al., 2004). These findings suggest that arrow-triggered attention arises not from evolutionary hardwiring but through extensive exposure and learned associations in cultural contexts (Brignani et al., 2009; Chacón-Candia et al., 2022).

Converging evidence from neuropsychological and neuroimaging studies underscores the dissociation between arrow- and biologically driven attention. Lesion studies reveal that damage to the rSTG and amygdala selectively impairs gaze-triggered attention while sparing arrow-induced orienting (Akiyama et al., 2006, 2007, 2008). Supporting this dissociation, neuroimaging studies indicate that, while arrows can evoke reflexive attentional shifts, they predominantly engage the dorsal attention network (Callejas et al., 2014), including the medial temporal gyrus, inferior parietal lobule, precuneus, and frontal eye fields (Salera et al., 2024). These findings align with behavioral evidence suggesting that arrow cueing relies more on learned, goal-directed systems than on reflexive social attention mechanisms (Liu et al., 2021; Chacón-Candia et al., 2022). Event-related potential studies further reveal distinct temporal dynamics: gaze cues elicit early occipital components (N170) linked to face perception (Brignani et al., 2009), while arrows evoke a later parietal positivity (P300) associated with conceptual analysis and response preparation (Hietanen et al., 2008). This temporal divergence reinforces the view that arrows engage higher-order cognitive systems distinct from the innate mechanisms driving gaze-triggered attention. Research on arrow-triggered attention offers an alternative framework for understanding the mechanisms underlying referential orienting. This mechanism is believed to depend on extensive learning of associations between directional cues and meaningful outcomes. Through this process, arrow cues acquire the ability to guide attention in a largely automatic manner, thereby enhancing the efficient interpretation of symbolic referential signals.

Interactively constructed meaning and coordinated abstraction

Many referential behaviors are not merely the transmission of preestablished meanings but rather dynamic processes through which meaning is actively negotiated between interacting brains. Examples include task-specific terminology shared among coworkers and personalized nicknames created within specific social circles. Such referential conventions emerge from interactive negotiation within a group of individuals, shaped by social goals and contexts, refined by partner-specific feedback and mutual adaptation (Garrod and Doherty, 1994; Camerer, 2003; Efferson et al., 2008; Franke and Wagner, 2014; Goodman and Frank, 2016; Lazaridou et al., 2017; Hawkins et al., 2019). The ability to develop such referential conventions is a hallmark of human social intelligence, enabling efficient collaboration, learning, and innovation. Beyond its functional significance, this process provides a unique window into the cognitive and neural mechanisms of adaptation that gives rise to shared social meaning. How do the brains gradually agree on shared meaning? What neural processes allow interactants to encode, decode, and update referential signals over time? How are communicative conventions represented in the brain?

A widely used experimental approach to studying referential adaptation is the repeated referential game, in which communicators must develop shared conventions through iterative interactions (Lewis, 1969). Across variations of this paradigm, a sender needs to generate a signal to refer to a target, and a receiver needs to identify the correct referent from a set of alternatives based on the received signal. Crucially, the absence of preestablished or conventionalized signals for distinguishing the target forces participants to develop a shared naming system through repeated referential interactions, during which the sender can observe the receiver's choice and the receiver can

track how the sender's signals evolve over time (Fig. 3A; Krauss and Weinheimer, 1964; Clark and Wilkes-Gibbs, 1986; Hupet and Chantraine, 1992; Brennan and Clark, 1996; Weber and Camerer, 2003; Feiler and Camerer, 2010; Stolk et al., 2013, 2014; Hawkins et al., 2020b, 2023; Boyce et al., 2024). A robust body of research shows that these interactions lead to increasing communicative efficiency—that is, receivers identify targets more quickly and accurately over time, while misunderstandings gradually diminish (Krauss and Weinheimer, 1964; Clark and Wilkes-Gibbs, 1986; Weber and Camerer, 2003; Feiler and Camerer, 2010; Stolk et al., 2014; Hawkins et al., 2023; Boyce et al., 2024).

A key behavioral feature of referential adaptation is that referential signals within each dyad tend to converge and become progressively simplified over repeated interactions. This has been observed across diverse communicative modalities, including sketches, verbal descriptions, gestures, and object movements (Fig. 3A,B; Krauss and Weinheimer, 1964; Clark and Wilkes-Gibbs, 1986; Hupet and Chantraine, 1992; Garrod et al., 2007; Feiler and Camerer, 2010; Hawkins et al., 2023; Boyce et al., 2024). Despite this systematic reduction in complexity, converging evidence shows that these conventions remain highly variable across dyads, with different pairs developing distinct referring strategies for the same targets (Fig. 3B; Hawkins et al., 2020b, 2023; Boyce et al., 2024). This suggests that referential adaptation is shaped not only by general communicative constraints but also by the specific interaction histories of communicators, reflecting a socially grounded flexibility in referential adaptation.

Intriguingly, the progressive simplification of referential signals within dyads does not appear to be a random elimination of signal complexity but rather a process of structured abstraction (Carroll, 1980; Brennan and Clark, 1996; Hawkins et al., 2020b, 2023). In some experiments, communicators selectively preserved features that effectively distinguished the target from alternatives, while filtering out elements that carried less discriminative value (Fig. 3A; Hawkins et al., 2020b, 2023). This observation implicates that the overt signal simplification likely reflects a covert process of goal-directed abstraction, in which the brain evaluates candidate referents, extracts relational structures among candidates, and encodes task-relevant distinctions into referential signals to enhance communicative efficiency while maintaining precision.

This interpretation aligns with recent insights from cognitive neuroscience into the neural bases of task-specific abstraction in nonsocial contexts. Studies in various individual inferential and decision-making tasks have identified abstracted representations in the medial temporal lobe (TL) and medial prefrontal cortex (mPFC) that capture core relational structures of task problems and disregard irrelevant details (Ho et al., 2019; Konidaris, 2019; Peer et al., 2021; Vaidya and Badre, 2022; De Martino and Cortese, 2023). In referential adaptation, however, abstraction may not be solely an individual cognitive process but rather a socially coordinated one (Brennan and Clark, 1996; Metzinger, 2003; Ibarra and Tanenhaus, 2016; Pozzi et al., 2024). While individuals may initially construct distinct task representations, repeated interaction likely facilitate the alignment of these abstractions across brains, giving rise to the convergence to a common conceptualization in support of mutual understanding (Stephens et al., 2010; Hasson et al., 2012).

Neuroimaging evidence offers initial evidence for the hypothesis. A magnetoencephalography study compared communicative interactions, where participants had to generate and interpret novel referential signals, with instrumental interactions,

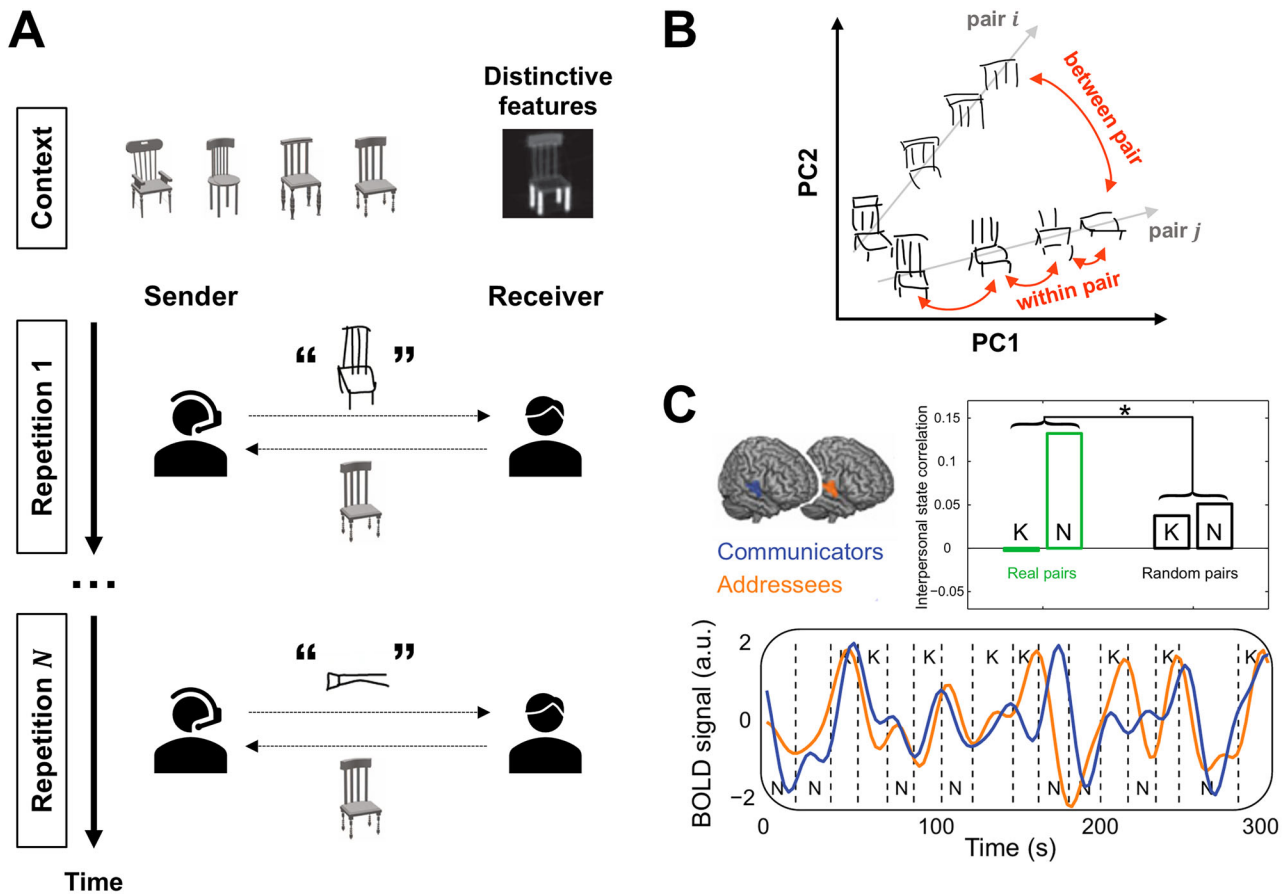


Figure 3. Experience-driven alignment in referential communication. **A, B**, In a repeated referential communication task, a sender produces hand-drawn sketches to indicate a target object among similar distractors. The receiver selects the intended referent based on the sketch. With repeated interactions, signals within each sender–receiver pair become increasingly simplified while retaining distinctive features necessary for target discrimination. Across pairs, referential signals diverge, reflecting idiosyncratic conventions (Hawkins et al., 2023). **C**, Top, Neural activity in the rSTG and other regions was more tightly coupled between two individuals when they formed an actual communicative pair (green bars), compared with randomly shuffled pairings (black bars). This coupling was especially pronounced when communicating about entities that lacked preestablished signal–meaning associations (indicated by N), comparing to those with established meaning (indicated by K), suggesting that mutual adaptation under novel communicative demands fosters dynamic alignment of brain states. Bottom, fMRI time courses from an example pair of participants engage in referential communication. Dashed lines indicate the timing of communicative episodes, with “K” and “N” marking trials involving problems with or without preestablished meaning, respectively (Stolk et al., 2014). **A, B**, Adapted from Hawkins et al. (2023), under the permission of a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>). **C**, Adapted from Stolk et al. (2014).

which required coordinated but noncommunicative actions (Stolk et al., 2013). The data revealed that both senders and receivers in referential interactions exhibited similar spectral changes in neural activity, localized to the right TL and ventromedial prefrontal cortex (vmPFC), previously found implicated in processing semantic information and flexibly decision-making, respectively (Sabbagh, 1999; Milne and Grafman, 2001; Lambon Ralph et al., 2010). These neural patterns appeared before the production of specific communicative signals and persisted throughout the interaction rather than being tied to discrete communicative cues or actions. This sustained neural pattern was interpreted as the evidence for high-level, conceptual representations of communicators, beyond sensorimotor processing of communicative events. Further supporting the notion that the internal representations are socially aligned, an fMRI hyperscanning study examined power spectral densities of BOLD signals and found that senders and receivers exhibited synchronized neural activity in the rSTG—a region previously found implicated in linguistic event processing, mental simulation, and human-agent motion recognition (Han et al., 2013; Bašnáková et al., 2014; Nummenmaa et al., 2014; Silbert et al., 2014; Stolk et al., 2014)—when processing novel referential

signals compared with preestablished ones (Fig. 3C; Stolk et al., 2014). Such intersubject synchronization was not related to transient communicative signals and was only observed within pairs with previous communicative history.

These findings provide evidence for the possibility that high-level cortical regions such as TL and vmPFC may be involved in representing task-related abstractions during referential adaptation, and activity in the rSTG may be involved in the alignment of such abstractions across brains. Recent studies of communication in naturalistic settings like lectures and classroom interactions show that interbrain synchronization between signal senders and receivers predicts communicative success (Stephens et al., 2010; Hasson et al., 2012; Jiang et al., 2015; Sievers et al., 2024). This raises an intriguing question about the extent to which internal abstractions observed in referential adaptation can be extended to complex, real-world communication, where abstract representations are needed to reflect information structure for effective delivery. Future studies are needed to elucidate the neural and computational mechanisms underlying how dynamic inter-brain coordination supports aligned abstractions. A growing body of social neuroscience research has applied formal models of behavior to characterize the neurocomputational mechanisms

of adaptive social behaviors like trust, reciprocity, and competition (King-Casas et al., 2005; Zhu et al., 2012, 2019; Park et al., 2019; Rusch et al., 2020; Jiang et al., 2022, 2023). Separately, significant progress has been made in cognitive neuroscience in investigating the neural instantiation of abstraction in various personal, nonsocial tasks (Behrens et al., 2018; Peer et al., 2021; Vaidya and Badre, 2022; De Martino and Cortese, 2023; Mishchanchuk et al., 2024). Integrating these approaches, future research could investigate how abstraction operates in referential communication and, crucially, how social feedback dynamically shapes the alignment of these abstractions across interacting brains, providing mechanistic insights into the neural bases of social signal generation and interpretation.

Deliberative inferences: strategic modeling of others' minds in referential signal processing

While adaptive referential behavior captures how communicative conventions emerge dynamically through repeated interactions, referential decisions are not always driven by such adaptive processes alone. In many contexts, individuals must engage in deliberative processing to actively construct and interpret social signals, particularly when facing novel communicative challenges or ambiguous contexts. For instance, a speaker explaining a scientific concept needs to adjust her wording based on the audience's assumed expertise, selecting terms differently for a colleague than for a child. Likewise, a traveler encountering an unfamiliar local term must infer its meaning by integrating relevant prior knowledge with contextual information. These behaviors engage flexible, context-sensitive inference to construct the meaning of social signals on the fly. We refer to this as the deliberative level of referential processing. The term "deliberative" here does not denote a general processing mode (such as controlled vs automatic) but rather marks a level of processing that becomes relevant when preestablished mappings, whether innate or learned, are absent, ambiguous, or contextually inappropriate. In such cases, referential processing requires explicit cognitive efforts to anticipate partners' perspectives, model their inferential process, and select signals optimizing communicative efficiency. This level of processing supports the strategic and adaptive deployment of referential behavior, particularly in novel or high-uncertain situations (Clark, 1996; Noveck and Rebol, 2008; Crawford, 2019).

Deliberative referential interaction provides a controlled yet ecologically valid testbed for investigating the neurocomputations of interactive social inferences. In a typical referential communication task, a signal sender operates under communication constraints (e.g., selecting from a predefined set of signals) and must encode their intent within those limitations, while a receiver—who has some knowledge about the sender's constraints—needs to disambiguate the intended meaning (Fig. 4A; Frank and Goodman, 2012). These interactions have two key properties that make them particularly valuable for studying interactive social inferences. First, referential communication is inherently goal-directed and cooperative: Both the sender and the receiver benefit from successful message transmission, distinguishing it from more complex communicative interactions involving competitive or deceptive incentives (Bhatt et al., 2010; Zhu et al., 2014; Oey et al., 2023). Second, referential signal selection and interpretation depend on modeling how social partners would reason. This recursive process makes deliberative referential interaction a valuable paradigm for studying the neural processes of interactive social inferences. Unlike

observational learning, during which the brain often passively learns about others' goals and beliefs as statistical regularities, interactive social engagement such as referential communication is inherently bidirectional—decisions both shape and are shaped by others' choices and beliefs (Rusch et al., 2020; Konovalov et al., 2021). A key challenge in social neuroscience is to understand how the representations of others' goals and inferences are leveraged to predict and influence others' decisions in such interactive contexts.

In the broad area of social decision-making, computational models of cognitive hierarchy and level- k reasoning are instrumental in guiding the investigation of such recursive social reasoning (Stahl, 1993; Camerer, 2003). Initially developed in behavioral game theory, these models have been applied across disciplines, including psychology and artificial intelligence (Lee and Seo, 2015; Rabinowitz et al., 2018; Rusch et al., 2020). They propose that interacting agents differ in their levels of strategic sophistication: a naive or level-0 agent acts nonstrategically, while a level- k agent models the inferred reasoning of a level- $(k-1)$ counterpart and optimizes their response accordingly. Under this framework, interactive social inferences can be conceptualized as an iterative process, such as "I think that you think that I think..." with deeper recursion indicating greater strategic sophistication. Large-scale behavioral studies using a classic economic game called the P-beauty contest, conducted with newspaper readers as a proxy for a representative population, suggest that human decision-makers, on average, engage in level-2 reasoning, yet with substantial individual variations (Camerer, 2003). At the neural level, a growing body of research suggests that differential levels of strategic reasoning is associated with distinct activation patterns in regions such as the dmPFC, TPJ, and vmPFC (Hampton et al., 2008; Coricelli and Nagel, 2009; Bhatt et al., 2010; Yoshida et al., 2010; Hill et al., 2017; Konovalov et al., 2021). Within these regions, the vmPFC exhibits greater activation during higher-order reasoning and is thought to integrate inputs from the dmPFC and TPJ (Coricelli and Nagel, 2009), combining strategic computations with valuation processes to guide behavior (Hampton et al., 2008; Hill et al., 2017). For the dmPFC and TPJ, in comparison, while these areas have long been complicated in theory of mind, recent findings start to suggest functional dissociations in their contributions to interactive social inferences, with the TPJ likely involved in tracking opponent-specific feedbacks and the dmPFC supporting strategy implementation (Konovalov et al., 2021).

Recent research has extended this framework to explore the neurocomputation of deliberative referential interactions. To capture the cooperative, recursive nature of such inferences, studies have drawn on insights from the rational speech act (RSA) model, which formalizes intentional communication as a goal-directed, Bayesian decision-making process under communicative uncertainty (Goodman and Frank, 2016). Originally developed in experimental pragmatics as a predictive model of language use, its core idea aligns closely with active inference frameworks in theoretical neuroscience (Friston and Frith, 2015a,b) and inverse reinforcement learning in artificial intelligence (Jara-Ettinger, 2019). The model posits that a sender selects a message that maximizes informativeness while minimizing communication cost, ensuring that the receiver can efficiently infer the intended meaning. Conversely, a receiver interprets the signal under the assumption that the sender is rational and cooperative, using Bayesian inference to resolve ambiguity. These models allow for flexible assumptions about the strategic sophistication levels of communicators, generating testable behavioral and neural predictions. At

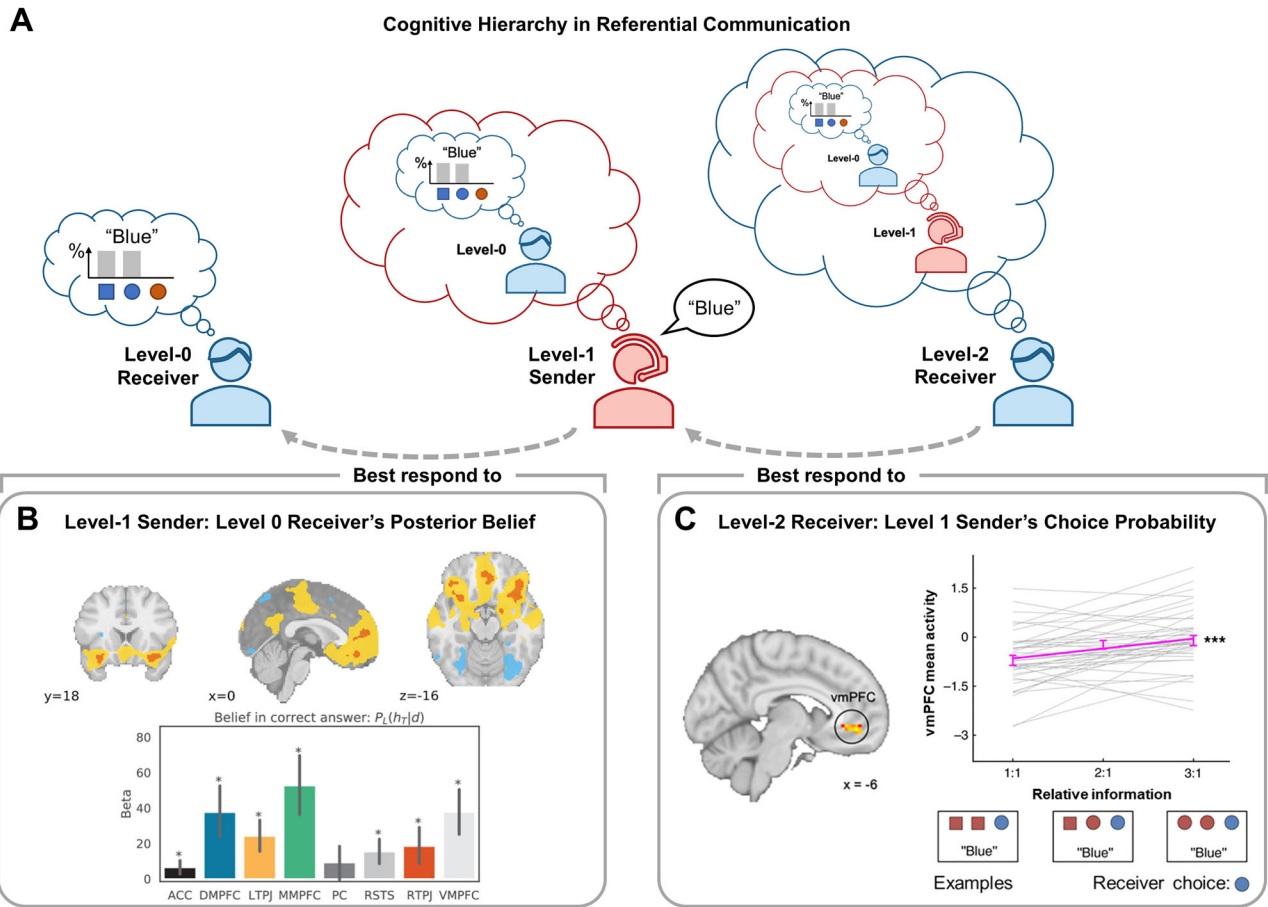


Figure 4. Deliberative inferences and cognitive hierarchy in referential communication. **A**, In the illustrated task, a sender must choose a signal (e.g., “blue”) to help a receiver identify the correct referent (e.g., the blue circle) from three candidate options. According to the cognitive hierarchy and RSA model, a level-0 receiver interprets the signal (“blue”) literally, choosing randomly among items that match the word. A level-1 sender anticipates this literal interpretation and selects the signal that most effectively narrows down the receiver’s options. A level-2 receiver goes one step further by inferring the sender’s reasoning—modeling how a rational sender would choose a signal given the available alternatives. This recursive structure allows researchers to isolate distinct levels of communicative reasoning. **B**, Neuroimaging evidence consistent with level-1 sender reasoning shows that dorsomedial prefrontal cortex (dmPFC) and TPJ and other regions encode the sender’s estimate of the literal receiver’s belief. **C**, Neuroimaging evidence consistent with level-2 receiver reasoning shows that the vmPFC tracks the inferred likelihood of a level-1 sender choosing a particular signal to refer to the intended object. **B**, Modified from Véléz et al. (2023), **C**, Adapted from Mi et al. (2021) and Jiang et al. (2023), under the permission of a Creative Commons Attribution-NonCommercial 4.0 International License (<https://creativecommons.org/licenses/by-nc/4.0/>).

the behavioral level, models related to RSA successfully predict social behaviors involving nonliteral messages, such as indirect speech, intentional ambiguity, and delays in social interactions (Stuhlmüller and Goodman, 2014; Jordan et al., 2016).

Guided by the framework, recent research examines the neural implementations of deliberative referential interaction in both signal senders and receivers. Findings from these research consistently suggest that referential communication is not merely a passive exchange of signals but a structured, multilevel inferential process. For senders, a recent investigation examines how signals (examples) are selected by senders to facilitate receivers’ (learners’) comprehension in a multiple-choice format (Véléz et al., 2023). Findings suggest that senders likely engage in level-1 reasoning, selecting signals that balance expected information gain for a literal, level-0 receiver against the cost of signaling. Neural evidence supports this cooperative, utility maximization strategy for senders, with activity in senders’ bilateral TPJ and dmPFC reflecting senders’ predictions about how a message will update a level-0 receiver’s belief (Fig. 4B). For receivers, a neuroimaging investigation examines a referential task in which senders are restricted to ambiguous signals, requiring the receivers to rely on contextual information for disambiguation

(Mi et al., 2021). Behavioral evidence suggests that receivers likely engage in level-2 reasoning, who assume that senders are level-1 and utility-maximizers and mentally invert senders’ decision-making process to derive the most likely meaning (Frank and Goodman, 2012; Franke and Degen, 2016). Activity in receivers’ vmPFC activity reflects the likelihood that a level-1 sender would produce the received message to convey a specific meaning (Fig. 4C). Receivers’ TPJ and dmPFC, in comparison, do not directly reflect receivers’ predictions about senders’ choice probabilities. Instead, these regions show stronger functional connectivity with the vmPFC when receivers’ choices are consistent with a level-2 receiver than when their choices are inconsistent.

Taken together, these findings align with the framework proposed by cognitive hierarchy theory, supporting distinct levels of social inferences in deliberative referential interactions. Senders act informatively, with the TPJ and dmPFC activity reflecting their beliefs about literal receivers, while receivers engage in higher-level reasoning, with the vmPFC encoding expectations about goal-directed decisions by a less strategic (level-1) sender. However, an important question remains: Do these neural correlates primarily reflect the communicative roles of senders and receivers, or do they instead correspond to the level of reasoning

each communicator engages in (level-1 vs level-2)? For example, if the involvement of the TPJ and dmPFC in level-1 reasoning is not specific to senders, these regions may also contribute to receivers' recursive reasoning, such as supporting the simulation from a level-1 sender's perspective about a literal receiver. This could explain why these regions in receivers do not directly encode predictions about senders but instead play an indirect role—likely influence signal interpretation by modulating vmPFC activity (Mi et al., 2021). Similarly, if the vmPFC's in level-2 reasoning extends beyond receivers, then prior evidence showing that vmPFC damage disrupts senders' ability to tailor messages to different audiences (e.g., adults vs children; Stolk et al., 2015) would suggest that audience design may rely on broader neural computations beyond the mere modeling of a literal listener. This raises the possibility that the vmPFC contributes to flexible referential behavior by integrating contextual inferences with mental modeling, enabling senders and receivers to adjust their strategies. Future research is needed to clarify the precise computational roles of these regions across communicative roles and levels of recursive inferences.

While these studies provide insight into the neural computations underlying deliberative referential interactions, they focus on controlled and stylized experimental paradigms in which communicative strategy spaces are constrained, and the dynamic, real-time nature of natural communication is largely reduced. Moreover, instead of relying on recursive inferences grounded in rational, utility-maximizing principles, real-world communication may in some cases favor alternative strategies, such as habitual responses or heuristic-based reasoning. Whether the neurocognitive processes identified in these settings generalize to such naturalistic contexts remains an open question. Additionally, further investigation is needed to elucidate the extent to which consensus—either established through stable conventions or interactively negotiated in real time—shapes deliberative communicative reasoning.

Discussion

Understanding how the brain detects, interprets, and responds to referential signals provides valuable insights into the broader question of social signal processing. The evidence discussed here supports the notion that referential signal processing spans multiple levels of neurocognitive operations. Early-emerging sensitivity to biological signals, such as eye gaze and BM, provides a foundation for referential processing. These evolutionarily conserved mechanisms prioritize socially salient stimuli, enabling rapid attentional orienting and fostering the development of more complex referential signal processing. While these innate responses shape early referential sensitivity, acquired referential signals such as symbolic gestures illustrate the role of social learning and adaptation in constructing referential meaning. Moreover, referential signals can also derive meaning from deliberative reasoning, where communicators strategically select and interpret signals based on their expectations of how others encode and decode communicative intent in context. Together, these processes suggest that the processing of referential signal is not a unitary process but supported by distinct yet complementary systems. The integration of innate sensitivities, acquired conventions, and deliberative reasoning ensures that referential meaning remains both stable and adaptable, facilitating effective information exchange in dynamic environments.

While the multilevel architecture is likely conserved across individuals, cultural context shapes how each level is expressed.

For example, even biologically grounded sensitivities to gaze or facial expressions can be modulated by cultural norms (Ekman et al., 1987; Altvater-Mackensen and Mani, 2013); symbolic gestures and linguistic conventions vary widely across communities (Wang and Vallotton, 2016); and deliberative inferences draw on culturally specific assumptions about perspective-taking, emotion perception, and shared knowledge (Kita and Özyürek, 2003; Misyak et al., 2014; Li, 2021). These differences reflect variation in the content and deployment of referential processes, not in their structural organization. The present framework thus provides a generalizable model of referential cognition, distinguishing system-level architecture from the culturally contingent contents it may take.

Notably, boundaries between these levels of referential processing are not discrete. At the neural level, while these levels of referential processing may engage distinct substrates, particularly along subcortical–cortical gradients, they also recruit overlapping regions, most notably in the temporal cortices (Fig. 1). At the cognitive level, communicative behavior often involves dynamic interactions across systems. For example, when individuals encounter a novel referential situation, they may initially engage in strategic or deliberative inference to establish shared meaning. With repeated exposure to similar contexts, the signal–meaning mapping may become conventionalized (e.g., a nickname or gesture), and the processing of that signal can then become routinized and supported by the learned system (Brennan and Clark, 1996; Stolk et al., 2014; Hawkins et al., 2020a). This transition suggests the involvement of an evaluation mechanism that flexibly arbitrates between levels of processing, selecting whether to engage deliberative inference or rely on established associations depending on contextual demands. Beyond such transitions, these multiple levels of processing may also operate in coordination. Behavioral evidence from both children and adults suggests that deliberative interpretation of referential acts is influenced by the perceptual and social saliency of cues, in a manner consistent with the possibility that these salient features function as Bayesian priors in support of inferential reasoning (Bohn et al., 2021; Mi et al., 2021). These findings suggest that innate and acquired responses are not purely stimulus-bound but can be flexibly incorporated into higher-order inference. Such coordination among different systems may rely on top–down modulation from cortical areas involved in mentalizing or executive control, which regulate subcortical responses and reshape the processing of preestablished signals in a context-sensitive manner. Future work is needed to clarify the mechanisms by which these systems interact, including how neural control systems arbitrate across levels of processing during social signal processing.

From a developmental perspective, these levels of referential processing likely emerge along a broad trajectory, reflecting increasing abstraction and cognitive flexibility. In early infancy, referential processing is primarily guided by innate mechanisms that bias attention toward socially salient cues. However, the attentional shifts elicited by such cues are largely reflexive, and infants exhibit limited understanding of the communicative intentions underlying others' behaviors at this stage (Butterworth and Jarrett, 1991). With accumulating social experience, learning-based referential processing gradually becomes more dominant. Infants progressively extract more nuanced referential meaning from others' gaze and bodily cues. For example, they increasingly rely on explicit directional cues, such as pointing gestures, rather than gaze alone, to guide their attention toward external objects (Butterworth and Itakura, 2000). By ~12 months of age, they begin

to infer the intended link between referential gestures and target objects, demonstrating emerging sensitivity to others' communicative intentions (Gliga and Csibra, 2009). In addition to responsive referential processing, longitudinal evidence suggests that the developmental trajectory of pointing initiated by infants themselves is also supported by both early-emerging gaze-following abilities and ongoing social interactions (Matthews et al., 2012). The onset and frequency of index-finger pointing were not significantly accelerated by increased parental modeling alone but rather were predicted by infants' prior ability to follow gaze. In contrast, socialization processes, such as caregivers' responsiveness and increased initiative pointing frequency during daily routines, played a stronger role in shaping the quality and communicative use of pointing (e.g., gaze alternation during pointing). These findings suggest that while the ability to point referentially depends on core social-cognitive competencies, it is further elaborated through experience and interaction. During this process, innate predispositions and social learning jointly establish a foundational referential system based on gaze and bodily cues, which scaffolds the later development of more complex language- and symbol-based referential abilities (Armstrong and Wilcox, 2007; Liszkowski et al., 2012). In contrast, the emergence of deliberative-level referential processing likely occurs later in development. This level depends on the ability to represent others' mental states and integrate contextual information to flexibly infer the communicative meaning of referential signals. Such processing is closely linked to the maturation of theory of mind, which typically becomes more robust after the age of 4, as reflected in consistent success on classic false-belief tasks (Callaghan et al., 2005; Rakoczy, 2022). Together, these developmental patterns suggest that the flexible, context-sensitive interpretation of referential signals is shaped by a dynamic interplay between innate biases, social learning, and higher-order reasoning across ontogeny.

Dysfunction within this multilevel processing framework has profound implications for neurodevelopmental conditions such as autism spectrum disorder (ASD). Individuals with ASD are widely found to exhibit atypical processing of verbal and nonverbal referential signals, with deficits traditionally attributed to impairments in higher-order functions such as mentalization (Malkin et al., 2018; Sidera et al., 2018). However, growing evidence links social deficits in ASD to early impairments in subcortical pathways (Baron-Cohen, 2000; Jure, 2019), particularly atypical activation in regions such as the amygdala (Ibrahim et al., 2024) and SC (Hadjikhani et al., 2017) during the processing of biologically significant referential signals like eye gaze. One hypothesis posits that individuals with ASD exhibit a reduced intrinsic preference for social information, resulting in diminished exposure to referential intent conveyed through eye gaze or BM. This, in turn, may limit opportunities to learn how to use these signals to guide other's attention, potentially hindering the development of higher-order communicative functions across cognitive levels. Consistent with this hypothesis, extensive eye-tracking research in ASD has shown that infants with ASD demonstrate a reduced preference for biological referential signals (Klin et al., 2009, 2015; Jones and Klin, 2013). Moreover, early attention-following behavior in response to biological cues has been found to predict later language development (Çetinçelik et al., 2021). Despite these findings, there remains a significant gap in understanding how atypical subcortical activity influences early-stage referential signal processing and how such influences cascade into more complex referential behavior over development. Future research should explore how innate, acquired, and deliberative processes interact over developmental

timescales and during real-time decision-making. Such work will shed light on the neurocognitive architecture of referential interactions in neurotypical and neurodivergent populations.

References

- Akiyama T, Kato M, Muramatsu T, Saito F, Umeda S, Kashima H (2006) Gaze but not arrows: a dissociative impairment after right superior temporal gyrus damage. *Neuropsychologia* 44:1804–1810.
- Akiyama T, Kato M, Muramatsu T, Umeda S, Saito F, Kashima H (2007) Unilateral amygdala lesions hamper attentional orienting triggered by gaze direction. *Cereb Cortex* 17:2593–2600.
- Akiyama T, Kato M, Muramatsu T, Maeda T, Hara T, Kashima H (2008) Gaze-triggered orienting is reduced in chronic schizophrenia. *Psychiatry Res* 158:287–296.
- Altwater-Mackensen N, Mani N (2013) Word-form familiarity bootstraps infant speech segmentation. *Dev Sci* 16:980–990.
- Armstrong DF, Wilcox SE (2007) *The gestural origin of language*. New York: Oxford University Press.
- Astor K, Gredebäck G (2022) Gaze following in infancy: five big questions that the field should answer. In: *Advances in child development and behavior* (Lockman JJ, ed), pp 191–223. Cambridge, MA: Elsevier.
- Babinet M-N, Cublier M, Demily C, Michael GA (2022) Eye direction detection and perception as premises of a social brain: a narrative review of behavioral and neural data. *Cogn Affect Behav Neurosci* 22:1–20.
- Baron-Cohen S (1995) The eye direction detector (EDD) and the shared attention mechanism (SAM): two cases for evolutionary psychology. In: *Joint attention: its origins and role in development* (Moore C, Dunham PJ, Dunham P, eds), pp 41–59. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Baron-Cohen S (2000) Theory of mind and autism: a review. In: *International review of research in mental retardation* (Glidden LM, ed), pp 169–184. Elsevier.
- Bašnáková J, Weber K, Petersson KM, Van Berkum J, Hagoort P (2014) Beyond the language given: the neural correlates of inferring speaker meaning. *Cereb Cortex* 24:2572–2578.
- Behrens TEJ, Muller TH, Whittington JCR, Mark S, Baram AB, Stachenfeld KL, Kurth-Nelson Z (2018) What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron* 100:490–509.
- Bhatt MA, Lohrenz T, Camerer CF, Montague PR (2010) Neural signatures of strategic types in a two-person bargaining game. *Proc Natl Acad Sci U S A* 107:19720–19725.
- Birmingham E, Kingstone A (2009) Human social attention: a new look at past, present, and future investigations. *Ann N Y Acad Sci* 1156:118–140.
- Bohn M, Tessler MH, Merrick M, Frank MC (2021) How young children integrate information sources to infer the meaning of words. *Nat Hum Behav* 5:1046–1054.
- Boyce V, Hawkins RD, Goodman ND, Frank MC (2024) Interaction structure constrains the emergence of conventions in group communication. *Proc Natl Acad Sci U S A* 121:e2403888121.
- Brennan SE, Clark HH (1996) Conceptual pacts and lexical choice in conversation. *J Exp Psychol Learn Mem Cogn* 22:1482–1493.
- Brignani D, Guzzon D, Marzi CA, Miniussi C (2009) Attentional orienting induced by arrows and eye-gaze compared with an endogenous cue. *Neuropsychologia* 47:370–381.
- Butterworth G, Itakura S (2000) How the eyes, head and hand serve definite reference. *Br J Dev Psychol* 18:25–50.
- Butterworth G, Jarrett N (1991) What minds have in common is space: spatial mechanisms serving joint visual attention in infancy. *Br J Dev Psychol* 9: 55–72.
- Callaghan T, Rochat P, Lillard A, Claux ML, Odden H, Itakura S, Tapanya S, Singh S (2005) Synchrony in the onset of mental-state reasoning: evidence from five cultures. *Psychol Sci* 16:378–384.
- Callejas A, Shulman GL, Corbetta M (2014) Dorsal and ventral attention systems underlie social and symbolic cueing. *J Cogn Neurosci* 26:63–80.
- Camerer C (2003) *Behavioral game theory: experiments in strategic interaction*. Princeton, NJ: Princeton University Press.
- Carlin JD, Calder AJ, Kriegeskorte N, Nili H, Rowe JB (2011) A head view-invariant representation of gaze direction in anterior superior temporal sulcus. *Curr Biol* 21:1817–1821.
- Carlin JD, Rowe JB, Kriegeskorte N, Thompson R, Calder AJ (2012) Direction-sensitive codes for observed head turns in human superior temporal sulcus. *Cereb Cortex* 22:735–744.

- Carroll JM (1980) Naming and describing in social communication. *Lang Speech* 23:309–322.
- Çetinçelik M, Rowland CF, Snijders TM (2021) Do the eyes have it? A systematic review on the role of eye gaze in infant language development. *Front Psychol* 11:589096.
- Chacón-Candia JA, Román-Caballero R, Aranda-Martín B, Casagrande M, Lupiáñez J, Marotta A (2022) Are there quantitative differences between eye-gaze and arrow cues? A meta-analytic answer to the debate and a call for qualitative differences. *Neurosci Biobehav Rev* 144:104993.
- Chong I, Ramezanzpour H, Thier P (2023) Causal manipulation of gaze-following in the macaque temporal cortex. *Prog Neurobiol* 226:102466.
- Clark HH (1996) *Using language*. Cambridge, UK: Cambridge University Press.
- Clark HH, Wilkes-Gibbs D (1986) Referring as a collaborative process. *Cognition* 22:1–39.
- Constantino JN, Kennon-McGill S, Weichselbaum C, Marrus N, Haider A, Glowinski AL, Gillespie S, Klaiman C, Klin A, Jones W (2017) Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature* 547:340–344.
- Coricelli G, Nagel R (2009) Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proc Natl Acad Sci U S A* 106:9163–9168.
- Crawford VP (2019) Experiments on cognition, communication, coordination, and cooperation in relationships. *Annu Rev Econ* 11:167–191.
- Dalmaso M (2014) Social modulators of social attention. Available at: <https://www.semanticscholar.org/paper/af539d4938cf74e0d44f7d3a6aa3d8a8005e7126>
- De Martino B, Cortese A (2023) Goals, usefulness and abstraction in value-based choice. *Trends Cogn Sci* 27:65–80.
- Driver J, Davis G, Ricciardelli P, Kidd P, Maxwell E, Baron-Cohen S (1999) Gaze perception triggers reflexive visuospatial orienting. *Vis Cogn* 6:509–540.
- Efferson C, Lalive R, Fehr E (2008) The coevolution of cultural groups and ingroup favoritism. *Science* 321:1844–1849.
- Ekman P, et al. (1987) Universals and cultural differences in the judgments of facial expressions of emotion. *J Pers Soc Psychol* 53:712–717.
- Engell AD, Nummenmaa L, Oosterhof NN, Henson RN, Haxby JV, Calder AJ (2010) Differential activation of frontoparietal attention networks by social and symbolic spatial cues. *Soc Cogn Affect Neurosci* 5:432–440.
- Farroni T, Csibra G, Simion F, Johnson MH (2002) Eye contact detection in humans from birth. *Proc Natl Acad Sci U S A* 99:9602–9605.
- Farroni T, Massaccesi S, Pividori D, Johnson MH (2004) Gaze following in newborns. *Infancy* 5:39–60.
- Farroni T, Johnson MH, Menon E, Zulian L, Faraguna D, Csibra G (2005) Newborns' preference for face-relevant stimuli: effects of contrast polarity. *Proc Natl Acad Sci U S A* 102:17245–17250.
- Feiler L, Camerer CF (2010) Code creation in endogenous merger experiments. *Econ Inq* 48:337–352.
- Frank MC, Goodman ND (2012) Predicting pragmatic reasoning in language games. *Science* 336:998.
- Frank M, Degen J (2016) Reasoning in reference games: individual- vs. population-level probabilistic modeling Allen P, ed. *PLoS One* 11:e0154854.
- Frank M, Wagner EO (2014) Game theory and the evolution of meaning. *Lang Linguist Compass* 8:359–372.
- Friesen CK, Ristic J, Kingstone A (2004) Attentional effects of counter-predictive gaze and arrow cues. *J Exp Psychol Hum Percept Perform* 30:319.
- Friesen CK, Kingstone A (1998) The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychon Bull Rev* 5:490–495.
- Friston KJ, Frith CD (2015a) A duet for one. *Conscious Cogn* 36:390–405.
- Friston KJ, Frith CD (2015b) Active inference, communication and hermeneutics. *Cortex* 68:129–143.
- Gangopadhyay P, Chawla M, Dal Monte O, Chang SWC (2021) Prefrontal-amygdala circuits in social decision-making. *Nat Neurosci* 24:5–18.
- Garrod S, Fay N, Lee J, Oberlander J, MacLeod T (2007) Foundations of representation: where might graphical symbol systems come from? *Cogn Sci* 31:961–987.
- Garrod S, Doherty G (1994) Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition* 53:181–215.
- Gliga T, Csibra G (2009) One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychol Sci* 20:347–353.
- Goodman ND, Frank MC (2016) Pragmatic language interpretation as probabilistic inference. *Trends Cogn Sci* 20:818–829.
- Grossmann T (2015) The development of social brain functions in infancy. *Psychol Bull* 141:1266.
- Hadjikhani N, Johnels JÅ, Zürcher NR, Lassalle A, Guillon Q, Hippolyte L, Billstedt E, Ward N, Lemonnier E, Gillberg C (2017) Look me in the eyes: constraining gaze in the eye-region provokes abnormally high subcortical activation in autism. *Sci Rep* 7:1–7.
- Hampton AN, Bossaerts P, O'Doherty JP (2008) Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc Natl Acad Sci U S A* 105:6741–6746.
- Han Z, Bi Y, Chen J, Chen Q, He Y, Caramazza A (2013) Distinct regions of right temporal cortex are associated with biological and human-agent motion: functional magnetic resonance imaging and neuropsychological evidence. *J Neurosci* 33:15442–15453.
- Hasson U, Ghazanfar AA, Galantucci B, Garrod S, Keysers C (2012) Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn Sci* 16:114–121.
- Hawkins RD, Kwon M, Sadigh D, Goodman N (2020a) Continual adaptation for efficient machine communication. In: *Proceedings of the 24th conference on computational natural language learning*, pp 408–419. Online: Association for Computational Linguistics.
- Hawkins RD, Frank MC, Goodman ND (2020b) Characterizing the dynamics of learning in repeated reference games. *Cogn Sci* 44:e12845.
- Hawkins RD, Sano M, Goodman ND, Fan JE (2023) Visual resemblance and interaction history jointly constrain pictorial meaning. *Nat Commun* 14:2199.
- Hawkins RXD, Goodman ND, Goldstone RL (2019) The emergence of social norms and conventions. *Trends Cogn Sci* 23:158–169.
- Hietanen JK, Leppänen JM, Nummenmaa L, Astikainen P (2008) Visuospatial attention shifts by gaze and arrow cues: an ERP study. *Brain Res* 1215:123–136.
- Hill CA, Suzuki S, Polanía R, Moisa M, O'Doherty JP, Ruff CC (2017) A causal account of the brain network computations underlying strategic social behavior. *Nat Neurosci* 23:1–27.
- Hirai M, Senju A (2020) The two-process theory of biological motion processing. *Neurosci Biobehav Rev* 111:114–124.
- Hirata S, Fuwa K, Sugama K, Kusunoki K, Fujita S (2010) Facial perception of conspecifics: chimpanzees (*Pan troglodytes*) preferentially attend to proper orientation and open eyes. *Anim Cogn* 13:679–688.
- Ho MK, Abel D, Griffiths TL, Littman ML (2019) The value of abstraction. *Curr Opin Behav Sci* 29:111–116.
- Hommel B, Pratt J, Colzato L, Godijn R (2001) Symbolic control of visual attention. *Psychol Sci* 12:360–365.
- Hupet M, Chantraine Y (1992) Changes in repeated references: collaboration or repetition effects? *J Psycholinguist Res* 21:485–496.
- Ibarra A, Tanenhaus MK (2016) The flexibility of conceptual pacts: referring expressions dynamically shift to accommodate new conceptualizations. *Front Psychol* 7:561.
- Ibrahim K, Iturmendi-Sabater I, Vasisht M, Barron DS, Guardavaccaro M, Funaro MC, Holmes A, McCarthy G, Eickhoff SB, Sukhodolsky DG (2024) Neural circuit disruptions of eye gaze processing in autism spectrum disorder and schizophrenia: an activation likelihood estimation meta-analysis. *Schizophr Res* 264:298–313.
- Jakobsen KV, Frick JE, Simpson EA (2013) Look here! The development of attentional orienting to symbolic cues. *J Cogn Dev* 14:229–249.
- Jara-Ettinger J (2019) Theory of mind as inverse reinforcement learning. *Curr Opin Behav Sci* 29:105–110.
- Ji H, Wang L, Jiang Y (2020) Cross-category adaptation of reflexive social attention. *J Exp Psychol Gen* 149:2145–2153.
- Jiang J, Chen C, Dai B, Shi G, Ding G, Liu L, Lu C (2015) Leader emergence through interpersonal neural synchronization. *Proc Natl Acad Sci U S A* 112:4274–4279.
- Jiang Y, Wu H, Mi Q, Zhu L (2022) Neurocomputations of strategic behavior: from iterated to novel interactions. *Wiley Interdiscip Rev Cogn Sci* 13:e1598.
- Jiang Y, Mi Q, Zhu L (2023) Neurocomputational mechanism of real-time distributed learning on social networks. *Nat Neurosci* 26:506–516.
- Johnson MH, Senju A, Tomalski P (2015) The two-process theory of face processing: modifications based on two decades of data from infants and adults. *Neurosci Biobehav Rev* 50:169–179.
- Jones W, Klin A (2013) Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature* 504:427–431.
- Jonides J (1981) Voluntary versus automatic control over the mind's eye's movements. In: *Attention and performance* (Baddeley A, ed), pp 187–203. Hillsdale, NJ: Erlbaum.

- Jordan JJ, Hoffman M, Nowak MA, Rand DG (2016) Uncalculating cooperation is used to signal trustworthiness. *Proc Natl Acad Sci U S A* 113:8658–8663.
- Jure R (2019) Autism pathogenesis: the superior colliculus. *Front Neurosci* 12:1029.
- King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR, Montague PR (2005) Getting to know you: reputation and trust in a two-person economic exchange. *Science* 308:78–83.
- Kingstone A, Tipper C, Ristic J, Ngan E (2004) The eyes have it! An fMRI investigation. *Brain Cogn* 55:269–271.
- Kita S, Özyürek A (2003) What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *J Mem Lang* 48:16–32.
- Klin A, Lin DJ, Gorrindo P, Ramsay G, Jones W (2009) Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature* 459:257–261.
- Klin A, Shultz S, Jones W (2015) Social visual engagement in infants and toddlers with autism: early developmental transitions and a model of pathogenesis. *Neurosci Biobehav Rev* 50:189–203.
- Knudsen EI (2018) Neural circuits that mediate selective attention: a comparative perspective. *Trends Neurosci* 41:789–805.
- Kobylykov D, Rosa-Salva O, Zanon M, Vallortigara G (2024) Innate face-selectivity in the brain of young domestic chicks. *Proc Natl Acad Sci U S A* 121:e2410404121.
- Kobylykov D, Vallortigara G (2024) Face detection mechanisms: nature vs. nurture. *Front Neurosci* 18:1404174.
- Konidaris G (2019) On the necessity of abstraction. *Curr Opin Behav Sci* 29:1–7.
- Kononov A, Hill C, Daunizeau J, Ruff CC (2021) Dissecting functional contributions of the social brain to strategic behavior. *Neuron* 109:3323–3337.e5.
- Krauss RM, Weinheimer S (1964) Changes in reference phrases as a function of frequency of usage in social interaction: a preliminary study. *Psychon Sci* 1:113–114.
- Lambon Ralph MA, Sage K, Jones RW, Mayberry EJ (2010) Coherent concepts are computed in the anterior temporal lobes. *Proc Natl Acad Sci U S A* 107:2717–2722.
- Lazaridou A, Peysakhovich A, Baroni M (2017) Multi-agent cooperation and the emergence of (natural) language. In: 5th International conference on learning representations, ICLR 2017. Toulon, France.
- Le QV, Le QV, Nishimaru H, Matsumoto J, Takamura Y, Hori E, Maior RS, Tomaz C, Ono T, Nishijo H (2020) A prototypical template for rapid face detection is embedded in the monkey superior colliculus. *Front Syst Neurosci* 14:5.
- Leadner K, Sekely L, Klein RM, Gabay S (2021) Evolution of social attentional cues: evidence from the archerfish. *Cognition* 207:104511.
- Lee D, Seo H (2015) Neural basis of strategic decision making. *Trends Neurosci* 39:40–48.
- Leonard TK, Blumenthal G, Gothard KM, Hoffman KL (2012) How macaques view familiarity and gaze in conspecific faces. *Behav Neurosci* 126:781.
- Lewis DK (1969) Convention: a philosophical study. *The philosophical quarterly*. Available at: <https://academic.oup.com/pq/article-lookup/doi/10.2307/2218418> Accessed March 9, 2025.
- Li J (2021) The origin of cross-cultural differences in referential intuitions: perspective taking in the Gödel case. *J Semant* 38:415–440.
- Ling S, Pratte MS, Tong F (2015) Attention alters orientation processing in the human lateral geniculate nucleus. *Nat Neurosci* 18:496–498.
- Liszkowski U, Brown P, Callaghan T, Takada A, De Vos C (2012) A prelinguistic gestural universal of human communication. *Cogn Sci* 36:698–713.
- Liu W, Yuan X, Liu D, Wang L, Jiang Y (2021) Social attention triggered by eye gaze and walking direction is resistant to temporal decay. *J Exp Psychol Hum Percept Perform* 47:1237.
- Lorenzi E, Nadalin G, Morandi-Raikova A, Mayer U, Vallortigara G (2024) Noncortical coding of biological motion in newborn chicks' brain. *Cereb Cortex* 34:bhae262.
- Lu X, Hu Z, Xin Y, Yang T, Wang Y, Zhang P, Liu N, Jiang Y (2024) Detecting biological motion signals in human and monkey superior colliculus: a subcortical-cortical pathway for biological motion perception. *Nat Commun* 15:9606.
- Lyon DC, Nassi JJ, Callaway EM (2010) A disinaptic relay from superior colliculus to dorsal stream visual cortex in macaque monkey. *Neuron* 65:270–279.
- Ma X, Yuan X, Liu J, Shen L, Yu Y, Zhou W, Liu Z, Jiang Y (2022) Gravity-dependent animacy perception in zebrafish. *Research* 2022:9829016.
- Malkin L, Abbot-Smith K, Williams D (2018) Is verbal reference impaired in autism spectrum disorder? A systematic review. *Autism Dev Lang Impair* 3:2396941518763166.
- Marciniak K, Atabaki A, Dicke PW, Thier P (2014) Disparate substrates for head gaze following and face perception in the monkey superior temporal sulcus. *Elife* 3:e03222.
- Matthews D, Behne T, Lieven E, Tomasello M (2012) Origins of the human pointing gesture: a training study. *Dev Sci* 15:817–829.
- Metzing C (2003) When conceptual pacts are broken: partner-specific effects on the comprehension of referring expressions. *J Mem Lang* 49:201–213.
- Mi Q, Wang C, Camerer CF, Zhu L (2021) Reading between the lines: listener's vmPFC simulates speaker cooperative choices in communication games. *Sci Adv* 7:eabe276.
- Milne E, Grafman J (2001) Ventromedial prefrontal cortex lesions in humans eliminate implicit gender stereotyping. *J Neurosci* 21:RC150.
- Mishchanchuk K, Gregoriou G, Qü A, Kastler A, Huys QJM, Wilbrecht L, MacAskill AF (2024) Hidden state inference requires abstract contextual representations in the ventral hippocampus. *Science* 386:926–932.
- Misyak JB, Melkonyan T, Zeitoun H, Chater N (2014) Unwritten rules: virtual bargaining underpins social interaction, culture, and society. *Trends Cogn Sci* 18:512–519.
- Mizzi R, Michael GA (2016) Exploring visual attention functions of the human extrageniculate pathways through behavioral cues. *Psychol Rev* 123:740.
- Morton J, Johnson MH (1991) CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychol Rev* 98:164.
- Muschinski J, Feczko E, Brooks JM, Collantes M, Heitz TR, Parr LA (2016) The development of visual preferences for direct versus averted gaze faces in infant macaques (*Macaca mulatta*). *Dev Psychobiol* 58:926–936.
- Myowa-Yamakoshi M, Tomonaga M, Tanaka M, Matsuzawa T (2003) Preference for human direct gaze in infant chimpanzees (Pan troglodytes). *Cognition* 89:113–124.
- Noveck IA, Reboul A (2008) Experimental pragmatics: a Gricean turn in the study of language. *Trends Cogn Sci* 12:425–431.
- Nummenmaa L, Passamonti L, Rowe J, Engell AD, Calder AJ (2010) Connectivity analysis reveals a cortical network for eye gaze perception. *Cereb Cortex* 20:1780–1787.
- Nummenmaa L, Smirnov D, Lahnakoski JM, Glerean E, Jääskeläinen IP, Sams M, Hari R (2014) Mental action simulation synchronizes action-observation circuits across individuals. *J Neurosci* 34:748–757.
- Oey LA, Schachner A, Vul E (2023) Designing and detecting lies by reasoning about other agents. *J Exp Psychol Gen* 152:346–362.
- Palazzolo G (2024) A case for animal reference: beyond functional reference and meaning attribution. *Synthese* 203:59.
- Park SA, Sestito M, Boorman ED, Dreher J-C (2019) Neural computations underlying strategic social decision-making in groups. *Nat Commun* 10:5287.
- Peer M, Brunec IK, Newcombe NS, Epstein RA (2021) Structuring knowledge with cognitive maps and cognitive graphs. *Trends Cogn Sci* 25:37–54.
- Pitcher D, Japee S, Rauth L, Ungerleider LG (2017) The superior temporal sulcus is causally connected to the amygdala: a combined TBS-fMRI study. *J Neurosci* 37:1156–1161.
- Portugal AM, Viktorsson C, Taylor MJ, Mason L, Tammimies K, Ronald A, Falck-Ytter T (2023) Infants' looking preferences for social versus non-social objects reflect genetic variation. *Nat Hum Behav* 8:115–124.
- Pozzi M, Bangerter A, Mazzarella D (2024) Does lexical coordination affect epistemic and practical trust? The role of conceptual pacts. *Cogn Sci* 48:e13372.
- Rabinowitz NC, Perbet F, Song HF, Zhang C, Eslami SMA, Botvinick M (2018) Machine theory of mind. Available at: <http://arxiv.org/abs/1802.07740> Accessed Dec. 29, 2022.
- Rakoczy H (2022) Foundations of theory of mind and its development in early childhood. *Nat Rev Psychol* 1:223–235.
- Ramezanpour H, Thier P (2020) Decoding of the other's focus of attention by a temporal cortex module. *Proc Natl Acad Sci U S A* 117:2663–2670.
- Ramsey R, Cross ES, Hamilton AF (2011) Eye can see what you want: posterior intraparietal sulcus encodes the object of an actor's gaze. *J Cogn Neurosci* 23:3400–3409.

- Ristic J, Friesen CK, Kingstone A (2002) Are eyes special? It depends on how you look at it. *Psychon Bull Rev* 9:507–513.
- Ristic J, Wright A, Kingstone A (2007) Attentional control and reflexive orienting to gaze and arrow cues. *Psychon Bull Rev* 14:964–969.
- Rosati AG, Arre AM, Platt ML, Santos LR (2016) Rhesus monkeys show human-like changes in gaze following across the lifespan. *Proc Biol Sci* 283:20160376.
- Rusch T, Steixner-Kumar S, Doshi P, Spezio M, Gläscher J (2020) Theory of mind and decision science: towards a typology of tasks and computational models. *Neuropsychologia* 146:107488.
- Sabbagh MA (1999) Communicative intentions and language: evidence from right-hemisphere damage and autism. *Brain Lang* 70:29–69.
- Salera C, Boccia M, Pecchinenda A (2024) Segregation of neural circuits involved in social gaze and non-social arrow cues: evidence from an activation likelihood estimation meta-analysis. *Neuropsychol Rev* 34:496–510.
- Salva OR, Mayer U, Vallortigara G (2015) Roots of a social brain: developmental models of emerging animacy-detection mechanisms. *Neurosci Biobehav Rev* 50:150–168.
- Sato W, Okada T, Toichi M (2007) Attentional shift by gaze is triggered without awareness. *Exp Brain Res* 183:87–94.
- Sato W, Kochiyama T, Uono S, Yoshikawa S (2009) Commonalities in the neural mechanisms underlying automatic attentional shifts by gaze, gestures, and symbols. *Neuroimage* 45:984–992.
- Saxe R, Houlihan SD (2017) Formalizing emotion concepts within a Bayesian model of theory of mind. *Curr Opin Psychol* 17:15–21.
- Schilbach L (2015) The neural correlates of social cognition and social interaction. In: *Brain mapping* (Toga AW, ed), pp 159–164. Cambridge, MA: Elsevier.
- Schurz M, Radua J, Aichhorn M, Richlan F, Perner J (2014) Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci Biobehav Rev* 42:9–34.
- Senju A, Vernetti A, Ganea N, Hudry K, Tucker L, Charman T, Johnson MH (2015) Early social experience affects the development of eye gaze processing. *Curr Biol* 25:3086–3091.
- Shen L, Li S, Tian Y, Wang Y, Jiang Y (2025) Cortical tracking of hierarchical rhythms orchestrates the multisensory processing of biological motion. *Elife* 13:RP98701.
- Shepherd SV, Klein JT, Deaner RO, Platt ML (2009) Mirroring of attention by neurons in macaque parietal cortex. *Proc Natl Acad Sci U S A* 106:9489–9494.
- Shi J, Weng X, He S, Jiang Y (2010) Biological motion cues trigger reflexive attentional orienting. *Cognition* 117:348–354.
- Sidera F, Perpiñá G, Serrano J, Rostan C (2018) Why is theory of mind important for referential communication? *Curr Psychol* 37:82–97.
- Stevens B, Welker C, Hasson U, Kleinbaum AM, Wheatley T (2024) Consensus-building conversation leads to neural alignment. *Nat Commun* 15:3936.
- Silbert LJ, Honey CJ, Simony E, Poeppel D, Hasson U (2014) Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proc Natl Acad Sci U S A* 111:E4687–E4696.
- Simion F, Regolin L, Bulf H (2008) A predisposition for biological motion in the newborn baby. *Proc Natl Acad Sci U S A* 105:809–813.
- Simpson EA, Paukner A, Pedersen EJ, Ferrari PF, Parr LA (2019) Visual preferences for direct-gaze faces in infant macaques (*Macaca mulatta*) with limited face exposure. *Dev Psychobiol* 61:228–238.
- Slater A, Quinn PC, Hayes R, Brown E (2000) The role of facial orientation in newborn infants' preference for attractive faces. *Dev Sci* 3:181–185.
- Stahl DO (1993) Evolution of smartn players. *Games Econ Behav* 5:604–617.
- Stephens GJ, Silbert LJ, Hasson U (2010) Speaker-listener neural coupling underlies successful communication. *Proc Natl Acad Sci U S A* 107:14425–14430.
- Stolk A, Verhagen L, Schoffelen J-M, Oostenveld R, Blokpoel M, Hagoort P, Van Rooij I, Toni I (2013) Neural mechanisms of communicative innovation. *Proc Natl Acad Sci U S A* 110:14574–14579.
- Stolk A, Noordzij ML, Verhagen L, Volman I, Schoffelen J-M, Oostenveld R, Hagoort P, Toni I (2014) Cerebral coherence between communicators marks the emergence of meaning. *Proc Natl Acad Sci U S A* 111:18183–18188.
- Stolk A, D'Imperio D, di Pellegrino G, Toni I (2015) Altered communicative decisions following ventromedial prefrontal lesions. *Curr Biol* 25:1469–1474.
- Stuhlmüller A, Goodman ND (2014) Reasoning about reasoning by nested conditioning: modeling theory of mind with probabilistic programs. *Cogn Syst Res* 28:80–99.
- Tipper CM, Handy TC, Giesbrecht B, Kingstone A (2008) Brain responses to biological relevance. *J Cogn Neurosci* 20:879–891.
- Tipples J (2002) Eye gaze is not unique: automatic orienting in response to uninformative arrows. *Psychon Bull Rev* 9:314–318.
- Tomasello M, Hare B, Lehmann H, Call J (2007) Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. *J Hum Evol* 52:314–320.
- Troje NF, Westhoff C (2006) The inversion effect in biological motion perception: evidence for a “life detector”? *Curr Biol* 16:821–824.
- Uono S, Sato W, Kochiyama T (2014) Commonalities and differences in the spatiotemporal neural dynamics associated with automatic attentional shifts induced by gaze and arrows. *Neurosci Res* 87:56–65.
- Vaidya AR, Badre D (2022) Abstract task representations for inference and control. *Trends Cogn Sci* 26:484–498.
- Vallortigara G, Regolin L, Marconato F (2005) Visually inexperienced chicks exhibit spontaneous preference for biological motion patterns. *PLoS Biol* 3:e208.
- Vélez N, Chen AM, Burke T, Cushman FA, Gershman SJ (2023) Teachers recruit mentalizing regions to represent learners' beliefs. *Proc Natl Acad Sci U S A* 120:e2215015120.
- Versace E, Damini S, Stancher G (2020) Early preference for face-like stimuli in solitary species as revealed by tortoise hatchlings. *Proc Natl Acad Sci U S A* 117:24047–24049.
- Waddington CH (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150:563–565.
- Wang L, Yang X, Shi J, Jiang Y (2014) The feet have it: local biological motion cues trigger reflexive attentional orienting in the brain. *Neuroimage* 84:217–224.
- Wang L, Wang Y, Xu Q, Liu D, Ji H, Yu Y, Hu Z, Yuan P, Jiang Y (2020) Heritability of reflexive social attention triggered by eye gaze and walking direction: common and unique genetic underpinnings. *Psychol Med* 50:475–483.
- Wang R, Yuan T, Wang L, Jiang Y (2024) A common and specialized neural code for social attention triggered by eye gaze and biological motion. *Neuroimage* 301:120889.
- Wang W, Vallotton C (2016) Cultural transmission through infant signs: objects and actions in U.S. and Taiwan. *Infant Behav Dev* 44:98–109.
- Wang Y, Wang L, Xu Q, Liu D, Chen L, Troje NF, He S, Jiang Y (2018) Heritable aspects of biological motion perception and its covariation with autistic traits. *Proc Natl Acad Sci U S A* 115:1937–1942.
- Wang Y, Zhang X, Wang C, Huang W, Xu Q, Liu D, Zhou W, Chen S, Jiang Y (2022) Modulation of biological motion perception in humans by gravity. *Nat Commun* 13:2765.
- Weber RA, Camerer CF (2003) Cultural conflict and merger failure: an experimental approach. *Manage Sci* 49:400–415.
- Xu S, Zhang S, Geng H (2011) Gaze-induced joint attention persists under high perceptual load and does not depend on awareness. *Vision Res* 51:2048–2056.
- Yang F, Tian J, Yuan P, Liu C, Zhang X, Yang L, Jiang Y (2024) Unconscious and conscious gaze-triggered attentional orienting: distinguishing innate and acquired components of social attention in children and adults with autistic traits and autism spectrum disorders. *Research* 7:0417.
- Yoshida W, Seymour B, Friston KJ, Dolan RJ (2010) Neural mechanisms of belief inference during cooperative games. *J Neurosci* 30:10744–10751.
- Zeiträg C, Jensen TR, Osvath M (2022) Gaze following: a socio-cognitive skill rooted in deep time. *Front Psychol* 13:950935.
- Zeiträg C, Reber SA, Osvath M (2023) Gaze following in Archosauria—alligators and palaeognath birds suggest dinosaur origin of visual perspective taking. *Sci Adv* 9:eadf0405.
- Zhu L, Mathewson KE, Hsu M (2012) Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proc Natl Acad Sci U S A* 109:1419–1424.
- Zhu L, Jenkins AC, Set E, Scabini D, Knight RT, Chiu PH, King-Casas B, Hsu M (2014) Damage to dorsolateral prefrontal cortex affects tradeoffs between honesty and self-interest. *Nat Neurosci* 17:1319–1321.
- Zhu L, Jiang Y, Scabini D, Knight RT, Hsu M (2019) Patients with basal ganglia damage show preserved learning in an economic game. *Nat Commun* 10:802.