Representation of others' beliefs

Jingmin Qin and Haiyan Wu, Centre for Cognitive and Brain Sciences and Department of Psychology, University of Macau, Macau, China

Key points

- The distinction between self and others plays a vital role in social interactions, including shaping self- and other-representations
- Self- and other-representations interplay with each other, in social interactions.
- Representation of others affected by both individual and environment factors, from micro to macro level.
- People may use Bayesian inference to update their own self-representation, and belief to others.
- The representation of other's belief may be formed and updated by observation and active interaction.
- Such capacity of representation of others suggests individual difference, and evolves across the lifespan.

Abstract

This article delves into the diverse aspects in which individuals make inferences about the beliefs and values held by others. By reviewing the psychological factors underlying representing the beliefs of others, as well as the individual differences for both the individuals being represented and those undertaking the representation, this article sheds light on the intricate nature of representation of other's belief. Furthermore, it discusses the ways and considerations involved in updating these beliefs (such as observation, active interaction, and Bayesian inference) and offers suggestions for future research.

Introduction

Background and significance of studying representation of others' beliefs

The ability to represent the beliefs, attitudes, and values of others is a critical aspect of human communication and understanding (Rokeach, 1970). Our capacity to interpret the thoughts and feelings of others is essential for building meaningful relationships and resolving conflicts (Karniol, 1990). Moreover, the ability to understand and empathize with others' perspectives can lead to more effective decision-making in various domains, including business, politics, and education.

Studying representation of others' beliefs provides valuable insights into how individuals process and make sense of information from diverse sources. For example, researchers have found that when people are exposed to information that contradicts their existing beliefs or values, they may experience cognitive dissonance and seek to resolve this conflict by adjusting their beliefs or values accordingly (Hart et al., 2009). This phenomenon highlights the importance of considering multiple perspectives when interpreting the beliefs of others.

Understanding representation of others' beliefs can also help us better understand the factors that influence people's decisions and actions (Frith and Frith, 1999). By examining how individuals construct and justify their beliefs, researchers can identify patterns and trends that may be relevant to different contexts and domains (Festinger, 1954). For instance, studies have shown that people tend to favor information that confirms their preexisting beliefs (Kahneman and Tversky, 1979), which suggests that understanding how individuals construct and justify their beliefs can provide useful insights into decision-making.

The study of representation of others' beliefs has gained significant attention across multiple disciplines, including psychology, economic (Kim, 2009), philosophy (Zynda, 2000), neuroscience (Ramsey et al., 2021), education and politics science (Elcheroth et al., 2011). Researchers from these diverse fields are interested in unraveling the complex cognitive processes that underlie our ability to reason about the thoughts and beliefs of other individuals. Educators, for example, need to have a deep understanding of how students perceive themselves and others (i.e., represent learners' beliefs), in order to create effective and inclusive learning environments (Vélez et al., 2023). In clinical psychology, the ability to accurately interpret and understand the mental states of patients is crucial for providing appropriate interventions and treatments (Salvatore et al., 2012). Similarly, social workers often encounter individuals from diverse backgrounds and must be skilled in navigating the complexities of others' beliefs and experiences. For example, the recent interactive mentalizing theory (IMT) proposes new components including co-mentalizing and group mentalizing in social contexts (Wu et al, 2020). These format of mentalizing are frequently observed in daily life and workplaces settings (**Fig. 1**).

Overview of the article structure

Previous introduction of the representation of others' belief have typically followed the traditional approach to understanding mentalizing, which assumes a limited number of agents and factors in the social interaction (Wu et al., 2020). They often present research that aims to find evidence for different physiological and behavioral response patterns associated with different mentalizing states. In contrast, this article takes a dynamic perspective on the simple and complex social interaction settings, focusing on the factors that affect both representation of oneself and others, and drive continuous process in formation, updating and biases in representations of others' mind (Griffiths et al., 2010). Categorization and labeling are considered to initiate first and there are dynamic changing steps in this view.

This theory-independent review focuses on the static versus dynamic representation of others' beliefs and the interaction between self-other representation and dynamics of updating. It provides a summary of empirical literature, adopting a self-other dimension and a static-dynamic update dimension. The review integrates classic mentalizing predictions with available empirical evidence, proposing an integration theory that considers micro-macro environment factors in other belief representation. It prioritizes hypothesis-driven approaches, examining how manipulations in controlled laboratory or culture settings modulate these representations. The review also highlights studies utilizing advanced computational modeling, such as Bayesian models, for objective assessments of different aspects of other belief representation beyond subjective inference reports or category labels.

Static representation of others

The development of distinguishing self from others

In order to engage in higher-order social cognition effectively, it is essential to integrate and differentiate between representations of oneself and others (Adolphs, 2009). This ability enables us to infer mental states of others as we need to recognize "self" at first (Payne and Tsakiris, 2017; Santiesteban et al., 2012; Wang et al., 2016). Having a self-representation means that we have a mental framework or cognitive structure that allows us to understand and interpret ourselves as an individual. As a common view, the process of gaining self-representation is developmental, basically from perception to representation.

Neisser (1991) demonstrated that infants acquire the self-perception first through transactions with the environment (ecological self) and others (interpersonal self), implying that self is shaped by physical and social information. Within a few weeks of birth, infants rudimentarily percept themselves by finger explosion (Kravitz et al., 1978) and are more sensitive to information related to their own experience (Filippetti et al., 2013). As cognitive ability matures, infants start to recognize themselves beyond sense perception. The best known paradigm to test the occurrence of self-representation is the mirror self-recognition task, developed by

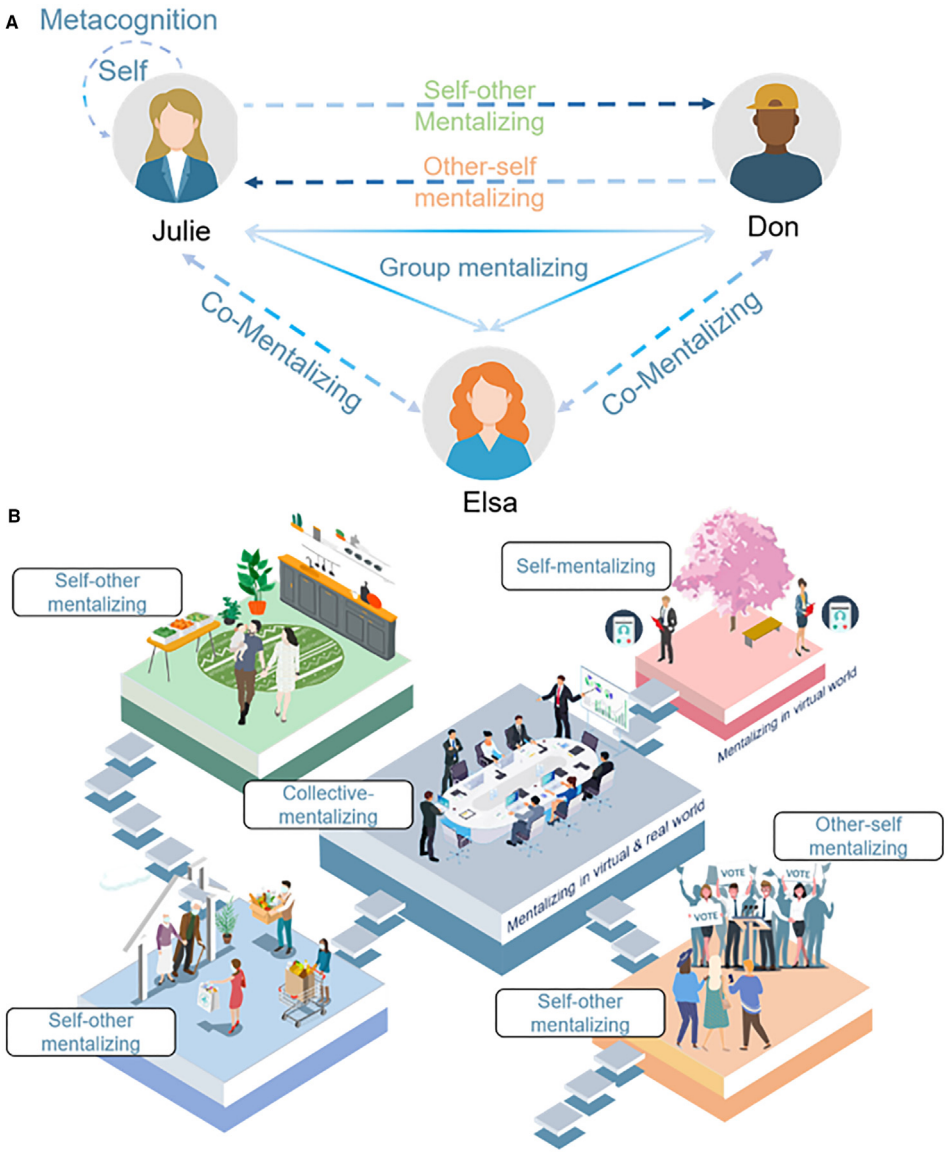

Fig. 1 Representation of other's mind in life (Wu et al., 2020) and the workspace.

Schematic illustration of different types of minds mentalizing (A) in both daily life and workspace (B). Note that these representations or mentalizing can also be referred to, respectively, as self-mentalizing, first order mentalizing, higher order mentalizing, and collective mentalizing.

- Self-Mentalizing: An individual reflecting on their own desires and goal before deciding.
- Collective Mentalizing: A team working together to come up with a strategy, considering each member's perspectives and aligning their collective goals.
- Self-other mentalizing: The ability to attribute mental states that a friend might have a different perspective on a political issue and try to understand their reasoning behind it. Trying to empathize with a coworker's critical comments by considering their previous experiences and understanding their position.
- Other-self mentalizing: Politicians mentalize when speaking, to understand how the audience perceives their attitudes, self-image, and intentions.
- Mentalizing in Virtual World: Role-playing a character in a virtual game and understanding their motives and goals within the game's narrative.

Figure created with Adobe.

Bertenthal and Fischer (1978), which uses serial mirror tasks to measure the development of self-recognition in infants. In their studies, they put a mirror in front of infants, and saw 15-month-old infants could recognize the rouge dot on their noses and attempting to touch their noses. Since there is an objective basis of self, the content of self-representation began to increase exponentially from childhood to adulthood, encompassing self-beliefs (Valentine et al., 2004), self-esteem (Butler and Gasson, 2005), personality traits (Caspi et al., 2004), and other characteristics that define who we are. The development of self-identification abilities also provides a cognitive foundation for recognizing and understanding others. For instance, rapid self-recognition is equivalent to effectively identifying the non-self (i.e., others; e.g., Ma and Han, 2010). Moreover, as the cognitive and emotional processes

involved are similar among individuals, individuals can use their own experiences to understand others, particularly in terms of empathy (Buie, 1981). Researchers have extensively investigated self-relevant beliefs, delving into the makeup of the self-concept and its corresponding neural representations (Hu et al., 2016). The meta-analysis revealed that processing self-faces involves bilateral regions with a right hemispheric dominance, whereas self-referential judgments are associated with the anterior cingulate cortex/superior frontal cortex (ACC/SFC). Furthermore, the ACC and the left inferior frontal gyrus (IFG) extending to the insula were identified as central components of the self.

As we explore our sense of self, we become increasingly aware of others and their unique differences from us. These distinctions encompass sensory, cognitive, and emotional aspects. In the early stages, infants display distinct reactions to their own crying compared to that of other infants (Martin and Clark, 1982). This period marks the beginning of infants' development of a basic understanding of others, which lays the groundwork for more intricate social interactions in the future (Sodian, 2011). Through observational studies and experiments, researchers have discovered that even within the first year of life, infants demonstrate an innate sensitivity to the intentions and desires of others (e.g., Woodward, 2009; Yott and Poulin-Dubois, 2016). Infants exhibit a remarkable ability to anticipate and interpret the behavior of those around them (e.g., Cannon et al., 2012; Fawcett and Liszkowski, 2012). During the preschool years, children begin to exhibit the emergence of belief reasoning. The recognition of other's different beliefs by children has been widely confirmed during this stage (Wellman et al., 2001). Additionally, during adolescence, the social reasoning abilities and social brain functions of individuals are further enhanced and refined (Burnett et al., 2011; Peterson and Wellman, 2019). This emergence of complex social cognition and belief attribution plays a crucial role in their social interactions, relationships, and decision-making processes (Kilford et al., 2016).

The brain organizes representations of others based on their connection to our own identity, with distinct clusters for the self, social network members, and celebrities (Andrea and Meghan, 2020). Neurological patterns associated with self-recognition differ from those involved in recognizing others, highlighting a perceptual distinction (Platek et al., 2008). Understanding the distinction between self and other is crucial for comprehending social cognition and interpersonal dynamics, enabling us to recognize that others possess mental states that may differ from our own, even within intimate relationships (Itzchakov et al., 2022).

The way to capture other's beliefs

Theory of mind and mentalizing

Theory of Mind (ToM) and mentalizing have originated from different psychological directions but share similar definitions. ToM, emerged from evolutionary and developmental psychology, refers to individuals' ability to understand and interpret the mental states of self and others. ToM was initially proposed by Premack and Woodruff in 1978, who asked whether chimpanzees possess the ability to understand the hidden mental states behind others' behaviors. This concept refers to cognitive capacity to attribute mental states, paving the way for understanding intentions, beliefs, and desires (Premack and Woodruff, 1978). Mentalizing, on the other hand, originated from clinical psychology, refers to a capacity to recognize and understand the mental states within oneself and others, especially in explaining feelings and thoughts (Bateman and Fonagy, 2019). From a conceptual standpoint, ToM and mentalizing share significant similarities, which incorporates higher cognitive processes involved in inferencing the mental processes of oneself and others.

One of the classic methods used to assess ToM is the false belief task, which assesses individuals' understanding of others' intentions and knowledge (Gopnik and Slaughter, 1991). The false belief refers to a belief that does not align with reality or the available evidence. In this task, the main idea is to present a story scenario, whereby manipulating the presence or absence of the false belief in the character, two beliefs from the character's perspective and the observer' perspective are constructed. One representative task typically administered to children involves presenting them with a scenario involving two characters, often named Sally and Anne (commonly known as the Sally-Anne task; Baron-Cohen et al., 1985). The child observes Sally hiding a ball in location A. Subsequently, Anne moves the ball from location A to location B while Sally is absent. The child is then asked where Sally will look for the ball when she returns. A correct response requires an understanding that Sally holds a belief about the object's location (a false belief) and will search for it where she last saw it (action prediction).

Researchers have developed various engaging tasks, including sensory stories, cartoons, and visual perspective tasks, to assess children's ToM abilities (see **Table 1**). For example, the ability to understand false beliefs emerges early in childhood, with infants as young as one year old demonstrating awareness of false beliefs in simple tasks such as false photographs or pretense acting (Kovács et al., 2010; Onishi et al., 2007; Scott and Baillargeon, 2017). Infants exhibit sustained attention when someone holds thoughts that differ from reality. As children develop attentional and linguistic skills, they can successfully navigate more complex false belief scenarios presented in question-and-answer story formats. By around the age of four, children acquire a stable ability to consider "what others think" (Lewis and Osborne, 1990; Rubio-Fernández and Geurts, 2012; Surian and Leslie, 1999). However, according to Quesque and Rossetti (2020), the paradigm that really measures ToM needs to meet the following two points: (a) success should be attributed to mental state understanding rather than lower-level processes (mentalizing criterion), and (b) the task should require distinguishing between one's own mental state and others' mental states (non-merging criterion).

Cognitive mechanisms underlying representation of others belief

Several cognitive mechanisms contribute to our ability to develop a representation of others' beliefs, such as perspective-taking, mental state attribution (Spengler et al., 2009), cognitive flexibility and self-awareness, including meta-cognition.

Table 1 Typical tasks measured Theory of Mind.

Task form	*Task name*	*Aims*	*Description*
Story sensoria	Sally-Anne task (e.g., Wimmer and Perner, 1983)	False beliefs attribution	Watching a puppet show and inferring the next movement or beliefs of others (who has a false believe)
	Faux-pas test (e.g., Baron-Cohen et al., 1999)	Faux-pas detection	Reading stories and answering whether someone in the story said something he/she should not have said
	Strange stories (e.g., Happé, 1994)	Action explanation	Reading stories and answering "why" questions for characters in the story
	Expected and unexpected test (e.g., Onishi et al., 2007)	Pretense detection	Seeing an experimenter with true belief but behaving wrongly
Cartoon	Comic strips (e.g., Sarfati et al., 1997)	Action prediction	Choosing the next scene of a comic strip
	Humorous cartoon (e.g., Happé et al., 1999)	Action explanation	Seeing a cartoon about misunderstanding and explain why
	Triangle task (e.g., Abell et al., 2000)	Action explanation	Seeing a big triangle interacting with a kid triangle and telling a story about them
Visual perspective taking	Piaget's three-mountain task (e.g., Piaget and Inhelder, 1967)	Perspective taking	Describing a look from the perspective of another person on different side
	The director task (e.g., Wu and Keysar, 2007)	Perspective taking	A "director" instructs another person who does not share the same view to move certain objects
	Picture test (e.g., Hegarty and Waller, 2004)	Perspective taking	Choosing the correct scene when standing in another person' shoes will see

Behavioral tasks are typically used to assess the capacity of Theory of Mind. In the Story sensoria task, participants are typically given a series of short stories or scenarios to read or observe. The task aims to assess an individual's understanding and interpretation of mental states, such as beliefs, desires, and intentions, in others. During the task, participants are typically asked questions or prompted to make judgments about the characters in the stories based on their mental states. They may be asked to infer what a character is thinking or feeling, predict their behavior, or make inferences about their intentions. In the "Cartoon" task, participants are typically presented with a series of cartoon-like images or short animations depicting social interactions between characters. During the task, participants are often asked questions or prompted to make judgments about the characters' mental states based on the visual information provided. They may be asked to infer what a character is thinking, feeling, or intending to do in each situation. They need to use this information to make accurate interpretations and predict the next scenario, and about the mental states of the characters. The "Visual Perspective Taking" task is commonly used to evaluate an individual's ability to engage in egocentric or decentered perspective-taking. For example, in the Three Mountains task, participants need to accurately ascertain what a scene would look like from another person's spatial viewpoint.
Table adapted from Quesque and Rossetti (2020).

Perspective-Taking. Perspective-taking is a fundamental cognitive process involved in representing others' beliefs. It refers to the ability to adopt someone else's viewpoint and understand their subjective experiences (Davis, 1983; see **Box 1**), e.g., we put ourselves in someone else's shoes. It allows individuals to consider alternative mental states, helping them anticipate and interpret others' beliefs accurately. Study shows that people can sense and react to how others perceive the appearance of an object or number

Box 1 Representation of Others' Beliefs

Mentalizing
The ability to understand and interpret the thoughts, intentions, beliefs, and emotions of oneself and others, and to use this understanding in social interactions.

Theory of Mind
Theory of Mind refers to the ability to attribute mental states (beliefs, desires, intentions etc.) to oneself and others, and to understand that these mental states can differ from person to person.

Intentionality
It refers to the capacity to attribute intentions or purpose to-actions. Recognizing intentionality is an essential aspect of mentalizing as it helps in understanding other people's actions and predicting their behavior.

Empathy
Empathy is the ability to understand and share the feelings of others. It involves putting oneself in someone else's shoes and experiencing their emotions, which is closely linked to mentalizing.

Perspectives-taking
Perspective-taking is the act of perceiving a situation or understanding a concept from an alternative point of view. This may involve adopting the perspective of another person or that associated with a particular social role.

(e.g., "6" or "9"), whether they share or conflict with that perspective (Surtees et al., 2016). Research has demonstrated that perspective-taking skills positively correlate with empathy and prosocial behaviors, indicating its significance in social interactions (Gutsell and Inzlicht, 2010).

Mental State Inference. Mental state inference involves making inferences about others' beliefs based on available cues, going beyond observable behaviors (Wellman et al., 2001). Facial expressions, vocal tone, body language, and verbal statements provide cues for accurate belief attribution (Bartsch and Wellman, 1995). This ability to attribute mental states, including beliefs, desires, intentions, and knowledge, to oneself and others has been studied since Heider and Simmel (1944) famous silent animation experiment. Recognizing that others have separate thoughts and beliefs allows us to understand their behavior and predict their actions based on their beliefs. In social interactions, high-level attribution becomes more crucial as complexity and depth of reasoning increase.

Cognitive flexibility. Cognitive flexibility, crucial for understanding others' beliefs, involves adapting thinking and behavior to new situations (Diamond, 2013). It includes task switching, creative thinking, and considering multiple perspectives (Uddin, 2021; Zühlsdorff et al., 2023). Cognitive flexibility allows adjusting beliefs and perspectives based on new information (Kienhues and Bromme, 2011), accommodating others' beliefs, and adapting mental representations. Mentalizing requires dynamic perspective shifting and understanding diverse mental states. Cognitive flexibility facilitates this by enabling consideration and switching between viewpoints and interpretations of situations.

Self-awareness, including **metacognition.** Self-awareness involves recognizing one's mental states and their influence on interpreting others' mental states (Boekaerts, 1999; Gilovich et al., 1998). While people are generally aware of their internal states, accurately estimating one's beliefs about others can be challenging. Meta-cognition, a key concept in cognitive psychology (Pasquali et al., 2010). Meta-cognition refers to the ability to think about and reflect on one's own cognitive processes, such like "thinking of thinking" (Carruthers and Chamberlain, 2000), as well as understand and reason about the mental states of others (Livingston, 2003). It enables us to notice the representation in our mental states. Meta-cognition allows us to observe our mental representations and distinguish them from others', providing higher-level cognitive supervision for representing others' beliefs (Jiang et al., 2022).

Roles and functions of core brain regions in mentalizing

The neural mechanisms underlying ToM and mentalizing involve a network of brain regions known as the **mentalizing network.** The mentalizing network refers to a collection of brain regions and neural circuits involved in understanding and attributing mental states to oneself and others. Frith and Frith (2003) defined the mentalizing system based on various paradigms that required participants to consider the mental states of others (e.g., reading stories, watching moving shapes, or playing interactive games). This network includes the medial prefrontal cortex (mPFC), temporoparietal junction (TPJ), superior temporal sulcus (STS), and posterior cingulate cortex (PCC; Frith and Frith, 2008), which brain regions interact to facilitate mental state inference, perspective-taking, and empathy. Molenberghs et al. (2016) conducted a meta-analysis on 144 fMRI studies using ToM tasks and depicted the brain network associated with mentalizing (see **Fig. 2**). The analysis revealed significant clusters in several regions, including mPFC extending into the medial orbitofrontal cortex and ACC, the precuneus, bilateral regions extending from the temporal pole into the posterior superior temporal gyrus and TPJ, as well as bilateral regions in the IFG.

Frith and Frith (2006) conducted a comprehensive review of the core brain regions involved in the mentalizing network and provided detailed insights into the potential functions and roles of each region. Specifically, posterior STS appears to play a crucial role in perspective-taking, as it tracks the eye movements of others and helps in representing the world from different visual perspectives (Aichhorn et al., 2006; Pelphrey et al., 2004).

TPJ serves as a convergence point for information about the social context of others, enabling us to make moment-to-moment inferences about their mental states. Quesque and Brass (2019) highlighted the importance of TPJ in distinguishing between self and other and suggested its involvement in controlling self and other representations in a domain-general manner. TPJ and the ACC consistently respond to mental states and can differentiate between information related to oneself and others (Schurz et al., 2014).

PFC areas also play multiple roles in mentalizing. Firstly, they are involved in attention and executive control processes when selecting between false and true beliefs (Hartwright et al., 2012; Sommer et al., 2007). Secondly, they are responsible for anticipating others' thoughts and adjusting our beliefs about them when their behavior does not align with our expectations (Burke

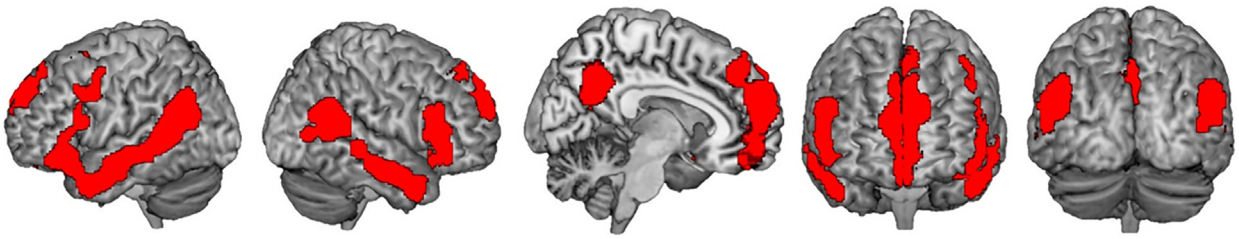

Fig. 2 Meta-analysis results of fMRI studies on theory of mind tasks (Molenberghs et al., 2016).

et al., 2010; Grèzes et al., 2004). The PFC serves as a key node in the social brain network, facilitating perspective-taking and self-referential processing (Mitchell et al., 2005). Ereira et al. (2020) found that the ventromedial prefrontal cortex (vmPFC) may monitor the identification of self and others and distinguish between self- and other-attributed signals. Additionally, the mPFC is implicated in metacognition, particularly in encoding decision confidence (Bang and Fleming, 2018).

During retrospective judgments of confidence, the lateral prefrontal cortex (lPFC) and ACC are activated, which may be related to performance monitoring function (Fleming and Dolan, 2012). In our brain model of representation of others, the lPFC and ACC receive input mental states from both oneself and others. By accessing metacognition, these two meta-level regions monitor and regulate confidence levels regarding mental states. For example, the ACC aids in detecting matched or mismatched signals between predicted intentions or desires of others and their actual outcomes. With input from the dorsolateral prefrontal cortex (DLPFC) to other mentalization brain regions, individuals adjust their inferences adaptively, which complements metacognition. According to the "simulation" and "projection" account, ToM and metacognition networks closely interact with each other (Cristiano et al., 2023; Jiang et al., 2022; Suzuki, 2022).

Additionally, studies investigating ToM using tasks like the photographs, videos, and animation tasks (excluding stories, cartoons, and interactive games) have consistently shown activation in the posterior IFG and adjacent premotor cortex. This activation pattern suggests that the tasks involving identification of observed actions or emotional expressions may impose demands on a simulation mechanism. This mechanism enables individuals to mentally simulate and understand the intentions, emotions, and mental states of others, contributing to their ToM abilities (Molenberghs et al., 2016).

As indicated, brain regions in the social brain network, play key role of in mentalizing (Wu et al., 2020). For instance, the default mode network (DMN) in recent models of belief representation for self and others, and shared reality building and representation that varies as a degree of similarity and closeness of others (e.g., Yeshurun et al., 2021). The DMN is considered as brain network involved in self-referential processing (Murray et al., 2012), and autobiographical memory (Zaki et al., 2009). Its link with empathy and ToM has also been demonstrated in fMRI studies, see a review of empathy and ToM (Schurz et al., 2021).

The integrative framework: interaction between self-representation, social representation and other's belief representation

Role of self-representation

When interpreting others' beliefs from a social perspective, it's crucial to consider self-representation, social influence, and cultural context (**Fig. 3**). Self-representation refers to how individuals perceive themselves and understand their thoughts, beliefs, and emotions, shaping their identity and influencing social interactions (Gergely, 2002). It can be actively updated during social exchanges, serving desirable social outcomes (Moutoussis et al., 2014).

Self-representation also involves inferring others' internal states by observing their behavior and connecting them to one's own mental states. Compelling evidence to support this viewpoint is the significant correlation between self-conscious emotions and

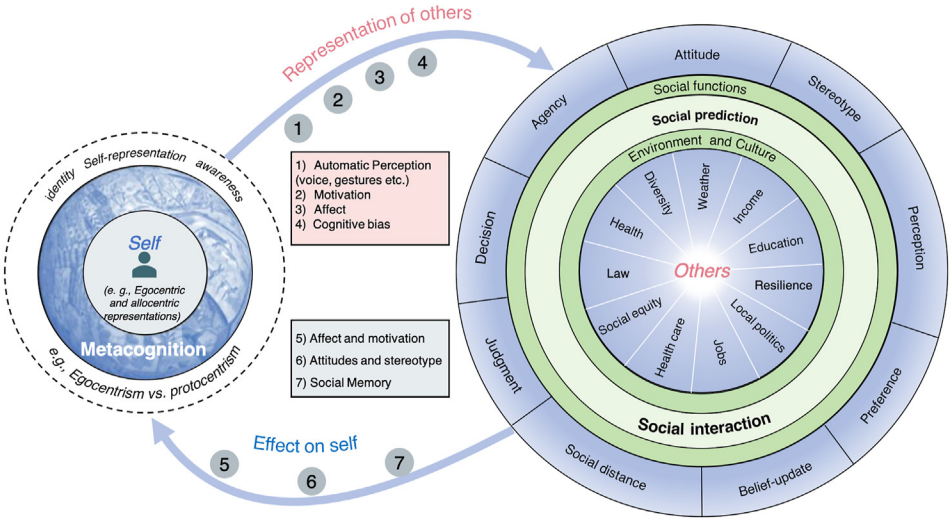

Fig. 3 The diagram of self-representation and representation of others.

Schematic illustration of interplay between an individual's self-representation and their representation of others. The central circle represents self-awareness and perception, while the surrounding circles symbolize external individuals or social groups. Lines connecting the circles depict the reciprocal influence of information, thoughts, and emotions. This interplay showcases how our self-representation shapes our understanding of others, influencing social interactions and relationships. The figure also highlights influential factors like cultural background, experiences, context, cognition, personality, emotions, and interpersonal dynamics. These factors impact how individuals perceive others and represent themselves. The figure emphasizes the complexity and multi-dimensional nature of this interaction, illustrating how these factors intertwine and impact social interactions and relationships.

others' opinions of us, indicating that our perception of ourselves has an impact on our inference to others (Leary, 2004). For instance, we may feel highly embarrassed when others inaccurately perceive us unfavorably, such as lying. A study suggests that liars tend to overestimate the detectability of their lies and has shown that people often overestimate the visibility of their feelings of disgust compared to how noticeable they actually are (Gilovich et al., 1998). This framework suggests that self-related variables, like desires and the surrounding world, can impact other-representation variables, such as emotions. Thus, even with limited knowledge, observers can make assumptions about others' beliefs that are not explicitly known.

Theory of social representation

Social representations, as proposed by social psychologists like Serge Moscovici, indicate that our understanding of the world is heavily influenced by the shared beliefs, values, and norms of the social groups we belong to (Moscovici, 2001). The role of culture within this theory of social representations is critical, as culture shapes the way individuals perceive and interpret the world around them, and it provides the social frameworks and collective meanings that influence social representations. These collectively held beliefs function as a cognitive framework through which individuals interpret and make sense of the thoughts, intentions, and behaviors of others.

Social representations are not just individual beliefs but rather are co-constructed and shared within a community or society (Howarth, 2001). They are shaped by various social factors, including social interactions, communication patterns, media, and dominant discourses. For example, media representations of certain groups or events can significantly influence how people perceive and understand the beliefs of others (Nelson et al., 1997). These shared social representations create a common understanding and provide a basis for interpretation in social interactions (Wu et al., 2016).

Integrative framework: representation of other's belief

By considering both individual internal process, and social context and cultural factors, we gain a deeper understanding of how representation of others' beliefs is socially constructed. We recognize that our understanding of others' minds is not solely based on individual cognitive processes but also on the influence of social groups and cultural contexts. This social perspective highlights the dynamic and contextual nature of representation, emphasizing the importance of considering social influence, cultural norms, and collective interpretations in understanding others' minds. In this regard, the emergence of a socioecological framework has rapidly gained traction, as it offers insights into the varying rates of formation for representations of other people's beliefs—both quick and gradual (Oishi, 2014; Oishi and Choi, 2017; Oishi and Graham, 2010).

Cultural context plays a pivotal role in the representation of others' beliefs, as belief of others are chargeable and can be specific to a person, a group, a religion or a culture (Hansen and Ryder, 2016). Culture encompasses the shared values, norms, attitudes, and practices of a particular group or society. It shapes our worldview and influences cognitive processes, including how we perceive and interpret the thoughts of others. Firstly, culture is associated with the timing of belief representation acquisition. For example, children in collectivist cultures exhibited a slight delay in successfully grasping false beliefs compared to their counterparts in individualistic cultures (Liu et al., 2008; Wellman et al., 2006). This difference may be attributed to variations in children's sociocultural experiences. Secondly, there are cultural variations in the sequential development of belief representations. More specifically, children from collectivist countries displayed a later understanding of diverse belief in sequence than did children from individualist countries (Shahaeian et al., 2011; Wellman et al., 2006). This may be because collectivist cultural norms emphasize mutual agreement, group harmony, and cohesion, which necessitate shared beliefs or actions. Furthermore, cultural similarity often leads individuals to represent other's belief more accurately and frequently (Selcuk et al., 2023). For instance, children had better mentalizing accuracy when it comes to their own group members compared to members of other groups (Glidden et al., 2021). Additionally, they use diverse mental state terms more frequently when describing their own cultural group than when describing other cultural groups (McLoughlin and Over, 2017).

The belief representation of others is shaped by internal psychological processes and the broader socio-ecological environment (**Fig. 3**). While understanding human perception, emotions, and behavior often emphasizes internal processes and brain activation, it tends to overlook the impact of the macro-level natural and social environment. Individuals interact with and make decisions within specific contexts, and the objective environment significantly influences their perception, judgments, and behavioral patterns.

The socio-ecological environment encompasses the natural and societal components of human living. It includes economic and political systems, education, demographics, geography, climate, and religion. Moreover, elements like cities, communities, housing, and family relationships contribute to the social and ecological environment. This environment gradually shapes individuals' beliefs, which can be inferred by others in social interactions. To interpret the representation of others' beliefs, we must consider self-representation, social influence, and cultural context. Social representations, influenced by social interactions and cultural norms, provide a shared cognitive framework for understanding others' beliefs. By examining these social factors, we gain a comprehensive understanding of how beliefs are socially constructed and influenced by collective values, norms, and social dynamics.

In conclusion understanding the representation of others' beliefs requires acknowledging the interplay between internal processes and the socio-ecological environment. By considering the broader social context, we can grasp how beliefs are shaped and shared within a society, contributing to the complex dynamics of social interactions.

The way to update the representation of other's beliefs

Research indicates that the use of probabilistic models based on Bayes' rule is becoming more common in exploring the representation of others. These models offer a flexible and complex way to understand representation update in human cognition. This article delves into the key mechanisms of Bayesian information processing and provides it as an example of how Bayesian approaches contribute to the representation of others' beliefs. We provide an overview of Bayesian modeling (**Box 2**), followed by a review of how Bayesian modeling has been linked to representation of other's beliefs. The significance of prior knowledge and active learning through social interaction is also highlighted.

Constantly updated beliefs of others from observation

Beliefs of others arise from both static and dynamic interaction settings. Traditional ToM research has focused on assessing individuals' ability to represent others' beliefs based on established facts. This kind of scenario-based approach involves assessing their capacity to understand and interpret the static beliefs of the characters. However, in real-life situations, beliefs are not static and can change due to evolving of events and other context factors (Yeager and Dweck, 2020).

With the dynamic updating perspective, the work by Baker et al. (2011) has proposed a framework for Bayesian ToM (BToM). This framework aims to capture the dynamic nature of belief inference processes in real-world contexts. The Bayesian framework acknowledges that individuals continuously update their representations of others' beliefs based on new information and contextual cues. Based on BToM, belief inferences are not binary, but continuous and probabilistic, and allow for quantitative variability in performance. Based on continuous observable behavior variables, observers can deduce the psychological state of actors that cannot be directly observed with posterior probability based on constantly updated experience.

Based on the BToM framework, observers' inferences are supported by probabilistically inverting a generative model of approximately rational agents perceiving, planning, and acting in a dynamic world. For example, Baker and et al. (2017) defined how an agent behaves in a grid world setting using a partially observable Markov decision process (POMDP) and prioritizing the rational perspective of maximizing expected utility. More recently, a similar model is proposed, the CogToM model (Nguyen and Gonzalez, 2022). The CogToM model shares similarities with the BToM framework as it observes an agent's actions in a gridworld environment and proposes a dynamic model that develops ToM from the observation of the other agent's actions. Usually, models that assume accurate observations or base their calculations on utility maximization tend to deviate from human behavior in decision-making tasks. However, humans are boundedly rational, and their decisions are constrained by cognitive capabilities for storing and retrieving information from memory. CogToM offers a cognitive model of the observer, and its predictions align with human observers.

Constantly updated beliefs of others from active interaction

The next level is the active updating level, which involves combining individual social interaction (e.g., interactions with others in social contexts) with updating individual beliefs, social knowledge, interaction motivation, and more (e.g., through memory, inferences, and social decision-making). This level, at least in part, relies on active social interaction and inference processes.

Box 2 Bayesian Model

Bayesian

Bayesian is a denoting an approach to statistical inference and probability that enables previously known (a priori) information about a population characteristic of interest to be incorporated into the analysis. In Bayesian methods, estimated quantities are based in part on empirical data (i.e., what was actually observed) and in part on collective or individual knowledge about what to expect in the population (as captured in a prior distribution).

Bayesian inference

Bayesian inference is a method of drawing conclusions about a characteristic of a population using both sample data and previously known information about that characteristic. It relies on bayes theorem to derive posterior distributions from obtained observations and prior distributions.

When $P(\Theta)$ is used to represent the prior probability, $P(y|\Theta)$ represents the probability of a particular outcome based on the prior of the parameter Θ, which is known as the likelihood, after the data y have been collected. And $P(y)$ represents the overall probability, which is independent of the parameters. Therefore, according to the principle of Bayes' theorem it can be concluded that $P(\Theta|y)$, which represents the posterior, is: $P(\Theta|y) = (P(y|\Theta) \times P(\Theta))/P(y)$.

Bayesian Theory of Mind

Bayesian theory of mind (BToM) is a theory-based Bayesian framework, which models the structured knowledge of human ToM at multiple levels of abstraction, representing the ontology and principles of the theory at an abstract level and mapping these concepts to particular domains and contexts to generate specific predictions and inferences.

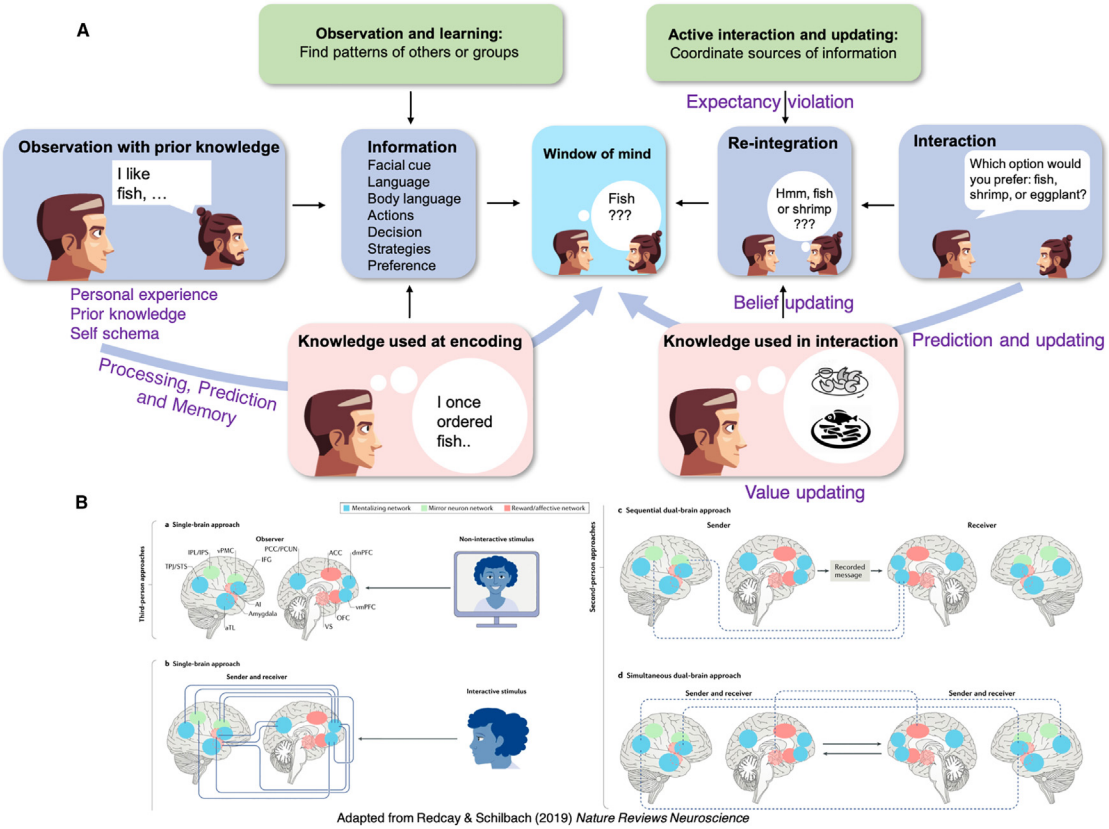

Fig. 4 Observation and active interaction in the formation of representations of others.

(A) Schematic illustrations of observation and active interaction facilitate the formation of representations of others. One example of observation and active interaction for the formation of representations of others can be seen in a social setting where two individuals, Mark and Ben, are engaged in an observation or direct conversation. During the conversation, Mark actively observes Ben's voice, and other perceptual cues, allowing him to gain inference of his food preference and intentions. He pays attention to his facial expressions, tone of voice, and body language, which helps him understand his emotional state and possible goals. Additionally, both Mark and Ben's memories of previous interactions influence their representations of each other. As they continue interacting, Mark forms a representation of Ben based on her observations and integration of memory. For example, Mark notices that Ben is speaking passionately about fish, which helps him infer his strong positive preference towards it. Additionally, both Mark and Ben's direct interaction can form memories and belief of in the interaction, which influence their representations of each other. During the interaction, recalling past conversations and experiences allows them to deepen their understanding of one another and adjust their mental models based on new information. This example illustrates how observation and real interaction contribute to the formation of representations of others. By actively engaging with perceptual cues, attitudes, actions, and memories, individuals can build more accurate and comprehensive mental models of those around them.

(B) Schematic illustrations of brain regions involved in observation and active interaction in the formation of representations of others. Adapted from Redcay and Schilbach (2019).

For example, take the scenario illustrated in Fig. 4A: when dining with a guest who refuses to eat a dish due to a specific ingredient that he finds repulsive, while you remember that he likes it. This situation might prompt you to inquire whether his food preferences have changed or if there are other factors influencing his reactions or preferences towards certain ingredients.

In the real interaction, changeable mental states of others can be understood through observation, learning, or possibly more rapidly through active interaction and updating, directly and indirectly (Wu et al., 2020).

Beliefs updating in game theory

With the development of decision neuroscience, an increasing number of studies investigate the mechanisms underlying social interactions, using tasks adapted from game theory. For example, some studies use the coordination game in which two players get positive payoffs if they both choose the same action collaboratively. This involves both mentalizing and meta-mentalization (Wu et al., 2020) in interacting minds (Camerer, 2011; McCubbins et al., 2012). When tested experimentally, people collect information and shift their decisions in strategy games such as the trust game across rounds, which indicates meta-mentalization is necessary for high level social interactions (Bhatt et al., 2010; McCabe et al., 2003). In the bargaining game, successful strategic deception relies on the confidence that a perceiver's thinks and feels about themselves and others.

Yoshida et al. (2008) adopted ideas from optimal control and game theory and provided a computational model for "game theory of mind," using "recursive sophistication." This entails self modeling of others' goals (ToM), and modeling of others' perceptions of self-goals, and so on to infinitum. Although this recursion can go to infinite levels, human behavior can often be modeled well by three levels, as humans have a limit on the degree of recursion, an example of "bounded rationality" (Kahneman, 2003a, 2003b; Simon, 1955).

Considering a Markov environment, the transition probability $p(S_{t+1} = i|S_t = j, v)$ is the probability of state j going to state i with a value v. The value v is the accumulative reward expected in the future. Based on optimal control theory and dynamic programming, it is

$$v(j) = l(j) + \sum_{i=1}^{n} v(i)p(S_{t+1} = i|S_t = j, v)$$

The policy is fully specified by value with logit distribution:

$$P(v)_{i,j} \propto P(0)_{i,j} e^{\lambda v(i)}$$

Where $P(0)_{i,j}$ is the probability of state j state i occurs autonomously and λ is the prevision parameter indicating the sensitivity to value differences. A level-k agent assumes the opponents adopt level-(k-1) strategies and selects actions accordingly (Yoshida et al., 2008).

Beliefs updating in the Bayesian framework

Based on the Bayesian framework and previous models, we propose the Bayesian inference model for mentalizing, where people integrate prior knowledge of oneself (K_S) and update their belief (Bel) with the evidence (E) from each interaction/communication. By this Bayesian inference model, $Bel_{t+1} = Bel_t * E_t$; the Bayesian update is $Bel_{t+1} - Bel_t$; and the posterior distribution is $(Post(t)) \propto prior * likelihood$. In mentalizing, people can either infer about others based on evidence ($Bel = E$), or totally simulate others based on self-knowledge ($Bel = K\_S$). Some work suggests that people are biased by their prior knowledge about themselves during mentalization, integrating both Bayesian updating and one's prior knowledge. This conforms to the model:

$$Bel(t) = (1 - a)Post(t) + aK\_S(t)$$

Regarding the representation updating and meta-cognition, we propose several alternative models, since people integrate inferences of others and others' mind to their own as well. Adapted from the existing computational models in social learning, social inferences, and social prediction, people first implement the reinforcement learning model for these components. Specifically, people compute the prediction error (i.e., the difference between the prediction about other agent's action and the actual outcome) and the success of the social interaction ($1 =$ successful, $-1 =$ unsuccessful) during each current trial t+1 by $V(t+1) = V(t) - a * PE(t)$. Notably, people's self-evaluation (metacognition or confidence) shifts upon the prediction and outcome, while their beliefs or mind-changes rely on the controllability of their own and the other's mental states. One important factor regarding metacognition in the mentalizing processes is the egocentric bias, i.e., the advantage of people's own perspective or self-reference effect leading to the exaggeration of their role in a situation. Such a bias can manifest on several aspects of mentalizing. For example, it is easier to perform first-perspective mentalizing, than conformational mentalizing.

An interesting view of interactive mentalizing is the idea of social Bayesian brain, which refers to the brain obeying the Bayesian rule (learning about others) with the priori assumption that others are Bayesian too (i.e., others also learn about us; Daunizeau et al., 2010). Research has proposed a k-ToM model, which predicts that the performance of agents engaged in competitive repeated interactions increases with their ToM sophistication (Devaine et al., 2014). In accordance with Yoshida et al. (2008), the ToM sophistication in the K-model was defined as the depth of recursive thinking. For example, 1-ToM is defined in terms of one's own recursive belief, i.e., the belief about 0-ToM's belief, and a k-ToM agent tries to understand how the opponent agent learns, with the assumption that the other agent is less sophisticated than oneself. In doing so, k-ToM forms high-order recursive beliefs.

In **Fig. 4**B, there is a demonstration of the different ways of representation formation and updating, with brain dynamic map of mentalization during social interaction, specifically highlighting the observation brain and interactive brain. For more detailed information, refer to the review paper on Redcay and Schilbach (2019).

Irrational updating beliefs

When updating the representation of others' beliefs from a social perspective, it is crucial to recognize the impact of social influence and other factors (e.g., prior belief, size of social influence, and cognitive bias etc.) that affect the updating process (Connors and Halligan, 2015). For example, as aforementioned, social representations indicate that our understanding of the world is heavily influenced by the shared beliefs, values, and norms of the social groups we belong to. These collectively held beliefs function as a cognitive framework through which individuals interpret and make sense of the thoughts, intentions, and behaviors of others. Usually, people can absorb external information and update the belief of others accurately through various ways, including social interactions, communication patterns, media, and dominant discourses (Choudhry et al., 2016; Modgil et al., 2021). These representations are not just individual beliefs but rather are co-constructed and shared within a community or society, and shaped by

various factors, which may not be updated in a rational way. For example, representations of mental state of certain one or groups or events can significantly affect by the strength of the prior belief.

Imagine the following situation in Fig. 5, the task involves guessing what others are working on, either a geometry or calculation assignment. It demonstrates the impact of influence from individuals and groups, with a focus on their size. The figure also highlights the strength of priors and the impact of irrational learning. To guess what someone is working on, one can ask classmates for information. If they confirm it's a geometry assignment on circles, it strengthens the belief. However, new information should be considered to update beliefs accurately. Some individuals may stubbornly stick to their initial beliefs, leading to irrational updating.

When individuals have strong prior beliefs, and cognitive bias, it can be challenging to update their beliefs accurately. This can lead to irrational updating, where individuals continue to hold onto their initial beliefs despite new information and cannot overcome the confirmation bias (Bang and Frith, 2017). For instance, if someone strongly believes that a particular political candidate is the best option, they may ignore or dismiss any new information that contradicts their belief. They may even interpret the new information in a way that reinforces their prior belief, resulting in irrational updating (Taber and Lodge, 2006). This can be problematic as it can lead to a lack of critical thinking and an inability to consider alternative perspectives. It is essential to recognize the impact of strong priors on belief updating, adjusting metacognition and strive to remain open-minded and receptive to new information (Moore and Healy, 2008).

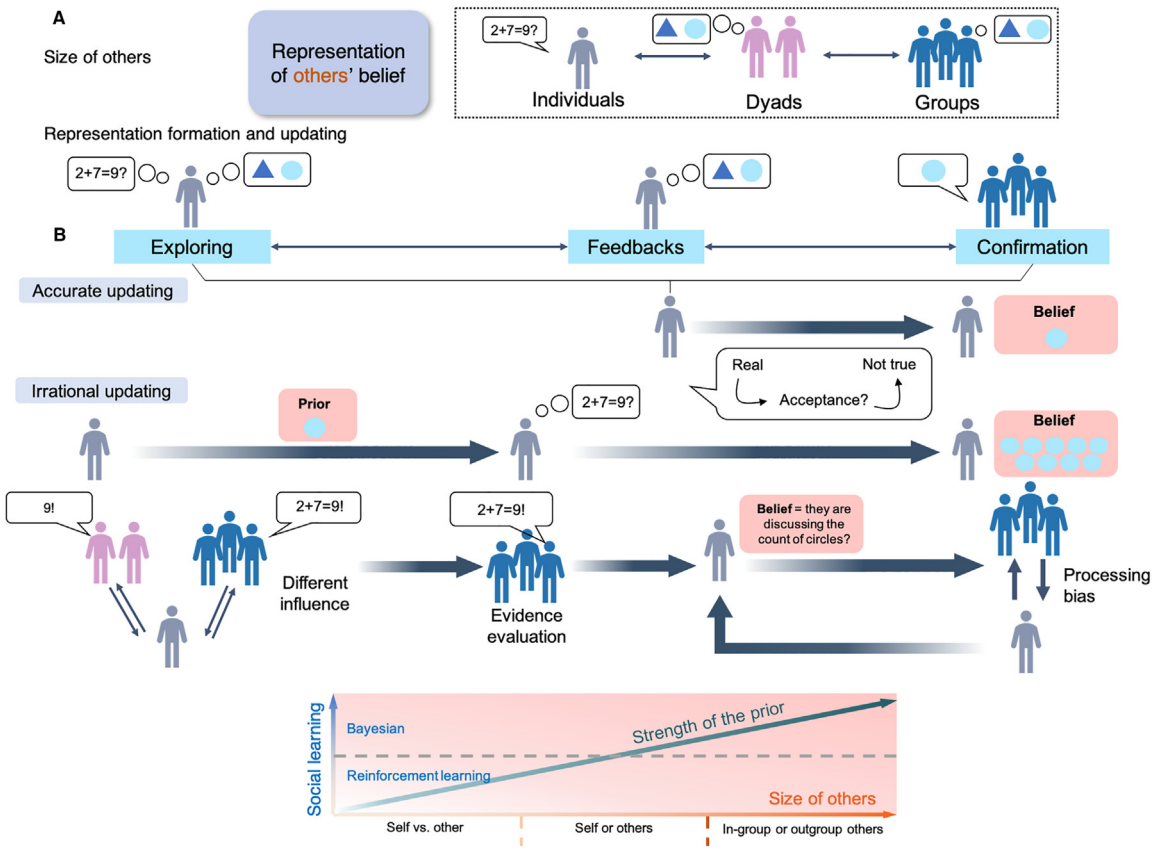

Fig. 5 Representation Formation and Updating: Influence from Others, Strength of Priors, and Irrational Learning Process.

The example task involves guessing the assignment of others, such as in a calculation or geometry assignment. (A) It demonstrates the essential role of influence from other individuals and groups, with a particular focus on the impact of their size; (B) This figure further illustrates the dynamics of strength of priors and the influence of an irrational learning process.

In this scenario, we have a situation where individuals are trying to guess what another person is thinking about: whether it is a geometry assignment related to shapes or an assignment related to number operations.

To build a belief about the other person doing geometry homework on circular shapes, one can ask someone in the class about the nature of the assignment. If that person confirms that the assignment is about shapes, specifically circles, then it strengthens the belief that the other person is indeed working on geometry homework related to circular shapes. However, even after forming this belief, it is possible to hear information from different sources and classmates on a new day about homework calculations. For example, there may be a discussion about whether 2 plus 7 equals 9. In such a situation, it is essential to consider the new information and update one's beliefs accordingly (accurate updating). However, some individuals may remain stubborn and continue to believe that the other person is discussing how many circles there are (i.e., irrational updating), rather than updating their beliefs to recognize that the assignment is about arithmetic rather than shapes. Adapted from Kwon and Telzer (2022).

Application

The representation of others' beliefs is influenced by various factors, including individual differences, situational factors, cognitive processes, emotional factors, and cultural factors. These factors play a crucial role in shaping how individuals understand and interpret the mental states of others and show individual variations in different populations. For instance, order adults have poorer performance on mentalizing tasks compared to younger adults (Duval et al., 2011; Henry et al., 2013). Additionally, gender is associated with mentalizing abilities. Prior research has demonstrated that male adolescents performed poorly on mentalizing tasks and made more hyper-mentalizing errors than female counterparts (Poznyak et al., 2019). Furthermore, socioeconomic status reflecting an individual's economic and social position relative to others, is linked to increased mentalizing-related neural value coding (Schulreich et al., 2023).

Here, one area of interest lies in exploring the implications of mentalization for various populations, it sheds light on both the stability and plasticity of mentalization and its implications for adaptive social behavior. It is interesting to understand the development of metacognition in both typical and atypical populations and how it influences social functioning across the lifespan. Additionally, researchers are keen on examining communication, decision-making processes and mentalization in subclinical and clinical samples. These investigations offer valuable insights into the interplay between mentalization, cognition, and social behavior.

Healthy

ToM plays a pivotal role in facilitating social interaction by enabling individuals to infer others' perspectives, emotions, and intentions. By employing ToM skills, people can navigate complex social situations better by considering others' beliefs and adjusting their behavior accordingly. The application of ToM promotes empathy and understanding, leading to more prosocial behaviors (Pavey et al., 2012; Reynolds and Scott, 1999). It allows individuals to anticipate and respond appropriately to the needs and desires of others, fostering trust and cooperation (Elliott et al., 2006; Etel and Slaughter, 2019; Fett et al., 2014).

Additionally, our ability to understand and connect with others plays a crucial role in communication. Developmentally, language and mentalizing ability are interdependent and support social communication (Miller, 2006). To facilitate smoother communication of social information, people must use others' viewpoints to guide their own communicative behaviors. Several studies support the idea that mentalizing ability contributes to effectively communicating information in conversations (Sidera et al., 2018). Conversely, individuals with difficulties in understanding others' minds experience challenges in social communication. McDonald and Flanagan (2004) conducted a study involving adults who had experienced severe traumatic brain injuries, most of whom exhibited injuries in frontal lobe regions. These individuals demonstrated significantly lower abilities in inferring the intentions of others, leading to difficulties in understanding social conversations.

Furthermore, the ability to read others' beliefs not only helps us establish automatic social causal inferences (i.e., understanding that behavior is based on specific intentions) but also provides us with the ability to predict and even evoke desired actions and reactions in other people while avoiding undesired ones (Ho et al., 2022). Planning involves considering the effects of different possible actions to select the action that maximizes expected value (Daw et al., 2011; Sutton and Barto, 2018). To accomplish this, planning requires a causal model that specifies the asymmetric dependence between causes and their effects, revealing what to expect from one variable following a potential intervention on another variable.

Autism

Although representing other people's beliefs or understanding their ideas is a basic ability for neurotypical individuals, it is challenging for people with autism spectrum disorder (ASD). Autism is one of the most debilitating developmental disorders, characterized by a profound lack of social understanding.

In 1943, autism was first recognized as a distinct disorder by Kanner. His seminal clinical description involved 11 boys with a condition characterized by a deficit in affective contact, known as autism (Kanner, 1943). Autistic individuals typically exhibit behaviors that differ from those of neurotypical individuals, such as preferring solitary activities, not engaging in play with other children, only participating when prompted and assisted by adults, and struggling to differentiate between "I" and "you" in verbal expressions.

The development of autism is strongly influenced by genetics (Muhle et al., 2004). Early evidence supporting the genetic nature of autism primarily comes from twin studies. In the case of identical twins who share the same DNA, if one twin has autism, the other twin has approximately a 60% chance of also being affected by the disorder (Bailey et al., 1995). Similarly, family members of individuals with autism are more likely to exhibit autistic-like traits compared to the general population. These traits may manifest as increased anxiety, impulsivity, aloofness, shyness, oversensitivity, irritability, or eccentricity (Murphy et al., 2000). Additionally, certain environmental factors, such as advanced parental age (Wu et al., 2017) and low birth weight (Lampi et al., 2012), contribute to the risk of autism.

Autism is characterized by significant impairments in social functioning. According to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), autism is characterized by impairments in social functions, including social communication, social interaction, and repetitive patterns of behavior, interests, or activities. Based on the symptoms indicative of impaired social functioning, Baron-Cohen et al. (1985) hypothesized that the ToM hypothesis of autism is simple: Autism is the result of

an absent or impaired ToM. To test this hypothesis, researchers employed Wimmer and Perner's (1983) puppet play paradigm, comparing a group of autistic children to both typically developing children and those with Down's syndrome. Despite the autistic children having a higher mental age than the control groups, they were the only ones who struggled to attribute beliefs to others. Since then, evidence has confirmed that autistic individuals perform significantly more poorly in mind-reading tasks than neurotypical individuals, supporting the hypothesis of lacking ToM in autism (Chung et al., 2014). Moreover, functional impairments in brain regions, especially the TPJ and TPJ-related neural networks, have been found in autistic individuals (Igelström et al., 2017; Kana et al., 2014; Maximo et al., 2014).

While researchers have made efforts to improve the social cognitive abilities of individuals with ASD through ToM training, the results have not consistently shown strong efficacy (Fletcher-Watson et al., 2014). Cognitive training has limitations in improving social communication skills in individuals with ASD. Although some studies indicate that cognitive training can temporarily improve specific social skills in individuals with ASD, such as eye contact and nonverbal expression, these changes often lack durability and are difficult to generalize to everyday life. However, neurostimulation interventions show promise in enhancing the social cognitive abilities of individuals with ASD (Camacho-Conde et al., 2022; García-González et al., 2021). For example, some studies have explored non-invasive neurostimulation techniques such as transcranial direct current stimulation (tDCS) and transcranial magnetic stimulation (TMS) to improve social cognitive functioning in individuals with ASD. These methods modulate brain activity to promote neuroplasticity and have shown positive outcomes in some research studies.

Summary and looking forward

Given the potential power for the human social cognition and social artificial intelligence (AI), studies of representation of others' mind, which refers to the ability to understand one's own and others' mental states, has garnered significant attention in recent years. Inspired by the existing studies, partly summarized in this article, future research should explore more on factors mentioned, or predictions of under the proposed framework.

Researchers have highlighted the importance of incorporating diverse methodologies in exploring mentalization across different populations. Therefore, it is of interest to investigate mentalization in clinical samples and provide valuable insights into the impairments observed in various mental health conditions. Mental health disorders such as borderline personality disorder (BPD) have been associated with difficulties in accurately perceiving and understanding others' mental states.

The integration of behavioral tasks, neuroimaging techniques, and self-report measures offers a holistic approach to comprehending the cognitive processes at play in mentalization. For instance, functional magnetic resonance imaging (fMRI) can elucidate the neural networks involved in different false belief tasks, combining with computational models and questionnaires or big data from different cultures, can shed light on the brain regions associated with accurate mental state attributions. Such multidimensional approaches allow for a more nuanced understanding of the mechanisms underlying social cognition.

As AI technology advances, it is becoming increasingly important to understand how humans and AI can interact in social situations. Future work can also integrate AI and human perspectives to study the representation of other people's beliefs from two aspects: (1) using AI agents as social partners in belief-related tasks. For example, participants could complete a false belief task with an AI agent as the partner, where the agent's beliefs are either congruent or incongruent with the participant's beliefs; (2) using AI algorithms to analyze and predict human data collected during belief-related tasks. This can provide a more nuanced understanding of the behavioral mechanisms underlying belief understanding and AI to develop more human-like belief updating system.

Overall, investigating the implications of mentalization for different populations, data from different modalities, and integrating AI and human perspectives are promising lines of inquiry that has immense relevance for understanding social functioning. By employing diverse methodologies, researchers can gain a comprehensive understanding of the cognitive and neural mechanisms involved in mentalization, contributing to advancements in the field of social cognition and informing clinical practices.

Box 3 Questions for future research

How do the prior knowledge, social influence and interaction norm interact with each other to influence beliefs of others?

What are the potential benefits and drawbacks of using Bayesian modeling in the study of human representation of others?

How to identify specific brain regions and networks involved in processing others' beliefs, and how do these neural mechanisms differ across tasks and individuals?

Are there ways to enhance the accurate representation the beliefs and mental states of humans in social situations?

What considerations need to be considered when integrating AI and human perspectives to study the representation of other people's beliefs?

References

Abell, F., Happé, F., Frith, U., 2000. Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. Cogn. Dev. 15 (1), 1—16. https://doi.org/10.1016/S0885-2014(00)00014-9.

Adolphs, R., 2009. The social brain: neural basis of social knowledge. Annu. Rev. Psychol. 60, 693—716. https://doi.org/10.1146/annurev.psych.60.110707.163514.

Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., Ladurner, G., 2006. Do visual perspective tasks need theory of mind? Neuroimage 30 (3), 1059—1068. https://doi.org/10.1016/j.neuroimage.2005.10.026.

Andrea, L.C., Meghan, L.M., 2020. Self-other representation in the social brain reflects social connection. J. Neurosci. 40 (29), 5616. https://doi.org/10.1523/JNEUROSCI.2826-19.2020.

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., Rutter, M., 1995. Autism as a strongly genetic disorder: evidence from a British twin study. Psychol. Med. 25 (1), 63—77. https://doi.org/10.1017/S0033291700028099.

Baker, C., Saxe, R., Tenenbaum, J., 2011. Bayesian theory of mind: modeling joint belief-desire attribution. In: Proceedings of the Annual Meeting of the Cognitive Science Society.

Baker, C.L., Jara-Ettinger, J., Saxe, R., Tenenbaum, J.B., 2017. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. Nat. Human Behav. 1 (4), 0064. https://doi.org/10.1038/s41562-017-0064.

Bang, D., Fleming, S.M., 2018. Distinct encoding of decision confidence in human medial prefrontal cortex. Proc. Natl. Acad. Sci. U. S. A. 115 (23), 6082—6087. https://doi.org/10.1073/pnas.1800795115.

Bang, D., Frith, C.D., 2017. Making better decisions in groups. R. Soc. Open Sci. 4 (8), 170193. https://doi.org/10.1098/rsos.170193.

Baron-Cohen, S., Leslie, A.M., Frith, U., 1985. Does the autistic child have a "theory of mind"? Cognition 21 (1), 37—46. https://doi.org/10.1016/0010-0277(85)90022-8.

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., Plaisted, K., 1999. Recognition of faux pas by normally developing children and children with asperger syndrome or high-functioning autism. J. Autism Dev. Disord. 29 (5), 407—418. https://doi.org/10.1023/A:1023035012436.

Bartsch, K., Wellman, H.M., 1995. Children Talk about the Mind. Oxford University Press. https://doi.org/10.1093/oso/9780195080056.001.0001.

Bateman, A.W., Fonagy, P., 2019. Handbook of Mentalizing in Mental Health Practice. American Psychiatric Pub.

Bertenthal, B.I., Fischer, K.W., 1978. Development of self-recognition in the infant. Dev. Psychol. 14 (1), 44—50. https://doi.org/10.1037/0012-1649.14.1.44.

Bhatt, M.A., Lohrenz, T., Camerer, C.F., Montague, P.R., 2010. Neural signatures of strategic types in a two-person bargaining game. Proc. Natl. Acad. Sci. U. S. A. 107 (46), 19720—19725. https://doi.org/10.1073/pnas.1009625107.

Boekaerts, M., 1999. Self-regulated learning: where we are today. Int. J. Educ. Res. 31 (6), 445—457. https://doi.org/10.1016/S0883-0355(99)00014-2.

Buie, D.H., 1981. Empathy: its nature and limitations. J. Am. Psychoanal. Assoc. 29 (2), 281—307. https://doi.org/10.1177/000306518102900201.

Burke, C.J., Tobler, P.N., Baddeley, M., Schultz, W., 2010. Neural mechanisms of observational learning. Proc. Natl. Acad. Sci. U. S. A. 107 (32), 14431—14436. https://doi.org/10.1073/pnas.1003111107.

Burnett, S., Sebastian, C., Cohen Kadosh, K., Blakemore, S.-J., 2011. The social brain in adolescence: evidence from functional magnetic resonance imaging and behavioural studies. Neurosci. Biobehav. Rev. 35 (8), 1654—1664. https://doi.org/10.1016/j.neubiorev.2010.10.011.

Butler, R.J., Gasson, S.L., 2005. Self esteem/self concept scales for children and adolescents: a review. Child Adolesc. Ment. Health 10 (4), 190—201. https://doi.org/10.1111/j.1475-3588.2005.00368.x.

Camacho-Conde, J.A., Gonzalez-Bermudez, M.d.R., Carretero-Rey, M., Khan, Z.U., 2022. Brain stimulation: a therapeutic approach for the treatment of neurological disorders. CNS Neurosci. Ther. 28 (1), 5—18. https://doi.org/10.1111/cns.13769.

Camerer, C.F., 2011. Behavioral Game Theory: Experiments in Strategic Interaction. Princeton university press.

Cannon, E.N., Woodward, A.L., Gredebäck, G., von Hofsten, C., Turek, C., 2012. Action production influences 12-month-old infants' attention to others' actions. Dev. Sci. 15 (1), 35—42. https://doi.org/10.1111/j.1467-7687.2011.01095.x.

Carruthers, P., Chamberlain, A., 2000. Evolution and the Human Mind: Modularity, Language and Meta-Cognition. Cambridge University Press.

Caspi, A., Roberts, B.W., Shiner, R.L., 2004. Personality development: stability and change. Annu. Rev. Psychol. 56 (1), 453—484. https://doi.org/10.1146/annurev.psych.55.090902.141913.

Choudhry, F.R., Mani, V., Ming, L.C., Khan, T.M., 2016. Beliefs and perception about mental health issues: a meta-synthesis. Neuropsychiatric Dis. Treat. 12 (null), 2807—2818. https://doi.org/10.2147/NDT.S111543.

Chung, Y.S., Barch, D., Strube, M., 2014. A meta-analysis of mentalizing impairments in adults with achizophrenia and autism spectrum disorder. Schizophr. Bull. 40 (3), 602—616. https://doi.org/10.1093/schbul/sbt048.

Connors, M.H., Halligan, P.W., 2015. A cognitive account of belief: a tentative road map. Front. Psychol. 5. https://doi.org/10.3389/fpsyg.2014.01588.

Cristiano, A., Finisguerra, A., Urgesi, C., Avenanti, A., Tidoni, E., 2023. Functional role of the theory of mind network in integrating mentalistic prior information with action kinematics during action observation. Cortex 166, 107—120. https://doi.org/10.1016/j.cortex.2023.05.009.

Daunizeau, J., den Ouden, H.E.M., Pessiglione, M., Kiebel, S.J., Stephan, K.E., Friston, K.J., 2010. Observing the observer (I): meta-bayesian models of learning and decision-making. PLoS One 5 (12), e15554. https://doi.org/10.1371/journal.pone.0015554.

Davis, M.H., 1983. Measuring individual differences in empathy: evidence for a multidimensional approach. J. Pers. Soc. Psychol. 44 (1), 113—126. https://doi.org/10.1037/0022-3514.44.1.113.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., Dolan, R.J., 2011. Model-based influences on humans' choices and striatal prediction errors. Neuron 69 (6), 1204—1215. https://doi.org/10.1016/j.neuron.2011.02.027.

Devaine, M., Hollard, G., Daunizeau, J., 2014. The social bayesian brain: does mentalizing make a difference when we learn? PLoS Comput. Biol. 10 (12), e1003992. https://doi.org/10.1371/journal.pcbi.1003992.

Diamond, A., 2013. Executive functions. Annu. Rev. Psychol. 64 (1), 135—168. https://doi.org/10.1146/annurev-psych-113011-143750.

Duval, C., Piolino, P., Bejanin, A., Eustache, F., Desgranges, B., 2011. Age effects on different components of theory of mind. Conscious. Cogn. 20 (3), 627—642. https://doi.org/10.1016/j.concog.2010.10.025.

Elcheroth, G., Doise, W., Reicher, S., 2011. On the knowledge of politics and the politics of knowledge: how a social representations approach helps us rethink the subject of political psychology. Polit. Psychol. 32 (5), 729—758. https://doi.org/10.1111/j.1467-9221.2011.00834.x.

Elliott, R., Völlm, B., Drury, A., McKie, S., Richardson, P., William Deakin, J.F., 2006. Co-operation with another player in a financially rewarded guessing game activates regions implicated in theory of mind. Soc. Neurosci. 1 (3—4), 385—395. https://doi.org/10.1080/17470910601041358.

Ereira, S., Hauser, T.U., Moran, R., Story, G.W., Dolan, R.J., Kurth-Nelson, Z., 2020. Social training reconfigures prediction errors to shape Self-Other boundaries. Nat. Commun. 11 (1), 3030. https://doi.org/10.1038/s41467-020-16856-8.

Etel, E., Slaughter, V., 2019. Theory of mind and peer cooperation in two play contexts. J. Appl. Dev. Psychol. 60, 87—95. https://doi.org/10.1016/j.appdev.2018.11.004.

Fawcett, C., Liszkowski, U., 2012. Infants anticipate others' social preferences. Infant Child Dev. 21 (3), 239—249. https://doi.org/10.1002/icd.739.

Festinger, L., 1954. A theory of social comparison processes. Hum. Relat. 7 (2), 117—140. https://doi.org/10.1177/001872675400700202.

Fett, A.-K.J., Shergill, S.S., Gromann, P.M., Dumontheil, I., Blakemore, S.-J., Yakub, F., Krabbendam, L., 2014. Trust and social reciprocity in adolescence—a matter of perspective-taking. J. Adolesc. 37 (2), 175—184. https://doi.org/10.1016/j.adolescence.2013.11.011.

Filippetti, M.L., Johnson, M.H., Lloyd-Fox, S., Dragovic, D., Farroni, T., 2013. Body perception in newborns. Curr. Biol. 23 (23), 2413—2416. https://doi.org/10.1016/j.cub.2013.10.017.

Fleming, S.M., Dolan, R.J., 2012. The neural basis of metacognitive ability. Phil. Trans. Biol. Sci. 367 (1594), 1338—1349. https://doi.org/10.1098/rstb.2011.0417.
Fletcher-Watson, S., McConnell, F., Manola, E., McConachie, H., 2014. Interventions based on the Theory of Mind cognitive model for autism spectrum disorder (ASD). Cochrane Database Syst. Rev. (3). https://doi.org/10.1002/14651858.CD008785.pub2.
Frith, C.D., Frith, U., 1999. Interacting minds–a biological basis. Science 286 (5445), 1692—1695. https://doi.org/10.1126/science.286.5445.1692.
Frith, C.D., Frith, U., 2006. The neural basis of mentalizing. Neuron 50 (4), 531—534. https://doi.org/10.1016/j.neuron.2006.05.001.
Frith, C.D., Frith, U., 2008. Implicit and explicit processes in social cognition. Neuron 60 (3), 503—510. https://doi.org/10.1016/j.neuron.2008.10.032.
Frith, U., Frith, C.D., 2003. Development and neurophysiology of mentalizing. Phil. Trans. Roy. Soc. Lond. B Biol. Sci. 358 (1431), 459—473. https://doi.org/10.1098/rstb.2002.1218.
García-González, S., Lugo-Marín, J., Setien-Ramos, I., Gisbert-Gustemps, L., Arteaga-Henríquez, G., Díez-Villoria, E., Ramos-Quiroga, J.A., 2021. Transcranial direct current stimulation in Autism Spectrum Disorder: a systematic review and meta-analysis. Eur. Neuropsychopharmacol 48, 89—109. https://doi.org/10.1016/j.euroneuro.2021.02.017.
Gergely, G., 2002. The development of understanding self and agency. In: Goswami, U. (Ed.), Blackwell Handbook of Childhood Cognitive Development. Blackwell Publishers Ltd, pp. 26—46. https://doi.org/10.1002/9780470996652.ch2.
Gilovich, T.S., Medvec, K., Husted, V., 1998. The illusion of transparency: biased assessments of others' ability to read one's emotional states. J. Pers. Soc. Psychol. 75, 332—346. https://doi.org/10.1037/0022-3514.75.2.332.
Glidden, J., D'Esterre, A., Killen, M., 2021. Morally-relevant theory of mind mediates the relationship between group membership and moral judgments. Cogn. Dev. 57, 100976. https://doi.org/10.1016/j.cogdev.2020.100976.
Gopnik, A., Slaughter, V., 1991. Young children's understanding of changes in their mental states. Child Dev. 62 (1), 98—110. https://doi.org/10.1111/j.1467-8624.1991.tb01517.x.
Grèzes, J., Frith, C.D., Passingham, R.E., 2004. Inferring false beliefs from the actions of oneself and others: an fMRI study. Neuroimage 21 (2), 744—750. https://doi.org/10.1016/S1053-8119(03)00665-7.
Griffiths, T.L., Chater, N., Kemp, C., Perfors, A., Tenenbaum, J.B., 2010. Probabilistic models of cognition: exploring representations and inductive biases. Trends Cogn. Sci. 14 (8), 357—364. https://doi.org/10.1016/j.tics.2010.05.004.
Gutsell, J.N., Inzlicht, M., 2010. Empathy constrained: prejudice predicts reduced mental simulation of actions during observation of outgroups. J. Exp. Soc. Psychol. 46 (5), 841—845. https://doi.org/10.1016/j.jesp.2010.03.011.
Hansen, I.G., Ryder, A., 2016. In search of "religion proper": intrinsic religiosity and coalitional rigidity make opposing predictions of intergroup hostility across religious groups. J. Cross Cult. Psychol. 47 (6), 835—857. https://doi.org/10.1177/0022022116644983.
Happé, F., Brownell, H., Winner, E., 1999. Acquired 'theory of mind' impairments following stroke. Cognition 70 (3), 211—240. https://doi.org/10.1016/S0010-0277(99)00005-0.
Happé, F.G.E., 1994. An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. J. Autism Dev. Disord. 24 (2), 129—154. https://doi.org/10.1007/BF02172093.
Hart, W., Albarracín, D., Eagly, A.H., Brechan, I., Lindberg, M.J., Merrill, L., 2009. Feeling validated versus being correct: a meta-analysis of selective exposure to information. Psychol. Bull. 135 (4), 555—588. https://doi.org/10.1037/a0015701.
Hartwright, C.E., Apperly, I.A., Hansen, P.C., 2012. Multiple roles for executive control in belief—desire reasoning: distinct neural networks are recruited for self perspective inhibition and complexity of reasoning. Neuroimage 61 (4), 921—930. https://doi.org/10.1016/j.neuroimage.2012.03.012.
Hegarty, M., Waller, D., 2004. A dissociation between mental rotation and perspective-taking spatial abilities. Intelligence 32 (2), 175—191. https://doi.org/10.1016/j.intell.2003.12.001.
Heider, F., Simmel, M., 1944. An experimental study of apparent behavior. Am. J. Psychol. 57 (2), 243—259. https://doi.org/10.2307/1416950.
Henry, J.D., Phillips, L.H., Ruffman, T., Bailey, P.E., 2013. A meta-analytic review of age differences in theory of mind. Psychol. Aging 28 (3), 826—839. https://doi.org/10.1037/a0030677.
Ho, M.K., Saxe, R., Cushman, F., 2022. Planning with theory of mind. Trends Cogn. Sci. 26 (11), 959—971. https://doi.org/10.1016/j.tics.2022.08.003.
Howarth, C., 2001. Towards a social psychology of community: a social representations perspective. J. Theor. Soc. Behav. 31 (2), 223—238. https://doi.org/10.1111/1468-5914.00155.
Hu, C., Di, X., Eickhoff, S.B., Zhang, M., Peng, K., Guo, H., Sui, J., 2016. Distinct and common aspects of physical and psychological self-representation in the brain: a meta-analysis of self-bias in facial and self-referential judgements. Neurosci. Biobehav. Rev. 61, 197—207. https://doi.org/10.1016/j.neubiorev.2015.12.003.
Igelström, K.M., Webb, T.W., Graziano, M.S.A., 2017. Functional connectivity between the temporoparietal cortex and cerebellum in autism spectrum disorder. Cerebr. Cortex 27 (4), 2617—2627. https://doi.org/10.1093/cercor/bhw079.
Itzchakov, G., Reis, H.T., Weinstein, N., 2022. How to foster perceived partner responsiveness: high-quality listening is key. Soc. Pers. Psychol. Compass 16 (1), e12648. https://doi.org/10.1111/spc3.12648.
Jiang, S., Wang, S., Wan, X., 2022. Metacognition and mentalizing are associated with distinct neural representations of decision uncertainty. PLoS Biol. 20 (5), e3001301. https://doi.org/10.1371/journal.pbio.3001301.
Kahneman, D., 2003a. Maps of bounded rationality: psychology for behavioral economics. Am. Econ. Rev. 93 (5), 1149—1475. https://doi.org/10.1257/000282803322655392.
Kahneman, D., 2003b. A psychological perspective on economics. Am. Econ. Rev. 93 (2), 162—168. https://doi.org/10.1257/000282803321946985.
Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47 (2), 263—291.
Kana, R.K., Libero, L.E., Hu, C.P., Deshpande, H.D., Colburn, J.S., 2014. Functional brain networks and white matter underlying theory-of-mind in autism. Soc. Cogn. Affect Neurosci. 9 (1), 98—105. https://doi.org/10.1093/scan/nss106.
Kanner, L., 1943. Autistic disturbances of affective contact. Nervous child 2 (3), 217—250.
Karniol, R., 1990. Reading people's minds: a transformation rule model for predicting other's thoughts and feelings. Adv. Exp. Soc. Psychol. 23, 211—247. Elsevier. doi.org/10.1016/S0065-2601(08)60320-2.
Kienhues, D., Bromme, R., 2011. Beliefs about abilities and epistemic beliefs: aspects of cognitive flexibility in information-rich environments. In: Elen, J., Stahl, E., Bromme, R., Clarebout, G. (Eds.), Links between Beliefs and Cognitive Flexibility: Lessons Learned. Springer, Netherlands, pp. 105—124. https://doi.org/10.1007/978-94-007-1793-0_6.
Kilford, E.J., Garrett, E., Blakemore, S.-J., 2016. The development of social cognition in adolescence: an integrated perspective. Neurosci. Biobehav. Rev. 70, 106—120. https://doi.org/10.1016/j.neubiorev.2016.08.016.
Kim, D.H., 2009. The link between individual and organizational learning. In: The Strategic Management of Intellectual Capital. Routledge, pp. 41—62.
Kovács, Á.M., Téglás, E., Endress, A.D., 2010. The social sense: susceptibility to others' beliefs in human infants and adults. Science 330 (6012), 1830—1834. https://doi.org/10.1126/science.1190792.
Kravitz, H., Goldenberg, D., Neyhus, A., 1978. Tactual exploration by normal infants. Dev. Med. Child Neurol. 20, 720—726. https://doi.org/10.1111/j.1469-8749.1978.tb15302.x.
Kwon, S.-J., Telzer, E.H., 2022. Social contextual risk taking in adolescence. Nat. Rev. Psychol. 1 (7), 393—406. https://doi.org/10.1038/s44159-022-00060-0.
Lampi, K.M., Lehtonen, L., Tran, P.L., Suominen, A., Lehti, V., Banerjee, P.N., Sourander, A., 2012. Risk of autism spectrum disorders in low birth weight and small for gestational age infants. J. Pediatr. 161 (5), 830—836. https://doi.org/10.1016/j.jpeds.2012.04.058.
Leary, T., 2004. Interpersonal Diagnosis of Personality: A Functional Theory and Methodology for Personality Evaluation. Wipf and Stock Publishers.
Lewis, C., Osborne, A., 1990. Three-year-olds' problems with false belief: conceptual deficit or linguistic artifact? Child Dev. 61 (5), 1514—1519. https://doi.org/10.1111/j.1467-8624.1990.tb02879.x.
Liu, D., Wellman, H.M., Tardif, T., Sabbagh, M.A., 2008. Theory of mind development in Chinese children: a meta-analysis of false-belief understanding across cultures and languages. Dev. Psychol. 44 (2), 523—531. https://doi.org/10.1037/0012-1649.44.2.523.

Livingston, J.A., 2003. Metacognition: An Overview.
Ma, Y., Han, S., 2010. Why we respond faster to the self than to others? An implicit positive association theory of self-advantage during implicit face recognition. J. Exp. Psychol. Hum. Percept. Perform. 36 (4), 619—633. https://doi.org/10.1037/a0015797.
Martin, G.B., Clark, R.D., 1982. Distress crying in neonates: species and peer specificity. Dev. Psychol. 18 (1), 3—9. https://doi.org/10.1037/0012-1649.18.1.3.
Maximo, J.O., Cadena, E.J., Kana, R.K., 2014. The implications of brain connectivity in the neuropsychology of autism. Neuropsychol. Rev. 24 (1), 16—31. https://doi.org/10.1007/s11065-014-9250-0.
McCabe, K.A., Rigdon, M.L., Smith, V.L., 2003. Positive reciprocity and intentions in trust games. J. Econ. Behav. Organ. 52 (2), 267—275. https://doi.org/10.1016/S0167-2681(03)00003-9.
McCubbins, M.D., Turner, M.B., Weller, N., 2012. The theory of minds within the theory of games. In: Proceedings of the 2012 International Conference on Artificial Intelligence.
McDonald, S., Flanagan, S., 2004. Social perception deficits after traumatic brain injury: interaction between emotion recognition, mentalizing ability, and social communication. Neuropsychology 18, 572—579. https://doi.org/10.1037/0894-4105.18.3.572.
McLoughlin, N., Over, H., 2017. Young children are more likely to spontaneously attribute mental states to members of their own group. Psychol. Sci. 28 (10), 1503—1509. https://doi.org/10.1177/0956797617710724.
Miller, C.A., 2006. Developmental relationships between language and theory of mind. Am. J. Speech Lang. Pathol 15 (2), 142—154. https://doi.org/10.1044/1058-0360(2006/014.
Mitchell, J.P., Banaji, M.R., Macrae, C.N., 2005. General and specific contributions of the medial prefrontal cortex to knowledge about mental states. Neuroimage 28 (4), 757—762. https://doi.org/10.1016/j.neuroimage.2005.03.011.
Modgil, S., Singh, R.K., Gupta, S., Dennehy, D., 2021. A confirmation bias view on social media induced polarisation during Covid-19. Inf. Syst. Front. https://doi.org/10.1007/s10796-021-10222-9.
Molenberghs, P., Johnson, H., Henry, J.D., Mattingley, J.B., 2016. Understanding the minds of others: a neuroimaging meta-analysis. Neurosci. Biobehav. Rev. 65, 276—291. https://doi.org/10.1016/j.neubiorev.2016.03.020.
Moore, D.A., Healy, P.J., 2008. The trouble with overconfidence. Psychol. Rev. 115 (2), 502—517. https://doi.org/10.1037/0033-295X.115.2.502.
Moscovici, S., 2001. Social Representations: Essays in Social Psychology. NYU Press. https://books.google.com/books?id=0fA8DAAAQBAJ.
Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R.J., Friston, K.J., 2014. Bayesian inferences about the self (and others): a review. Conscious. Cogn. 25, 67—76. https://doi.org/10.1016/j.concog.2014.01.009.
Muhle, R., Trentacoste, S.V., Rapin, I., 2004. The genetics of autism. Pediatrics 113 (5), e472—e486. https://doi.org/10.1542/peds.113.5.e472.
Murphy, M., Bolton, P.F., Pickles, A., Fombonne, E., Piven, J., Rutter, M., 2000. Personality traits of the relatives of autistic probands. Psychol. Med. 30 (6), 1411—1424. https://doi.org/10.1017/S0033291799002949.
Murray, R.J., Schaer, M., Debbané, M., 2012. Degrees of separation: a quantitative neuroimaging meta-analysis investigating self-specificity and shared neural activation between self- and other-reflection. Neurosci. Biobehav. Rev. 36 (3), 1043—1059. https://doi.org/10.1016/j.neubiorev.2011.12.013.
Neisser, U., 1991. Two perceptually given aspects of the self and their development. Dev. Rev. 11 (3), 197—209. https://doi.org/10.1016/0273-2297(91)90009-D.
Nelson, T.E., Clawson, R.A., Oxley, Z.M., 1997. Media framing of a civil liberties conflict and its effect on tolerance. Am. Polit. Sci. Rev. 91 (3), 567—583. https://doi.org/10.2307/2952075.
Nguyen, T.N., Gonzalez, C., 2022. Theory of mind from observation in cognitive models and humans. Top. Cogn. Sci. 14 (4), 665—686. https://doi.org/10.1111/tops.12553.
Oishi, S., 2014. Socioecological psychology. Annu. Rev. Psychol. 65 (1), 581—609. https://doi.org/10.1146/annurev-psych-030413-152156.
Oishi, S., Choi, H., 2017. Chapter four—culture and motivation: a socio-ecological approach. In: Elliot, A.J. (Ed.), Advances in Motivation Science, vol. 4. Elsevier, pp. 141—170. https://doi.org/10.1016/bs.adms.2017.02.004.
Oishi, S., Graham, J., 2010. Social ecology: lost and found in psychological science. Perspect. Psychol. Sci. 5 (4), 356—377. https://doi.org/10.1177/1745691610374588.
Onishi, K.H., Baillargeon, R., Leslie, A.M., 2007. 15-month-old infants detect violations in pretend scenarios. Acta Psychol. 124 (1), 106—128. https://doi.org/10.1016/j.actpsy.2006.09.009.
Pasquali, A., Timmermans, B., Cleeremans, A., 2010. Know thyself: metacognitive networks and measures of consciousness. Cognition 117 (2), 182—190. https://doi.org/10.1016/j.cognition.2010.08.010.
Pavey, L., Greitemeyer, T., Sparks, P., 2012. "I help because I want to, not because you tell me to": empathy increases autonomously motivated helping. Pers. Soc. Psychol. Bull. 38 (5), 681—689. https://doi.org/10.1177/0146167211435940.
Payne, S., Tsakiris, M., 2017. Anodal transcranial direct current stimulation of right temporoparietal area inhibits self-recognition. Cogn. Affect Behav. Neurosci. 17, 1—8. https://doi.org/10.3758/s13415-016-0461-0.
Pelphrey, K.A., Morris, J.P., McCarthy, G., 2004. Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. J. Cogn. Neurosci. 16 (10), 1706—1716. https://doi.org/10.1162/0898929042947900.
Peterson, C.C., Wellman, H.M., 2019. Longitudinal theory of mind (ToM) development from preschool to adolescence with and without ToM delay. Child Dev. 90 (6), 1917—1934. https://doi.org/10.1111/cdev.13064.
Piaget, J., Inhelder, B., 1967. The Child's Conception of Space. W.W. Norton & Company, Inc.
Platek, S.M., Wathne, K., Tierney, N.G., Thomson, J.W., 2008. Neural correlates of self-face recognition: an effect-location meta-analysis. Brain Res. 1232, 173—184. https://doi.org/10.1016/j.brainres.2008.07.010.
Poznyak, E., Morosan, L., Perroud, N., Speranza, M., Badoud, D., Debbané, M., 2019. Roles of age, gender and psychological difficulties in adolescent mentalizing. J. Adolesc. 74, 120—129. https://doi.org/10.1016/j.adolescence.2019.06.007.
Premack, D., Woodruff, G., 1978. Does the chimpanzee have a theory of mind? Behav. Brain Sci. 1 (4), 515—526. https://doi.org/10.1017/S0140525X00076512.
Quesque, F., Brass, M., 2019. The role of the temporoparietal junction in self-other distinction. Brain Topogr. 32 (6), 943—955. https://doi.org/10.1007/s10548-019-00737-5.
Quesque, F., Rossetti, Y., 2020. What do theory-of-mind tasks actually measure? Theory and practice. Perspect. Psychol. Sci. 15 (2), 384—396. https://doi.org/10.1177/1745691619896607.
Ramsey, R., Kaplan, D.M., Cross, E.S., 2021. Watch and learn: the cognitive neuroscience of learning from others' actions. Trends Neurosci. 44 (6), 478—491. https://doi.org/10.1016/j.tins.2021.01.007.
Redcay, E., Schilbach, L., 2019. Using second-person neuroscience to elucidate the mechanisms of social interaction. Nat. Rev. Neurosci. 20 (8), 495—505. https://doi.org/10.1038/s41583-019-0179-4.
Reynolds, W.J., Scott, B., 1999. Empathy: a crucial component of the helping relationship. J. Psychiatr. Ment. Health Nurs. 6 (5), 363—370. https://doi.org/10.1046/j.1365-2850.1999.00228.x.
Rokeach, M., 1970. Beliefs, Attitudes and Values: A Theory of Organization and Change. Jossey-Bass.
Rubio-Fernández, P., Geurts, B., 2012. How to pass the false-belief task before your fourth birthday. Psychol. Sci. 24 (1), 27—33. https://doi.org/10.1177/0956797612447819.
Salvatore, G., Lysaker, P.H., Popolo, R., Procacci, M., Carcione, A., Dimaggio, G., 2012. Vulnerable self, poor understanding of others' minds, threat anticipation and cognitive biases as triggers for delusional experience in schizophrenia: a theoretical model. Clin. Psychol. Psychother. 19 (3), 247—259. https://doi.org/10.1002/cpp.746.
Santiesteban, I., Banissy, M.J., Catmur, C., Bird, G., 2012. Enhancing social ability by stimulating right temporoparietal junction. Curr. Biol. 22 (23), 2274—2277. https://doi.org/10.1016/j.cub.2012.10.018.
Sarfati, Y., Hardy-Baylé, M.-C., Besche, C., Widlöcher, D., 1997. Attribution of intentions to others in people with schizophrenia: a non-verbal exploration with comic strips. Schizophr. Res. 25 (3), 199—209. https://doi.org/10.1016/S0920-9964(97)00025-X.

Schulreich, S., Tusche, A., Kanske, P., Schwabe, L., 2023. Higher subjective socioeconomic status is linked to increased charitable giving and mentalizing-related neural value coding. Neuroimage 279, 120315. https://doi.org/10.1016/j.neuroimage.2023.120315.
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. https://doi.org/10.1016/j.neubiorev.2014.01.009.
Schurz, M., Radua, J., Tholen, M.G., Maliske, L., Margulies, D.S., Mars, R.B., Kanske, P., 2021. Toward a hierarchical model of social cognition: a neuroimaging meta-analysis and integrative review of empathy and theory of mind. Psychol. Bull. 147 (3), 293–327. https://doi.org/10.1037/bul0000303.
Scott, R.M., Baillargeon, R., 2017. Early false-belief understanding. Trends Cogn. Sci. 21 (4), 237–249. https://doi.org/10.1016/j.tics.2017.01.012.
Selcuk, B., Gonultas, S., Ekerim-Akbulut, M., 2023. Development and use of theory of mind in social and cultural context. Child Dev. Perspect. 17 (1), 39–45. https://doi.org/10.1111/cdep.12473.
Shahaeian, A., Peterson, C.C., Slaughter, V., Wellman, H.M., 2011. Culture and the sequence of steps in theory of mind development. Dev. Psychol. 47 (5), 1239–1247. https://doi.org/10.1037/a0023899.
Sidera, F., Perpiñà, G., Serrano, J., Rostan, C., 2018. Why is theory of mind important for referential communication? Curr. Psychol. 37 (1), 82–97. https://doi.org/10.1007/s12144-016-9492-5.
Simon, H.A., 1955. A behavioral model of rational choice. Q. J. Econ. 99–118. https://doi.org/10.2307/1884852.
Sodian, B., 2011. Theory of mind in infancy. Child Dev. Perspect. 5 (1), 39–43. https://doi.org/10.1111/j.1750-8606.2010.00152.x.
Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., Hajak, G., 2007. Neural correlates of true and false belief reasoning. Neuroimage 35 (3), 1378–1384. https://doi.org/10.1016/j.neuroimage.2007.01.042.
Spengler, S., von Cramon, D.Y., Brass, M., 2009. Control of shared representations relies on key processes involved in mental state attribution. Hum. Brain Mapp. 30 (11), 3704–3718. https://doi.org/10.1002/hbm.20800.
Surian, L., Leslie, A.M., 1999. Competence and performance in false belief understanding: a comparison of autistic and normal 3-year-old children. Br. J. Dev. Psychol. 17 (1), 141–155. https://doi.org/10.1348/026151099165203.
Surtees, A., Apperly, I., Samson, D., 2016. I've got your number: spontaneous perspective-taking in an interactive task. Cognition 150, 43–52. https://doi.org/10.1016/j.cognition.2016.01.014.
Sutton, R.S., Barto, A.G., 2018. Reinforcement Learning: An Introduction. MIT press.
Suzuki, S., 2022. Inferences regarding oneself and others in the human brain. PLoS Biol. 20 (5), e3001662. https://doi.org/10.1371/journal.pbio.3001662.
Taber, C.S., Lodge, M., 2006. Motivated skepticism in the evaluation of political beliefs. Am. J. Polit. Sci. 50 (3), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x.
Uddin, L.Q., 2021. Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. Nat. Rev. Neurosci. 22 (3), 167–179. https://doi.org/10.1038/s41583-021-00428-w.
Vélez, N., Chen, A.M., Burke, T., Cushman, F.A., Gershman, S.J., 2023. Teachers recruit mentalizing regions to represent learners' beliefs. Proc. Natl. Acad. Sci. U. S. A. 120 (22), e2215015120. https://doi.org/10.1073/pnas.2215015120.
Valentine, J.C., DuBois, D.L., Cooper, H., 2004. The relation between self-beliefs and academic achievement: a meta-analytic review. Educ. Psychol. 39 (2), 111–133. https://doi.org/10.1207/s15326985ep3902_3.
Wang, H., Callaghan, E., Gooding-Williams, G., McAllister, C., Kessler, K., 2016. Rhythm makes the world go round: an MEG-TMS study on the role of right TPJ theta oscillations in embodied perspective taking. Cortex 75, 68–81. https://doi.org/10.1016/j.cortex.2015.11.011.
Wellman, H.M., Cross, D., Watson, J., 2001. Meta-analysis of theory-of-mind development: the truth about false belief. Child Dev. 72 (3), 655–684. https://doi.org/10.1111/1467-8624.00304.
Wellman, H.M., Fang, F., Liu, D., Zhu, L., Liu, G., 2006. Scaling of theory-of-mind understandings in Chinese children. Psychol. Sci. 17 (12), 1075–1081. https://doi.org/10.1111/j.1467-9280.2006.01830.x.
Wimmer, H., Perner, J., 1983. Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition 13 (1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5.
Woodward, A.L., 2009. Infants' grasp of others' intentions. Curr. Dir. Psychol. Sci. 18 (1), 53–57. https://doi.org/10.1111/j.1467-8721.2009.01605.x.
Wu, H., Liu, X., Hagan, C.C., Mobbs, D., 2020. Mentalizing during social InterAction: a four component model. Cortex 126, 242–252. https://doi.org/10.1016/j.cortex.2019.12.031.
Wu, H., Luo, Y., Feng, C., 2016. Neural signatures of social conformity: a coordinate-based activation likelihood estimation meta-analysis of functional brain imaging studies. Neurosci. Biobehav. Rev. 71, 101–111. https://doi.org/10.1016/j.neubiorev.2016.08.038.
Wu, S., Keysar, B., 2007. The effect of culture on perspective taking. Psychol. Sci. 18 (7), 600–606. https://doi.org/10.1111/j.1467-9280.2007.01946.x.
Wu, S., Wu, F., Ding, Y., Hou, J., Bi, J., Zhang, Z., 2017. Advanced parental age and autism risk in children: a systematic review and meta-analysis. Acta Psychiatr. Scand. 135 (1), 29–41. https://doi.org/10.1111/acps.12666.
Yeager, D.S., Dweck, C.S., 2020. What can be learned from growth mindset controversies? Am. Psychol. 75 (9), 1269–1284. https://doi.org/10.1037/amp0000794.
Yeshurun, Y., Nguyen, M., Hasson, U., 2021. The default mode network: where the idiosyncratic self meets the shared social world. Nat. Rev. Neurosci. 22 (3), 181–192. https://doi.org/10.1038/s41583-020-00420-w.
Yoshida, W., Dolan, R.J., Friston, K.J., 2008. Game theory of mind. PLoS Comput. Biol. 4 (12), e1000254. https://doi.org/10.1371/journal.pcbi.1000254.
Yott, J., Poulin-Dubois, D., 2016. Are infants' theory-of-mind abilities well integrated? implicit understanding of intentions, desires, and beliefs. J. Cogn. Dev. 17 (5), 683–698. https://doi.org/10.1080/15248372.2015.1086771.
Zühlsdorff, K., Dalley, J.W., Robbins, T.W., Morein-Zamir, S., 2023. Cognitive flexibility: neurobehavioral correlates of changing one's mind. Cerebr. Cortex 33 (9), 5436–5446. https://doi.org/10.1093/cercor/bhac431.
Zaki, J., Weber, J., Bolger, N., Ochsner, K., 2009. The neural bases of empathic accuracy. Proc. Natl. Acad. Sci. U. S. A. 106 (27), 11382–11387. https://doi.org/10.1073/pnas.0902666106.
Zynda, L., 2000. Representation theorems and realism about degrees of belief. Philos. Sci. 67 (1), 45–69. https://doi.org/10.1086/392761.

Further reading

Gray, H.M., Gray, K., Wegner, D.M., 2007. Dimensions of mind perception. Science 315 (5812), 61.
Heyes, C.M., Frith, C.D., 2014. The cultural evolution of mind reading. Science 344 (6190), 1243091.
Ramnani, N., Miall, R.C., 2004. A system in the human brain for predicting the actions of others. Nat. Neurosci. 7 (1), 85–90.
Tamir, D.I., Thornton, M.A., 2018. Modeling the predictive social mind. Trends Cogn. Sci. 22 (3), 201–212.
Wu, H., Liu, X., Hagan, C.C., Mobbs, D., 2020. Mentalizing during social InterAction: a four component model. Cortex 126, 242–252.