





## Opinion

Socially evaluative contexts  
facilitate mentalizingBrandon M. Woo <sup>1,2,\*</sup>, Enda Tan <sup>3,4</sup>, Francis L. Yuen <sup>3</sup> and J. Kiley Hamlin <sup>3</sup>

**Our ability to understand others' minds stands at the foundation of human learning, communication, cooperation, and social life more broadly. Although humans' ability to mentalize has been well-studied throughout the cognitive sciences, little attention has been paid to whether and how mentalizing differs across contexts. Classic developmental studies have examined mentalizing within minimally social contexts, in which a single agent seeks a neutral inanimate object. Such object-directed acts may be common, but they are typically consequential only to the object-seeking agent themselves. Here, we review a host of indirect evidence suggesting that contexts providing the opportunity to evaluate prospective social partners may facilitate mentalizing across development. Our article calls on cognitive scientists to study mentalizing in contexts where it counts.**

## Introduction

Our ability to mentalize – to make sense of others' mental states – is central to human social life and forms an active area of research throughout the cognitive sciences [1–3]. Impactful studies have provided evidence that mentalizing abilities emerge early in human ontogeny: Infants and toddlers readily represent others' **goals** (see *Glossary*) [4,5], **knowledge** [6,7], and (controversially) **beliefs** [8–10]. Great attention has been paid to when in development children mentalize, to whom they attribute **mental states** (e.g., **agents** vs. non-agents [4,11]), what kinds of mental states they reason about (e.g., knowledge vs. beliefs [12,13]), and the nature of these representations (whether they reflect mentalizing or some lower-level approximation [14]). By contrast, **far less attention has been paid to whether there are particular contexts where mentalizing is likely to occur in human development.** This omission is surprising: whereas some mental states have direct relevance for survival and flourishing (e.g., whether an individual intends to hurt vs. help), others have less relevance (e.g., whether an individual intends to approach one neutral, unremarkable object vs. another). Further, mentalizing often carries cognitive cost [15], requiring inferences that go beyond the surface of others' actions. Studies have revealed that nonhuman primates are more likely to mentalize when contexts are ecologically relevant (e.g., involving food) [16]. Here, **we propose that humans may be particularly likely to mentalize in contexts where representing others' mental states impacts survival and flourishing.**

## Mentalizing allows for accurate social evaluation

To date, most of the developmental mentalizing literature has focused on children's abilities to reason about an individual agent who pursues a neutral object [4,8,9,17]. In a seminal study examining toddlers' understanding of false beliefs [8], an agent repeatedly sought a neutral object. The object's location changed either in the agent's presence or absence, leading the agent to hold either a true or false belief about the object's location. Toddlers' looking times suggested that they expected the agent to search for the object where she believed it to be. Such belief representations could support early learning about the environment [18]. Other studies, however, have failed to replicate this evidence for false-belief understanding in toddlers

## Highlights

Cognitive scientists have long studied the origins of our ability to mentalize. Remarkably little is known, however, about **whether there are particular contexts where humans are more likely to mentalize.**

**We propose that mentalizing is facilitated in contexts where others' actions shed light on their status as a good or bad social partner.** Mentalizing within socially evaluative contexts supports effective partner choice.

Our proposal **is based on three lines of evidence.** First, infants leverage their understanding of others' mental states to evaluate others' social actions. Second, infants, children, and adults demonstrate enhanced mentalizing within socially evaluative contexts. Third, infants, children, and adults are especially likely to mentalize when agents cause negative outcomes.

Direct tests of this proposal will contribute to a more comprehensive understanding of human mentalizing.

<sup>1</sup>Department of Psychology, Harvard University, Cambridge, MA, USA, 02138

<sup>2</sup>Center for Brains, Minds, and Machines, Cambridge, MA, USA, 02139

<sup>3</sup>Department of Psychology, University of British Columbia, Vancouver, BC, Canada, V6T 1Z4

<sup>4</sup>Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD, USA, 20742

\*Correspondence: [bmwoo@g.harvard.edu](mailto:bmwoo@g.harvard.edu) (B.M. Woo).

©Twitter: @brandonmwoo (B.M. Woo).

[13,19,20] (cf. [21]). Here, we note that although observers could track the agent's beliefs in such scenarios, it is unclear why they would consistently bother to, as knowing a stranger's beliefs about a neutral object may bring limited practical value to observers.

In contrast to the neutral, object-directed acts in such scenarios, human life is centered around interdependent, cooperative relationships in which people's actions impact others. From birth, humans show motivations to engage with others [22] and depend on nonkin to provide them with assistance [23], facilitate their goals [24], and teach them new skills and knowledge [25]. Because of this interdependence, humans face the task of selecting appropriate cooperative partners: those who will care for versus abandon them, help versus harm them, teach versus mislead them, etc. This task, known as partner choice [6], requires abilities to accurately assess others' cooperative potential, to maximize positive and minimize negative social interactions.

Although one could assess strangers' cooperative potential through trial-and-erroring first-person interactions, this could be time-consuming and costly. Thus, the task of partner choice is facilitated by capacities for making inferences based on information gained as an independent third party (e.g., by observing how individuals behave towards others). In particular, those who have behaved prosocially should be viewed as better potential partners than those who have behaved antisocially. A host of research suggests that rudimentary abilities for making such social inferences emerge early in development: Infants and toddlers prefer looking to and reaching for prosocial over antisocial agents, such as those who have helped versus hindered others' goals [26–30].

Importantly, however, purely outcome-based social inferences would be limited: people often cause or are associated with outcomes that they did not intend (e.g., trying but failing to help, accidentally causing harm, etc.). Thus, the inferences made from observations of third-party social behaviors should focus on the mental states that drive social action [31–33]. Although it may require more effort to reason about mental states versus outcomes [15,31], individuals able to accurately attribute cooperative and uncooperative mental states will likely be more successful at choosing appropriate cooperative partners.

Here, we propose that mentalizing is facilitated in contexts where agents' actions are relevant to their cooperative potential. We refer to these as socially evaluative contexts, which include, but are not limited to, contexts involving prosocial and antisocial acts such as helping versus hindering others' goals, providing physical protection or care versus harming, and behaving fairly versus unfairly. In contrast, we refer to contexts where agentive actions are irrelevant to cooperative potential (e.g., actions on neutral objects) as nonevaluative. Given that failures to identify bad cooperative partners may be particularly costly, within socially evaluative contexts, individuals may focus especially on antisocial mental states.

Although, to our knowledge, no research has directly tested the possibility that socially evaluative contexts facilitate mentalizing, here we identify three lines of evidence that already indirectly support it. First, at the earliest ages at which human infants have been shown to represent others' mental states, these representations already appear to inform their evaluation of prospective social partners. Second, infants, children, and adults may be more sensitive to mental states within socially evaluative contexts than within nonevaluative ones. Third, infants, children, and adults may overattribute mental states when agents are antisocial versus prosocial. These three lines of evidence suggest that mentalizing is facilitated in contexts relevant to selecting social partners.

## Glossary

**Agents:** entities capable of forming goals and acting to achieve them. Often contrasted with inanimate objects that do not engage in self-propelled motion nor possess/demonstrate additional cues to agency (e.g., eyes and contingent behavior).

**Belief:** an agent's attitude about some state of the world, which may or may not be true. True beliefs are consistent with reality; false beliefs are inconsistent with reality. In infant and toddler studies, false beliefs are often established by making an agent ignorant of some change in the state of the world (e.g., when a toy changes locations).

**Goals:** a desired target (e.g., an object, location, outcome, etc.) that an agent works toward. In infant studies, goals are often established by showing an agent repeatedly acting on or toward a particular target. Goals may be unfulfilled if agents are incapable of achieving them (e.g., due to distance, barriers, etc.).

**Knowledge:** an agent's awareness of a particular state of the world (e.g., of an event, a fact, another's mental state, etc.). In infant studies, knowledge is often established based on whether agents are present to observe an event. If an agent lacks knowledge, they are instead described as ignorant.

**Mental state:** a representation within an agent's mind. Mental states include (but are not limited to) an agent's goals, states of knowledge and ignorance, and beliefs. Mental states are not directly observable. Studies of infants and toddlers challenge participants to infer agents' mental states from their actions and circumstances.

**Socially evaluative contexts:** situations that allow for assessing agents' value as cooperative partners. Assessments are typically based on agents' actions toward other agents. Nonevaluative contexts, in contrast, are situations in which agents' actions are irrelevant to assessing their cooperative potential.

**Teleological reasoning:** a nonmentalistic, reality-based framework for representing agentive action in which agents' actions reflect the efficient pursuit of a goal-state in light of relevant situational constraints. Teleological reasoning, as compared to mentalizing, allows observers to reason effectively about others' intentional action in many situations, but breaks down in cases in which the relationship between action,

### Early mental state representations inform social evaluation

The claim that socially evaluative contexts facilitate mentalizing may appear counterintuitive, given the computational demands within these contexts. Indeed, socially evaluative contexts typically involve multiple agents, each with unique mental states, some of which may have others' mental states embedded within them (e.g., intending to facilitate others' goals) [34–36]. Nevertheless, growing evidence suggests that at the earliest ages infants reason about mental states in general, they do so within socially evaluative contexts (see [Box 1](#) for a review of infants' representations of states of knowledge). Here, we turn to goal understanding as a case study.

situation, and end-state is imperfect, including failed attempts, accidents, and actions carried out based on ignorance or false belief.

In a classic study, Woodward [4] found that after an agent repeatedly acted on one object over another (a nonevaluative context), 6- and 9-month-olds expected the agent to continue reaching for the same object after the objects' locations switched. These findings have been consistently replicated [37–39], with evidence from infants as young as 3 months of age [7,40–42]. These findings provide evidence that infants represent object-directed goals (though there is debate about whether these representations are mentalistic [14]).

Notably, 3-month-olds already appear to leverage their goal understanding to evaluate potential cooperative partners. In one study [27], 3-month-olds watched a protagonist agent who tried but failed to climb a hill. A helper pushed the protagonist up the hill, whereas a hinderer pushed the protagonist down. After infants were habituated to these events, the helper and the hinderer were presented side-by-side, and infants' attention to each was measured. Infants preferentially looked to the helper, suggesting that infants differently evaluated the agents. In another study [28], 3-month-olds watched events in which a protagonist played with, then lost a ball. A helper returned the ball, whereas a hinderer ran away with it. Here too, infants preferentially looked to the helper. Thus, infants appear to prefer prosocial over antisocial agents.

To see agents as helping or hindering a protagonist in the pursuit of a goal in these studies, one presumably must first identify the protagonist's goal. Growing evidence has linked infants' evaluations of helpers and hinderers to goal understanding [43–45]. For instance, within the aforementioned hill paradigm [43], 6- to 11-month-olds only preferred the agent who pushed the protagonist uphill when the protagonist's goal was clear (i.e., not when it looked downhill while moving up) ([Figure 1A](#)). Similarly, in another study [44], the more 5-month-olds looked to the top of the hill as the protagonist tried to climb, the more that the infants preferred the helper, presumably because those infants more strongly represented the protagonist's goal ([Figure 1B](#)). In summary, despite the complexity of representing the goals of multiple agents, even young infants appear to represent goals within socially evaluative contexts, and use those representations to inform their social choices.

### Mentalizing is increased within socially evaluative contexts

Although socially evaluative contexts introduce computational demands, growing evidence suggests that mental state understanding not only informs social choices but is enhanced within socially evaluative versus nonevaluative contexts. Here, we review behavioral evidence and neuroimaging evidence that support this possibility.

#### Developmental behavioral evidence

Psychologists have probed infants' and toddlers' representations of various mental states beyond goals, including intentions and beliefs. Critically, representing intentions and beliefs requires reasoning beyond the physically instantiated outcomes of others' actions, and is therefore more complex than representing others' fulfilled goals. Here, we explore evidence that infants and toddlers represent these mental states more strongly in socially evaluative versus nonevaluative contexts.

**Box 1. Representations of knowledge and ignorance in infancy**

People often have information that others lack. How do infants navigate such situations? In one study [6], 6-month-olds observed an agent who reached for one object over another when they could see both objects, but the agent could only see the reached-for object. When the objects later changed locations, infants did not expect the agent to reach to the same object (for converging evidence in 3-month-olds and toddlers, see [7,98]). These findings suggest that infants track what others know and are ignorant about, based on what others can see.

Within socially evaluative contexts, infants have similarly demonstrated sensitivity to others' states of knowledge and ignorance. States of knowledge and ignorance appear to moderate infants' expectations for how agents will respond to others who have engaged in prosocial and antisocial actions. Studies have found that infants and toddlers prefer an agent who distributes resources equally (i.e., fairly) over an agent who distributes resources unequally (unfairly) to others [54,99–101]. When a third agent observes such acts of distribution, 10-month-olds are surprised when that agent chooses to reward an unfair distributor over a fair distributor. However, infants do not hold the same expectations when an agent had not observed others' acts of distribution [102]. Likewise, when an agent observes a peer harm others, 13-month-olds are surprised when the agent continues to interact with the peer; when the agent did not observe the act of harm, infants instead are surprised when the agent avoids the peer [103].

Beyond moderating expectations of others' behavior, states of knowledge and ignorance moderate infants' own evaluations of helping. In one experiment [36], 10-month-olds observed a protagonist who sought one toy over another (Figure 1A), either in the presence or absence of two other agents. When the protagonist's access to the toys was later blocked, one agent provided access to the desired toy; the other instead provided access to the undesired toy (Figure 1B). Infants preferred the agent who provided access to the desired toy only when the agents had observed the protagonist demonstrate its preference (Figure 1C–F).

Thus, at all ages that have been tested within socially evaluative contexts, infants have demonstrated sensitivity to others' states of knowledge and ignorance. Most research within socially evaluative contexts, however, has been on older infants. Although there is evidence that 3- and 6-month-olds track others' knowledge and ignorance in nonevaluative contexts [6,7], it remains unclear whether these representations inform their social evaluations.

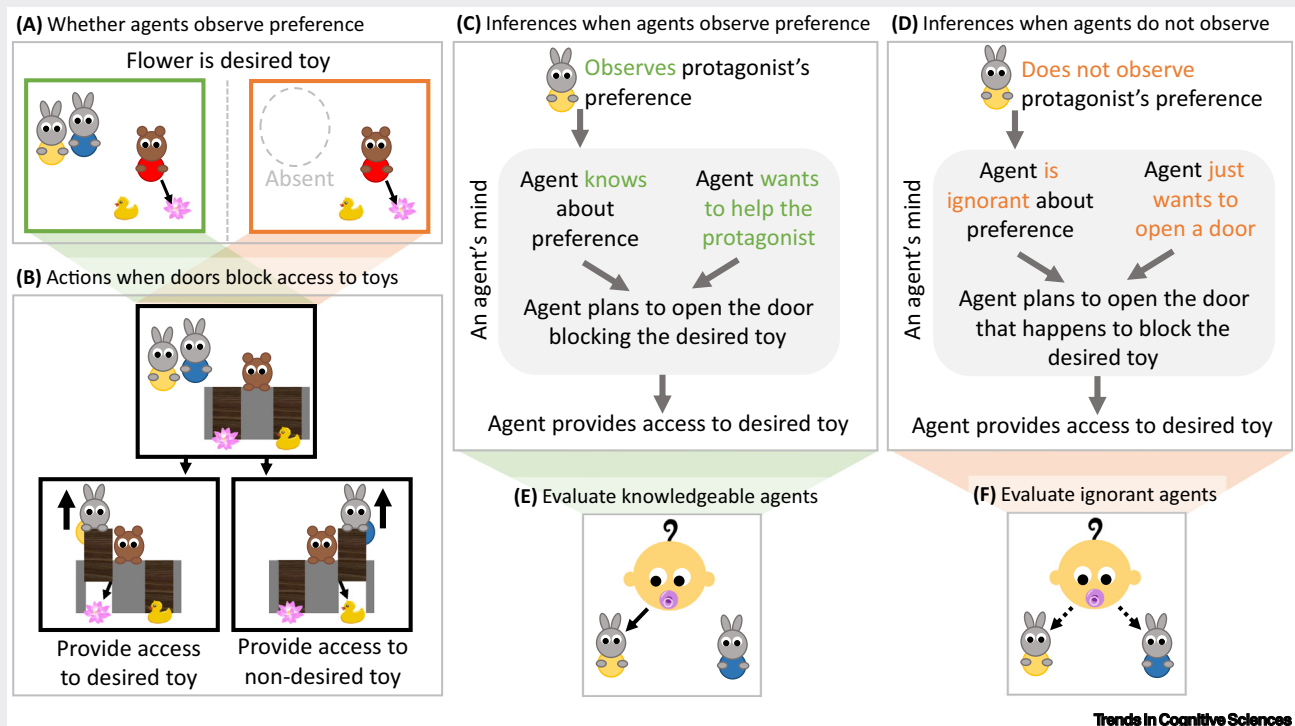
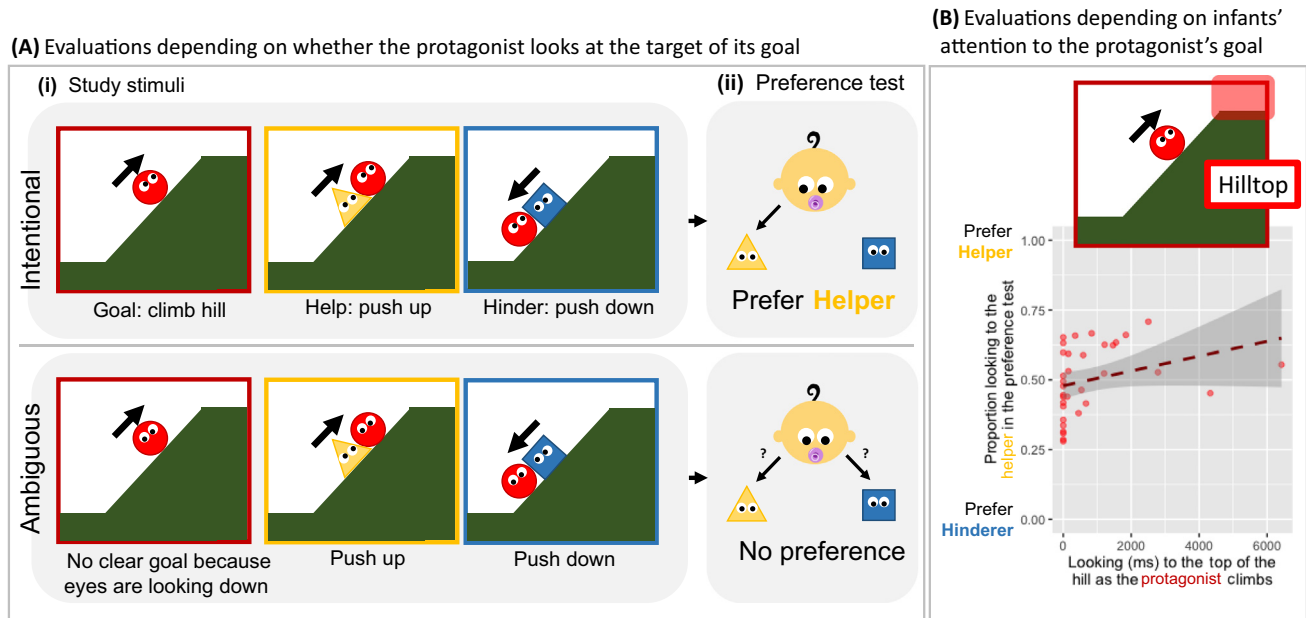


Figure 1. Role of knowledge and ignorance in early social evaluations. This is a schematic of the events that 10-month-olds observed in [36], and the inferences (C, D) and evaluations (E, F) that infants formed.

*Intentions: unfulfilled goals*

Adults appreciate that agents' goals often go unfulfilled (e.g., someone wants an out-of-reach object; Figure 2, Key figure). There is mixed evidence about whether young infants represent unfulfilled goals in nonevaluative contexts (e.g., when a person fails to reach a distant object):



Trends in Cognitive Sciences

**Figure 1. Evidence that goal representations inform infants' social evaluations.** (A) Infants aged 6–11 months preferred an agent who helped a protagonist climb a hill when the protagonist looked at the top of the hill [43], providing evidence that its actions were intentional and that it had the goal of climbing the hill. (B) The more 5-month-old infants looked to the hilltop, the target of the protagonist's goal, the more strongly they preferred the helper who facilitated that goal over the hinderer who prevented that goal [44]. The graph in (B) was created using data in [44].

Whereas two papers report evidence that 7- to 8-month-olds can do this [46,47], two other papers report that they do not [48,49] (Figure 3). We are aware of no studies demonstrating sensitivity to unfulfilled goals in nonevaluative contexts prior to 7 months.

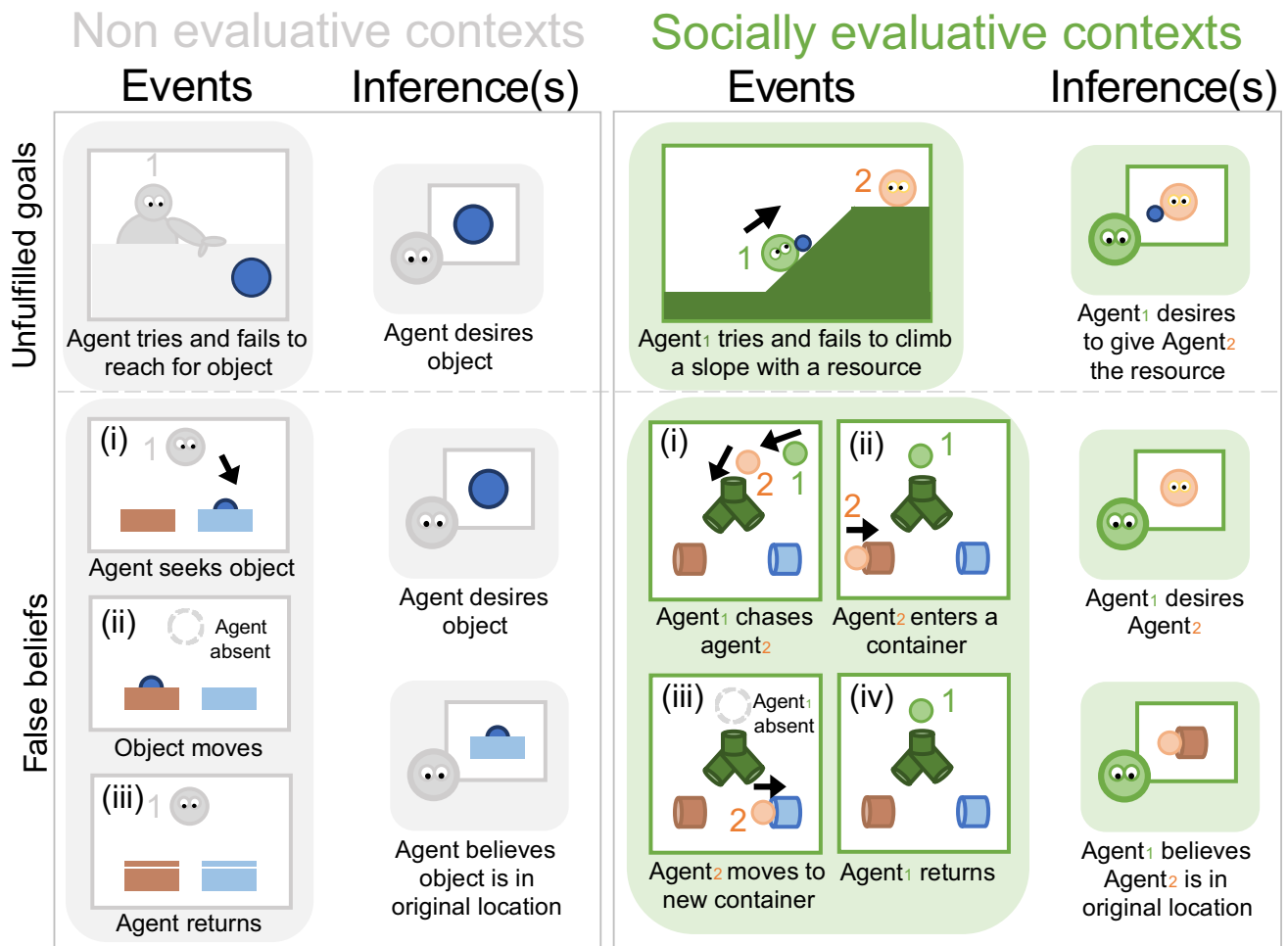
In contrast, evidence suggests that infants and toddlers represent unfulfilled goals months earlier within socially evaluative contexts (Figure 3). As reviewed above, studies have found that infants prefer helpers over hinderers by 3 months of age [27,28]. Importantly, these studies involve protagonists demonstrating unfulfilled goals (e.g., failing to climb a hill or retrieve a ball) who are then helped or hindered. Seeing helpers' and hinderers' actions as helping and hindering is presumably facilitated by representing the protagonist's unfulfilled goal.

Within these paradigms, however, protagonists achieve their goals whenever helpers intervene; thus, infants could appreciate the protagonist's unfulfilled goal after they see it being achieved, by reading backward from the outcome (i.e., **teleological reasoning** [14]). Evidence against this possibility comes from studies in which infants chose between hinderers and neutral agents [26,27], rather than hinderers and helpers. Here, the protagonist was only ever hindered, never achieving its goal: there was no opportunity to infer the goal *post hoc*. Yet, studies have shown that 10-, 6-, and 3-month-olds all prefer neutral agents over hinderers. **These findings suggest that infants represent unfulfilled goals in evaluative contexts by 3 months of age, at least 4 months before they do so in nonevaluative contexts.**

In addition to understanding unfulfilled nonsocial goals (e.g., to climb a hill), young infants appear sensitive to others' unfulfilled social goals, goals that indicate prosocial versus antisocial intentions. In first-party interactions, infants from 6 months express more impatience when adults are unwilling

**Key figure**

Comparisons of nonevaluative and socially evaluative contexts



Trends in Cognitive Sciences

**Figure 2.** This schematic depicts examples of events that infants and toddlers observe, and the mental states that they are challenged to infer, within studies involving nonevaluative and socially evaluative contexts. We highlight [48] and [8] as examples of designs within nonevaluative contexts studying early reasoning about unfulfilled goals and beliefs, respectively. Similarly, we highlight [55,56] and [60–63,97] as examples of designs within socially evaluative contexts studying early reasoning about unfulfilled goals and beliefs, respectively. Light grey and green indicate nonevaluative and socially evaluative contexts, respectively.

versus unable to share a toy [50–52]. In these situations, the outcomes of the adults' actions are equated (nobody shares), yet infants respond as though they are sensitive to adults' intentions. These results support the possibility that infants represent intentions within socially evaluative contexts.

Further, by late in the first year, infants demonstrate sensitivity to agents' unfulfilled goals to be prosocial versus antisocial towards third parties. While infants typically prefer helpers over hinderers, in one study [53], when two agents were unsuccessful in their attempts to help versus



**Figure 3. Timeline of children's understanding of unfulfilled goals.** This timeline depicts the ages, in months, for which there are findings that speak to whether infants and toddlers represent unfulfilled goals. Light grey and green indicate findings within nonevaluative contexts [46–49] and socially evaluative contexts [27,28,50–53,55,56], respectively. Full circles indicate evidence for an ability, and empty circles indicate a lack of evidence for (i.e., null findings) an ability.

hinder a protagonist, 8-month-olds preferred the agent who attempted to help over the agent who attempted to hinder, although the attempted helper was associated with a worse outcome: the protagonist failing to achieve its goal. Five-month-olds chose randomly in the same conditions, as though they noticed the inconsistency between intention and outcome, but were unsure which to focus on. Further evidence for infants' sensitivity to intention in socially evaluative contexts comes from the fairness domain [54]. Studies have reported that when agents were unsuccessful in their attempts to be fair versus unfair, 10-month-olds expected an observer to approach an agent who attempted to be fair [55] (Figure 2), and 9-month-olds preferred an agent who attempted to be fair over an agent who attempted to be unfair [56].

In summary, infants and toddlers appear sensitive to others' unfulfilled goals when interpreting others' social decisions and when evaluating other agents. Even when outcomes are equated (i.e., teleological reasoning is not possible), infants demonstrate sensitivity to others' unfulfilled goals to be prosocial versus antisocial (see [Box 2](#) for evidence that infants privilege intentions over outcomes in contexts involving accidents). A sensitivity to unfulfilled goals emerges both earlier in development and more robustly in socially evaluative contexts than in nonevaluative contexts.

### *False beliefs*

Situations involving false beliefs challenge observers to appreciate differences between their own and others' representations of the world [57]. Thus, false-belief understanding serves as a benchmark for mentalizing. Decades of research have demonstrated that it is not until 4 years of age that children reason about others' false beliefs in verbal tasks [58]. In nonverbal tests, by contrast, some studies have found that toddlers represent false beliefs.

In the first such study, Onishi and Baillargeon [8] presented 15-month-olds with an agent who sought a toy that moved between two boxes ([Figure 2](#)). For some toddlers, the toy moved while the agent was absent, resulting in the agent holding a false belief about the toy's location. Patterns of looking suggested that toddlers expected the agent to reach for the box where she last saw the toy, even if the toy had moved. Studies have since provided further evidence for nonverbal false-belief reasoning in toddlers [9,10].

There have been several failed attempts to replicate findings of nonverbal false-belief understanding in toddlers [12,13,19,20,59], including direct replications by the same laboratories reporting positive findings. This has inspired considerable debate about whether toddlers represent false beliefs. Although failed replications exist across other areas of cognitive development, including within social evaluation, the sheer number of failed attempts suggests that early nonverbal false-belief understanding may be especially fragile. That said, to our knowledge, these failed replications have only

#### **Box 2. Processing others' unintentional actions in infancy**

A child runs toward a desired toy, inadvertently tripping and knocking over a vase. Is the child blameworthy? Although accidents and failed attempts both present a conflict between intentions and outcomes, one key difference is whether agents cause the outcomes that they are associated with. In failed attempts, agents are simply associated with an outcome. For example, in the case of an agent who tries but fails to open a jam jar, the jar would have remained unopened even without the agent's intervention. In accidents, by contrast, agents cause outcomes. In the above scenario, the vase broke because the boy knocked it over. Thus, outcomes may be more salient for accidents versus failed attempts, and people may therefore find it more difficult to process mental states when evaluating accidents. Indeed, activity in brain regions associated with cognitive conflict is higher when adults evaluate accidental versus attempted harm [31], and in verbal tasks, children exculpate accidental harm later in development than they condemn attempted harm [104]. When agents accidentally cause positive or negative outcomes, how would infants evaluate agents: based on their outcomes or intentions?

Growing evidence suggests that by 10 months of age, infants' social evaluations are sensitive to accidental agents' lack of intentions [87,105]. In one experiment [87], 10-month-olds viewed a protagonist who struggled to place a toy on top of a high shelf, only succeeding after several attempts. During alternating events, an intentional harmer who had witnessed these struggles purposefully approached the shelf and pushed it over, and an accidental harmer who had not witnessed these struggles clumsily and inadvertently knocked the shelf over while running by. Although both agents caused a negative outcome for the protagonist, infants nevertheless distinguished between them, preferring the accidental over the intentional harmer. In another condition in which infants observed intentional and accidental helpers, infants preferred the intentional helper. Converging evidence has come from research on infants' evaluations of agents who prevent harm [105]: 10-month-olds (but not 6-month-olds) prefer agents who prevent harm intentionally over agents who do so accidentally. **These findings reveal that by late in the first year, infants are sensitive to whether social actions are intentional versus accidental. This is striking, given that accidents are more difficult to process than are failed attempts; these findings provide further support for the possibility that there is increased mentalizing within socially evaluative contexts.**



examined belief within nonevaluative contexts, in which an agent has beliefs about the location of an inanimate object. We propose that evidence for false-belief understanding may be mixed because nonevaluative contexts give observers little reason to care about agents' beliefs.

Socially evaluative contexts, by contrast, may facilitate belief representations by giving observers reason to care about agents' mental states. A growing number of studies have found evidence for nonverbal belief representations in toddlers, children, and adults when an agent chases another agent [60–63] – a social (typically harmful and/or antisocial) goal that may be relevant to social evaluation [64] – rather than seeking an inanimate object (for difficulty replicating these findings in 5-year-old children and adults, see [65,66]). In one paper [62], toddlers and adults viewed videos of a triangle that chased a disk (Figure 2). The triangle was either absent or present as the disk moved from one box to another. When the triangle returned, 17- to 23-month-olds and adults, but not 13- to 17-month-olds, looked first to the location where the triangle last 'saw' the disk. Similarly, evidence for belief representations in great apes has come from contexts where one agent chases another [67]. These findings contrast with failures to replicate findings of belief representations in 2-year-olds in nonevaluative contexts [59] when an agent instead seeks an object.

Research is now probing infants' and toddlers' social evaluations of agents who act on false beliefs. In one experiment [68], 15-month-olds viewed a protagonist who sought a desired toy that was located in one box, as two other puppets were present to observe. Then, either as the observers were present or absent, the desired toy was moved to a new location. When the observers were present and therefore knowledgeable about the toy's movement, toddlers valued an observer who directed the protagonist to the desired toy's new location. When the observers were absent and therefore had false beliefs about the toy's location, toddlers instead valued an observer who directed the protagonist to the location where the observers had last seen the desired toy: where they believed it to be. Further experiments with toddlers and 8-month-olds have conceptually replicated these findings [69]. Taken together, this evidence supports the possibility that mental states are represented more strongly and robustly within socially evaluative contexts.

#### Adult and infant neuroimaging evidence

Neuroscientists have identified key regions of the human brain that support mentalizing; these include the temporoparietal junction, the medial prefrontal cortex, the temporal poles, the precuneus, and the posterior superior temporal sulcus (pSTS) [70–72]. If socially evaluative contexts facilitate mentalizing relative to nonevaluative ones, then activity within this mentalizing network may be higher when humans observe actions that are more socially relevant. Indeed, nearly two decades of research have found that activity within these brain regions differs when adults view more versus less social actions [73–78]. In one fMRI study [78], adults viewed videos of agents who helped versus hindered a protagonist climb a hill (adapting stimuli from [26]) or engaged in independent actions (climbing independently). Here, videos of helping and hindering – actions that infants and adults evaluate – led to increased activity in the pSTS, relative to independent actions. Likewise, in another fMRI study [77], adults saw pictures of social (e.g., people cooperating or arguing) and independent (e.g., people facing away) actions. Here, pictures of social actions led to increased pSTS activity, relative to pictures of individuals acting independently. Recent research using functional near-infrared spectroscopy has provided converging evidence in 6- to 13-month-old infants [79]. The dorsomedial prefrontal cortex was higher in activity as infants observed social actions (e.g., clapping together) versus independent actions (e.g., clapping away from each other). These studies are striking: despite the social and independent actions being highly perceptually similar, activation in the mentalizing network differed for actions that were more versus less social. These findings support the possibility that adults and infants mentalize more strongly when viewing agents engaged in positive and negative interactions, because these provide opportunities for social evaluation.

Thus far, we have focused on neuroimaging studies that have examined how activity in the mentalizing network may differ for (i) individuals engaged in prosocial and antisocial interactions versus (ii) individuals engaged in independent actions. Both these kinds of stimuli are social in that they involve multiple agents, but only the former involves agents who act on other agents. In the next section, in addition to reviewing behavioral research, we review studies that contrast neural responses to positive versus negative social interactions.

### Negative social evaluation leads to an overattribution of mental states

Within socially evaluative contexts, failure to identify antisocial, noncooperative agents can present immediate risks to survival [80,81]. Given these risks, people may be especially likely to mentalize when agents harm (or attempt to harm) others, or when agents choose to violate established social norms. Consistent with this possibility, classic research has demonstrated that children reason about false beliefs more strongly when agents deceive versus play with others [58]. Here, we review evidence concerning a wider range of mental states, suggestive that adults, children, and infants are especially likely to mentalize when provided with opportunities for negative social evaluation.

#### Attributions of intentions

Perhaps the most well-known version of this phenomenon, first studied by Joshua Knobe, is the side-effect effect [82]. Here, adults are asked to imagine a company's chairman who starts a program knowing that it is likely to earn profits and either help or harm the environment. The chairman claims not to care about the environment and implements the program, which has its predicted effects. When asked whether the chairman intended to help/harm the environment, adults report that the chairman intended to harm the environment, but not help it. These and related findings [83] suggest that adults see negative side effects as more intentional than positive side effects. Some research suggests that the side-effect effect reflects a sensitivity to agents who have violated norms, be they moral or nonmoral [84]. **Because norm violations are highly relevant to cooperation, adults may overattribute intentions to agents who violate norms.**

Developmental research has revealed similar phenomena in children [85,86] and infants. As described in [Box 2](#), in one experiment [87], 10-month-olds preferred intentional to accidental helpers, but accidental to intentional hinderers. In an additional experiment, another group of 10-month-olds viewed events in which two agents both accidentally (clumsily) harmed or helped a protagonist. Critically, only one agent knew of the protagonist's goal when the agent accidentally facilitated or prevented it; only this agent could have foreseen the side effect of its action. When both agents accidentally harmed, infants preferred the ignorant over the knowledgeable agent. By contrast, infants chose randomly between the agents who accidentally helped. Thus, like adults, infants appear to find side effects caused by knowledgeable agents as more intentional when those outcomes are negative, rather than positive.

#### Attributions of agency

Opportunities for negative social evaluation may also lead to increased attributions of agency. For instance, adults typically view robots as being less agentic than human adults [88]. However, when a robot cheats to win a game (that is, acts antisocially), adults rate the robot as being more agentic and talk more to the robot, relative to when the robot behaved prosocially (helped the adult win) or neutrally [89,90]. Likewise, adults who lose economic games are more likely to guess that they were playing with another human (vs. a computer) than are adults who win [91], and adults attribute agency to malfunctioning but not to functioning computers [92]. Thus, negative outcomes can lead to attributions of agency, even in situations where no agents are present.

A tendency to attribute agency to nonagentic sources of negative outcomes may also occur in infancy. As described above, by 6 months of age, infants attribute goals to a hand that selectively acts on one object over another, showing surprise when the hand later acts on a different object [4]. Notably, infants in these studies do not attribute object-directed goals to inanimate tools, suggesting that their goal attribution is specific to agents. When an inanimate tool has first hindered an animate protagonist in its goal to open a box, however, 6-month-olds then attribute object-directed goals to the inanimate tool; by contrast, when an inanimate tool has first helped an animate protagonist, infants do not attribute goals to the tool [93]. These findings suggest that infants attributed agency to non-agentic tools that cause negative, but not positive, outcomes. Taken together, then, both adults and infants appear to attribute greater agency to the causes of negative outcomes.

#### Neural evidence of increased sensitivity to mental states

Complementary findings have come from research that has probed adults' and infants' neural activity. As reviewed above, fMRI research has presented adults with pictures of positive and negative social interactions (e.g., acts of cooperation vs. conflict) and noninteracting individuals [77]. In addition to examining how the pSTS (part of the mentalizing network) responds to more versus less social stimuli, this research has examined whether it responds differently to positive versus negative social stimuli. Here, activity in the pSTS was higher for negative than for positive social interactions (see [78] for qualitatively similar trends).

Research with infants has found convergent results. Tan and Hamlin [94] examined responses to helpers and hinderers in 6- and 12-month-olds using electroencephalography. They measured the P400 component, which has been associated with processing goal-directed actions [95,96] and is thought to reflect activity in the superior temporal sulcus. Here, event-related potentials were higher when 6- and 12-month-olds viewed images of a hinderer versus a helper. Thus, for both infants and adults, negative social interactions lead to stronger neural activity associated with mental state attribution than positive social interactions.

#### Concluding remarks

Mental state reasoning is not only used for social evaluation, but may be facilitated, and even overactivated, when humans engage in social evaluation. Human infants begin mentalizing in socially evaluative contexts as soon as they do so in nonevaluative contexts, if not earlier, and mental state representations across human development may be stronger in socially evaluative contexts, particularly when there are negative outcomes. This opinion article supports the possibility that mentalizing is privileged within socially evaluative contexts, perhaps due to its key role in facilitating the selection of appropriate cooperative partners. Effective partner choice may provide a strong foundation upon which humans' intensely interdependent and cooperative nature can flourish.

The work cited herein is highly suggestive, and more work is clearly needed to further explore this possibility (see [Outstanding questions](#)). We have mostly reviewed and compared data across experiments that have studied mentalizing in either socially evaluative or nonevaluative contexts, pulling from a wide range of ages and methods; to our knowledge, no research has directly compared both socially evaluative and nonevaluative contexts within the same experiment. Experiments using stringent minimal contrast designs would provide stronger tests of our central claims. In addition to such experiments, in the same way that meta-analyses have explored other predictors of mentalizing [13,58], we call on future researchers to conduct meta-analyses of findings that come from socially evaluative and nonevaluative contexts. We look forward to such research, which together will move us towards a more comprehensive understanding of humans' early mentalizing.

#### Outstanding questions

Is mentalizing facilitated when engaging in social evaluation more broadly? In our review of the developmental literature, we have focused on evaluations of agents who engage in morally relevant behaviors (e.g., helping, harming, and being fair), but infants and children also evaluate agents based on other social behaviors (e.g., imitating, yielding, and following norms) that may have consequences for cooperation.

How and why might socially evaluative contexts facilitate mentalizing? One possibility is that humans can mentalize in any context involving agents but are especially motivated within socially evaluative contexts. Alternatively, there may be capacities supporting mentalizing in socially evaluative contexts that may not readily apply to other contexts.

Might mentalizing be facilitated when there are immediate consequences for communication and learning? Beyond socially evaluative contexts, mentalizing is also important for effective communication and learning (e.g., understanding what others know).

Do younger infants appreciate others' knowledge and beliefs within socially evaluative contexts? Studies that have found evidence for sensitivity to such mental states within socially evaluative contexts have typically focused on older infants or toddlers.

Do infants consistently show enhanced mentalizing in socially evaluative contexts when their expectations, rather than their preferences, are assessed? Whereas infant studies involving socially evaluative contexts often assess preferences, studies involving nonevaluative contexts instead usually assess expectations.

Do socially evaluative contexts facilitate spontaneous belief representations? In some studies, others' false beliefs (e.g., about the location of an object) influence participants' behaviors (i.e., they respond as though they share those false beliefs), even when participants have true beliefs (i.e., they know where the object is).

## Acknowledgments

We thank the Cambridge Writing Group for feedback on an early draft. BW and FY were funded by Social Sciences and Humanities Research Council of Canada (SSHRC) Doctoral Fellowships (award 752-2020-0474 for BW, award 752-2022-2189 for FY). ET was funded by a SSHRC postdoctoral fellowship (award 756-2022-0589). JKH was funded by grants from SSHRC and the Natural Sciences and Engineering Research Councils of Canada (award RGPIN-2016-03775).

## Declaration of interests

No interests are declared.

## References

- Rabinowitz, N. *et al.* (2018) Machine theory of mind. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 4218–4227
- Horschler, D.J. *et al.* (2019) Do non-human primates really represent others' ignorance? A test of the awareness relations hypothesis. *Cognition* 190, 72–80
- Hyde, D.C. *et al.* (2018) Functional organization of the temporal-parietal junction for theory of mind in preverbal infants: a near-infrared spectroscopy study. *J. Neurosci.* 38, 4264–4274
- Woodward, A.L. (1998) Infants selectively encode the goal object of an actor's reach. *Cognition* 69, 1–34
- Cannon, E.N. and Woodward, A.L. (2012) Infants generate goal-based action predictions. *Dev. Sci.* 15, 292–298
- Luo, Y. and Johnson, S.C. (2009) Recognizing the role of perception in action at 6 months. *Dev. Sci.* 12, 142–149
- Choi, Y. *et al.* (2018) How do 3-month-old infants attribute preferences to a human agent? *J. Exp. Child Psychol.* 172, 96–106
- Onishi, K.H. and Baillargeon, R. (2005) Do 15-month-old infants understand false beliefs? *Science* 308, 255–258
- Southgate, V. *et al.* (2007) Action anticipation through attribution of false belief by 2-year-olds. *Psychol. Sci.* 18, 587–592
- Buttelmann, D. *et al.* (2009) Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition* 112, 337–342
- Weisman, K. *et al.* (2021) Similarities and differences in concepts of mental life among adults and children in five cultures. *Nat. Hum. Behav.* 5, 1358–1368
- Phillips, J. *et al.* (2021) Knowledge before belief. *Behav. Brain Sci.* 44, 1–37
- Holland, C. and Phillips, J.S. (2020) A theoretically driven meta-analysis of implicit theory of mind studies: the role of factivity. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*, pp. 1749–1755
- Gergely, G. and Csibra, G. (2003) Teleological reasoning in infancy: the naive theory of rational action. *Trends Cogn. Sci.* 7, 287–292
- Ferguson, H.J. *et al.* (2017) Eye tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. *Q. J. Exp. Psychol.* 70, 1646–1660
- Krupenye, C. and Call, J. (2019) Theory of mind in animals: current and future directions. *Wiley Interdiscip. Rev. Cogn. Sci.* 10, e1503
- Ganglmayer, K. *et al.* (2019) Infants' perception of goal-directed actions: a multi-lab replication reveals that infants anticipate paths and not goals. *Infant Behav. Dev.* 57, 101340
- Southgate, V. (2020) Are infants altercentric? The other and the self in early social cognition. *Psychol. Rev.* 127, 505
- Poulin-Dubois, D. *et al.* (2018) Do infants understand false beliefs? We don't know yet – a commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cogn. Dev.* 48, 302–315
- Powell, L.J. *et al.* (2018) Replications of implicit theory of mind tasks with varying representational demands. *Cogn. Dev.* 46, 40–50
- Scott, R.M. and Baillargeon, R. (2017) Early false-belief understanding. *Trends Cogn. Sci.* 21, 237–249
- Raz, G. and Saxe, R. (2020) Learning in infancy is active, endogenously motivated, and depends on the prefrontal cortices. *Annu. Rev. Dev. Psychol.* 2, 247–268
- Hrdy, S.B. (2009) *Mothers and Others: the Evolutionary Origins of Mutual Understanding*, Harvard University Press
- Tomasello, M. and Carpenter, M. (2007) Shared intentionality. *Dev. Sci.* 10, 121–125
- Gweon, H. (2021) Inferential social learning: cognitive foundations of human social learning and teaching. *Trends Cogn. Sci.* 25, 896–910
- Hamlin, J.K. *et al.* (2007) Social evaluation by preverbal infants. *Nature* 450, 557–559
- Hamlin, J.K. *et al.* (2010) Three-month-olds show a negativity bias in their social evaluations. *Dev. Sci.* 13, 923–929
- Hamlin, J.K. and Wynn, K. (2011) Young infants prefer prosocial to antisocial others. *Cogn. Dev.* 26, 30–39
- Margoni, F. and Surian, L. (2018) Infants' evaluation of prosocial and antisocial agents: a meta-analysis. *Dev. Psychol.* 54, 1445–1455
- Woo, B.M. *et al.* (2022) Human morality is based on an early-emerging moral core. *Annu. Rev. Dev. Psychol.* <https://doi.org/10.1146/annurev-devpsych-121020-023312>
- Young, L. *et al.* (2007) The neural basis of the interaction between theory of mind and moral judgment. *Proc. Natl. Acad. Sci.* 104, 8235–8240
- Curtin, C.M. *et al.* (2020) Kinship intensity and the use of mental states in moral judgment across societies. *Evol. Hum. Behav.* 41, 415–429
- Barrett, H.C. *et al.* (2016) Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proc. Natl. Acad. Sci.* 113, 4688–4693
- Powell, L.J. (2022) Adopted utility calculus: origins of a concept of social affiliation. *Perspect. Psychol. Sci.* 17, 1215–1233
- Ullman, T.D. *et al.* (2009) Help or hinder: Bayesian models of social goal inference. *Adv. Neural Inf. Process. Syst.* 22, 1874–1882
- Hamlin, J.K. *et al.* (2013) The mentalistic basis of core social cognition: experiments in preverbal infants and a computational model. *Dev. Sci.* 16, 209–226
- Biro, S. and Leslie, A.M. (2007) Infants' perception of goal-directed actions: development through cue-based bootstrapping. *Dev. Sci.* 10, 379–398
- Feiman, R. *et al.* (2015) Infants' representations of others' goals: representing approach over avoidance. *Cognition* 136, 204–214
- Cesana-Arrotti, N. *et al.* (2020) Infants recruit logic to learn about the social world. *Nat. Commun.* 11, 1–9
- Liu, S. *et al.* (2019) Origins of the concepts cause, cost, and goal in prereaching infants. *Proc. Natl. Acad. Sci.* 116, 17747–17752
- Woo, B.M. *et al.* (2021) Open-minded, not naïve: three-month-old infants encode objects as the goals of other people's reaches. In *Proc. Annu. Meeting Cogn. Sci. Soc.* (43), pp. 514–520
- Luo, Y. (2011) Three-month-old infants attribute goals to a non-human agent. *Dev. Sci.* 14, 453–460
- Hamlin, J.K. (2015) The case for social evaluation in preverbal infants: gazing toward one's goal drives infants' preferences for helpers over hinderers in the Hill paradigm. *Front. Psychol.* 5, 1563
- Tan, E. and Hamlin, J.K. (2022) Mechanisms of social evaluation in infancy: a preregistered exploration of infants' eye-movement and pupillary responses to prosocial and antisocial events. *Infancy* 27, 255–276
- Woo, B.M. and Spelke, E.S. (2020) How to help best: infants' changing understanding of multistep actions informs their evaluations of helping. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*, pp. 384–390
- Brandone, A.C. and Wellman, H.M. (2009) You can't always get what you want: Infants understand failed goal-directed actions. *Psychol. Sci.* 20, 85–91
- Brandone, A.C. *et al.* (2014) Infants' goal anticipation during failed and successful reaching actions. *Dev. Sci.* 17, 23–34

48. Hamlin, J.K. *et al.* (2008) Do as I do: 7-month-old infants selectively reproduce others' goals. *Dev. Sci.* 11, 487–494
49. Hamlin, J.K. *et al.* (2009) Eight-month-old infants infer unfulfilled goals, despite ambiguous physical evidence. *Infancy* 14, 579–590
50. Behne, T. *et al.* (2005) Unwilling versus unable: infants' understanding of intentional action. *Dev. Psychol.* 41, 328
51. Marsh, H.L. *et al.* (2010) Six- and 9-month-old infants discriminate between goals despite similar action patterns. *Infancy* 15, 94–106
52. Dunfield, K.A. and Kuhlmeier, V.A. (2010) Intention-mediated selective helping in infancy. *Psychol. Sci.* 21, 523–527
53. Hamlin, J.K. (2013) Failed attempts to help and harm: intention versus outcome in preverbal infants' social evaluations. *Cognition* 128, 451–474
54. Geraci, A. and Surian, L. (2011) The developmental roots of fairness: infants' reactions to equal and unequal distributions of resources. *Dev. Sci.* 14, 1012–1020
55. Strid, K. and Meristo, M. (2020) Infants consider the distributor's intentions in resource allocation. *Front. Psychol.* 11, 2806
56. Geraci, A. *et al.* (2022) Infants' intention-based evaluations of distributive actions. *J. Exp. Child Psychol.* 220, 105429
57. Dennett, D.C. (1989) *The Intentional Stance*, MIT press
58. Wellman, H.M. *et al.* (2001) Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev.* 72, 655–684
59. Kampis, D. *et al.* (2021) A two-lab direct replication attempt of Southgate, Senju and Csibra (2007). *R. Soc. Open Sci.* 8, 210190
60. Meristo, M. *et al.* (2012) Belief attribution in deaf and hearing infants. *Dev. Sci.* 15, 633–640
61. Surian, L. and Geraci, A. (2012) Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *Br. J. Dev. Psychol.* 30, 30–44
62. Surian, L. and Franchin, L. (2020) On the domain specificity of the mechanisms underpinning spontaneous anticipatory looks in false-belief tasks. *Dev. Sci.* 23, e12955
63. Kaltefleiter, L.J. *et al.* (2021) Evidence for goal- and mixed evidence for false belief-based action prediction in two- to four-year-old children: a large-scale longitudinal anticipatory looking replication study. *Dev. Sci.* 25, e13224
64. Gao, T. *et al.* (2009) The psychophysics of chasing: a case study in the perception of animacy. *Cognit. Psychol.* 59, 154–179
65. Kulke, L. *et al.* (2018) How robust are anticipatory looking measures of Theory of Mind? Replication attempts across the life span. *Cogn. Dev.* 46, 97–111
66. Kulke, L. *et al.* (2018) Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. *Psychol. Sci.* 29, 888–900
67. Krupenye, C. *et al.* (2016) Great apes anticipate that other individuals will act according to false beliefs. *Science* 354, 110–114
68. Woo, B.M. and Spelke, E. (2022) Toddlers' social evaluations of agents who act on false beliefs. *Dev. Sci.* Published online August 23, 2022. <https://doi.org/10.1111/desc.13314>
69. Woo, B.M. and Spelke, E. (2022) Eight-month-old infants' social evaluations of agents who act on false beliefs. In *Proceedings of the 44th Annual Conference of the Cognitive Science Society*, pp. 1184–1189
70. Amodio, D.M. and Frith, C.D. (2006) Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277
71. Schurz, M. *et al.* (2014) Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci. Biobehav. Rev.* 42, 9–34
72. DiNicola, L.M. *et al.* (2020) Parallel distributed networks dissociate episodic and social functions within the individual. *J. Neurophysiol.* 123, 1144–1179
73. Iacoboni, M. *et al.* (2004) Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *Neuroimage* 21, 1167–1173
74. Centelles, L. *et al.* (2011) Recruitment of both the mirror and the mentalizing networks when observing social interactions depicted by point-lights: a neuroimaging study. *PLoS One* 6, e15749
75. Becchio, C. *et al.* (2012) Social grasping: from mirroring to mentalizing. *Neuroimage* 61, 240–248
76. Eskenazi, T. *et al.* (2015) Neural correlates of observing joint actions with shared intentions. *Cortex* 70, 90–100
77. Arioli, M. *et al.* (2021) Increased pSTS activity and decreased pSTS-mPFC connectivity when processing negative social interactions. *Behav. Brain Res.* 399, 113027
78. Isik, L. *et al.* (2017) Perceiving social interactions in the posterior superior temporal sulcus. *Proc. Natl. Acad. Sci.* 114, E9145–E9152
79. Farris, K. *et al.* (2022) Processing third-party social interactions in the human infant brain. *Infant Behav. Dev.* 68, 101727
80. Kanouse, D.E. and Hanson, L.R., Jr (1987) Negativity in evaluations. In *Attribution: Perceiving the Causes of Behavior*, pp. 47–62
81. Vaish, A. *et al.* (2008) Not all emotions are created equal: the negativity bias in social-emotional development. *Psychol. Bull.* 134, 383–403
82. Knobe, J. (2003) Intentional action and side effects in ordinary language. *Analysis* 63, 190–194
83. Quillien, T. and German, T.C. (2021) A simple definition of 'intentionally'. *Cognition* 214, 104806
84. Uttich, K. and Lombrozo, T. (2010) Norms inform mental state ascriptions: a rational explanation for the side-effect effect. *Cognition* 116, 87–100
85. Leslie, A.M. *et al.* (2006) Acting intentionally and the side-effect effect: theory of mind and moral judgment. *Psychol. Sci.* 17, 421–427
86. Pellizzoni, S. *et al.* (2009) Foreknowledge, caring, and the side-effect effect in young children. *Dev. Psychol.* 45, 289
87. Woo, B.M. *et al.* (2017) Social evaluation of intentional, truly accidental, and negligently accidental helpers and harmers by 10-month-old infants. *Cognition* 168, 154–163
88. Weisman, K. *et al.* (2017) Rethinking people's conceptions of mental life. *Proc. Natl. Acad. Sci.* 114, 11374–11379
89. Litoiu, A. *et al.* (2015) Evidence that robots trigger a cheating detector in humans. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pp. 165–172
90. Yasuda, S. *et al.* (2020) Perceived agency of a social norm violating robot. In *Proc. 42nd Annu. Meet. Cogn. Sci. Soc.*
91. Morewedge, C.K. (2009) Negativity bias in attribution of external agency. *J. Exp. Psychol. Gen.* 138, 535–545
92. Waytz, A. *et al.* (2010) Making sense by making sentient: effectance motivation increases anthropomorphism. *J. Pers. Soc. Psychol.* 99, 410–435
93. Hamlin, J.K. and Baron, A.S. (2014) Agency attribution in infancy: evidence for a negativity bias. *PLoS One* 9, e96112
94. Tan, E. and Hamlin, J.K. (2022) Infants' neural responses to helping and hindering scenarios. *Dev. Cogn. Neurosci.* 54, 101095
95. Bakker, M. *et al.* (2015) Neural correlates of action perception at the onset of functional grasping. *Soc. Cogn. Affect. Neurosci.* 10, 769–776
96. Bakker, M. *et al.* (2016) Enhanced neural processing of goal-directed actions after active training in 4-month-old infants. *J. Cogn. Neurosci.* 28, 472–482
97. Grosse Wiesmann, C. *et al.* (2017) Implicit and explicit false belief development in preschool children. *Dev. Sci.* 20, e12445
98. Dunphy-Lelli, S. and Wellman, H.M. (2004) Infants' understanding of occlusion of others' line-of-sight: Implications for an emerging theory of mind. *Eur. J. Dev. Psychol.* 1, 49–66
99. Burns, M.P. and Sommerville, J. (2014) "I pick you": the impact of fairness and race on infants' selection of social partners. *Front. Psychol.* 5, 93
100. Lucca, K. *et al.* (2018) Fairness informs social decision making in infancy. *PLoS One* 13, e0192848
101. Ziv, T. *et al.* (2021) Toddlers' interventions toward fair and unfair individuals. *Cognition* 214, 104781
102. Meristo, M. and Surian, L. (2013) Do infants detect indirect reciprocity? *Cognition* 129, 102–113
103. Choi, Y. and Luo, Y. (2015) 13-month-olds' understanding of social interactions. *Psychol. Sci.* 26, 274–283
104. Cushman, F. *et al.* (2013) The development of intent-based moral judgment. *Cognition* 127, 6–21
105. Kanakogi, Y. *et al.* (2017) Preverbal infants affirm third-party interventions that protect victims from aggressors. *Nat. Hum. Behav.* 1, 1–7