Foundations of theory of mind and its development in early childhood

Hannes Rakoczy

Abstract | Theory of mind is the human conceptual capacity to understand other people as agents who have subjective mental states such as beliefs, desires, and intentions. It is the basis of distinctively human forms of social understanding and interaction that are essential for communication, cooperation, and culture. In this Review, I summarize the current state of research about the emergence and development of theory of mind in early childhood. I describe the typical developmental trajectory and review findings about the cognitive, linguistic, social and neural foundations of theory of mind development. Finally, I review an ongoing debate regarding whether there are different — implicit versus explicit — forms of theory of mind that develop independently, and conclude by providing an outlook on future challenges and perspectives for research in this area.

The way humans view agents, be it others or themselves, differs radically from the way humans view the rest of the world. This difference is because agents feel, perceive and think. Agents see the world from their own subjective perspective and they rationally plan, form intentions and act accordingly. This way of seeing others as rational subjects with individual perspectives on the world is termed theory of mind (ToM).

At the conceptual heart of ToM lies meta-representation: the capacity to represent that and how agents represent the world from their own points of view. Meta-representation enables a person to see, judge, or believe that others see, judge, or believe[1]. The prototypical form of meta-representational ToM is 'belief–desire psychology'[2]. Belief–desire psychology refers to the ability to represent how others believe the world is, how they desire it to be, and how they might rationally act to realize their desires in light of their beliefs. Such belief–desire reasoning is ubiquitous in everyday life. For example, it is easy to make sense of why someone engages in the seemingly unpleasant action of lifting heavy weights; it is a rational means of fulfilling a desire (to gain fitness) in light of a belief (that lifting weights fosters fitness).

In its mature form, subjective meta-representation comes with a cluster of four specific signatures: difference, incompatibility, mis-representation, and aspectuality (TABLE 1)[3]. First, when an observer meta-represents an agent's point of view, that point of view can differ from the observer's (for instance, representing 'They can see outside the window, but from where I stand, I cannot'). Second, the agent's and the observer's perspective might not only be different, but incompatible in their contents. This situation occurs, for example, when the desires of both agents cannot be fulfilled at the same time (for instance, 'She wants it to rain now, but I do not want it to rain'). Third, in the attribution of beliefs and other cognitive propositional attitudes, meta-representation implies the possible ascription of mis-representation. That is, one can represent another's false belief, which can still guide action (for instance, although it is not raining, 'He believes that it is raining and so he takes his umbrella'). Finally, in the assignment of propositional attitudes to other agents, it matters how — that is, under what aspects — that agent represents the relevant objects and situations. Suppose Eve sees Adam enter the house. Adam is the newly elected mayor of their town, but Eve does not know this yet. Eve therefore believes that Adam is in the house, but not that the mayor is in the house. This belief can drive Eve's behaviour: if she were searching for Adam, she would go into the house, but if she were searching for the mayor, she would not. In the development of ToM, these characteristic signatures of meta-representation do not necessarily emerge at the same time. Rather, development proceeds through intermediate stages in which children acquire various aspects of the meta-representational framework of ToM.

ToM has general and far-reaching consequences for people's social lives. With ToM, a second-person (embodied by the word 'you') perspective becomes possible[4]. I, as one agent, can now understand you as another subject with whom I can engage inter-subjectively. This second-person perspective supports novel types of social interactions: people can recognize each other as rational subjects that can share views on the world, talk to one another, criticize each other, and/or relate empathically towards each other. Relatedly, ToM enables first-person plural (embodied by the word 'we') perspectives. When people understand each other as rational agents, they can not only ascribe individual intentional states to each other, but also enter into shared intentionality, form

Institute of Psychology, University of Göttingen, Göttingen, Germany.

e-mail: hannes.rakoczy@psych.uni-goettingen.de

https://doi.org/10.1038/s44159-022-00037-z

Table 1 | **The four signatures of subjective meta-representation**

Signature	Definition	Example
Difference	Representing that others' perspectives may differ from one's own	Person A represents: Person A cannot see that it is raining, but Person B can see that it is raining
Incompatibility	Representing that the content of others' perspectives can be incompatible with one's own	Person A represents: Person A wants it to rain now, but Person B does not want it to rain now
Mis-representation	Representing that others' perspectives can be inaccurate or false	Person A represents: It is not raining but Person B thinks it is raining and brings their umbrella
Aspectuality	Representing that agents represent objects and situations under specific aspects	The umbrella carried by person B is a work of art made by a famous artist. Person A knows this, but person B does not. Person A thus represents: B believes they are carrying an umbrella but they do not believe they are carrying a work of art.

and pursue joint goals, and perform joint cooperative actions[5,6]. Finally, ToM grounds novel forms of subjectivity and a new first-person singular (embodied by the word 'I') perspective. As an agent with the capacity for meta-representation, I can apply my explicit mental state concepts to myself and consequently I gain new insights about and control over my own desires, and thus new forms of higher-order self-consciousness, reflection, and self-regulation[7,8].

ToM also has specific real-life consequences. First, the development of ToM competence goes along with general measures of children's peer social skills in early and middle childhood. These measures include abilities such as leadership, skills for joining new groups, welcoming new members into existing groups, and standing up for one's own opinion in exchanges with peers[9]. Second, ToM specifically predicts communicative competence. In particular, children who are more advanced in their ToM competence are better at conversation, argument and persuasion than children with less advanced ToM. Children with more advanced ToM are more sensitive to conversation partners' points of view and tailor their arguments and persuasion attempts accordingly[10,11]. Third, ToM competence is related to the quality of peer relationships: children with more advanced ToM are rated as more likeable and popular among their peers[12]. Fourth, children who are more proficient at ToM tasks tend to act more prosocially, including comforting, sharing or helping other individuals[13]. Finally, preschool ToM competence predicts achievement in primary school, a relationship that is possibly mediated by social competence, in that preschool ToM abilities enable subsequent social competence development, which in turn contributes to school achievement[14].

However, ToM does not have exclusively desirable real-life consequences. Insight into the subjective life of others can be used for all kinds of purposes from cooperative and empathic to selfish and manipulative. For example, advanced ToM reasoning enables children to engage in more sophisticated acts of deception and manipulation such as lying[15,16]. In addition, the aggressive and manipulative behaviour of some playground bullies might build on their sophisticated ToM functioning[17]. Overall, ToM is a foundational form of social cognition with wide-ranging consequences for virtually all domains of everyday social life.

In this Review, I describe milestones of development in early childhood through which children acquire the elements of ToM. I discuss determinants of ToM development, with a particular focus on language and executive function. Of special interest is newer research on implicit forms of ToM, demonstrated through simplified non-verbal tasks, that supposedly emerge very early in infancy. Original empirical findings on implicit ToM as well as problems with replication in this area are reviewed, and recommendations for future research directions are outlined.

Developmental milestones

ToM, in its most basic and rudimentary forms, develops in infancy and early childhood in the first four to five years of life. In the following, I review crucial developmental milestones (FIG. 1).

Emergence of basic ToM in infancy. Even in the first months of life, infants display social perception and interaction with other humans. For example, infants are specifically interested in and tuned to visually processing faces[18] and biological motion[19]. From two months on, infants socially smile at other people and engage in coordinated social proto-conversations with them[20]. There is some evidence that newborns are already capable of simple forms of imitation such as copying facial gestures[21]. However, these imitation findings have been called into question by failures to replicate the original findings in large-scale replication studies[22–24]. Even if reports of infant imitation are less reliable than originally thought, infants display considerable early social perception. But these behaviours do not necessarily involve any form of ascribing mental states to others and therefore are not ToM, even in the broadest sense.

Basic forms of ToM begin to emerge towards the end of the first year of life in the context of what is sometimes termed the '9-month revolution'[25]. Children at this age begin to exhibit a suite of new capacities: they represent what other agents perceive[26], what goals other agents pursue[27], and they form expectations about how other agents will rationally act to fulfill their goals in light of their perceptual access[28]. The emergence of these capabilities marks the advent of children's perception–goal psychology[29]. Perception–goal psychology is the capacity to track others' perceptions of their surroundings, the goals they have, and how they pursue the latter in light of the former.

Evidence for an emerging understanding of perception at 9 months of age comes from various sources. For example, children begin to follow the gaze of other agents in systematic and differential ways: they follow an agent's head turn only when the agent can actually see (has their eyes open rather than closed, or wears a transparent rather than an opaque blindfold[26]). Children do not follow all head movements, only those that are related to the attention and perception of others.

Similarly, when confronted with an ambiguous action by another agent that could be directed towards one of two objects, children at this age take the agent's perceptual access into account. For instance, if the infant can see two objects but only one is in the agent's view, infants expect the agent to act towards the only object that is visible for them[30]. This behaviour implies some representation of what the other agent can see.

Also at around 9 months, infants display an emerging understanding of goal-directedness. Evidence for this understanding comes in part from studies in which infants watch animated scenarios of geometrical figures that move in such a way that adults perceive them as agents acting in rich intentional ways[31]. Like adults, infants expect the animated agents to act rationally in the pursuit of their goals, as indicated by their looking times. For example, when a geometrical figure has repeatedly approached a particular object on a particular path, infants assume that the figure will pursue the same goal (reaching the object) in flexible and efficient ways when the paths change. Infants are surprised and look longer when the figure continues to take the same path as previously if the path no longer leads to the goal object or if the path leads there in inefficient ways[32].

In spontaneous interactions with other agents, infants also indicate sensitivity to goal-directed intentional action. When infants see another agent failing in their goal pursuit, they spontaneously intervene by offering help. For instance, if an adult accidentally drops an object needed to continue their action, the infant will collect and give the agent that object. However, infants do not do so in analogous control situations in which the agent does not pursue the goal in question (for example, if they voluntarily drop the object)[33]. That is, infants differentiate an agent who pursues a given goal and needs help to achieve that goal from an agent who behaves similarly but is not in need of help to achieve their goal.

Finally, a rich body of evidence regarding infants' goal understanding comes from imitation studies. Infants from around 9 months of age begin to engage in spontaneous imitation of others' instrumental, goal-directed actions with objects[25]. The way they imitate indicates that they do not just copy the superficial behaviour, but instead imitate the actions as goal-directed intentional activities. The clearest example of this phenomenon is so-called rational imitation. When infants see an ambiguous action — for example, an agent switches on a light using their head — the infant interprets the action as a means to a goal or as a goal in itself depending on the context and imitate accordingly. When the agent did not have their hands available, infants imitate the action by using their hand to turn on the light, indicating that they saw the use of the head as a means of turning on the light. By contrast, when the agent had their hands available but used their head anyway, infants imitate using their head, indicating that they saw this unusual act as somehow a goal in itself[34].

From the time it emerges, infants' perception–goal psychology does not only reveal itself in third-person interactions, but also in characteristic forms of first-person plural shared intentionality. Once they understand what others see, infants engage with them in joint attention by coordinating and aligning what they and another agent see to look at things together[35]. Once infants understand what goals others pursue and how they act intentionally towards them, they form joint intentions to act together towards a shared goal[5,36].

The conceptual scheme of perception–goal psychology that children develop from around 9 months of age qualifies as the most basic form of ToM. Perception–goal psychology involves the attribution of simple mental states, perceptions and goals, and principles of rational, teleological action explanation and prediction. Infants at this age can represent that agents act to fulfil their goals given their perceptual access. Perception–goal psychology is therefore a precursor, and potentially a foundation, for belief–desire psychology. Perception forms a precursor to belief and goals form precursors to desires. However, perception–goal psychology is characteristically limited in its conceptual power. With respect to the four characteristic signatures of meta-representational ToM, perception–goal psychology involves only the signature of difference. Thus, infants understand that different agents can see different things or can pursue different goals. They do not yet understand the other, more stringent aspects of fully fledged meta-representation: that representations can be mutually incompatible, can be false, and can involve fine-grained aspectual information.

The distinction between two types of perspective-taking helps illustrate that perception–goal psychology does not yet constitute full meta-representation (FIG. 2). When acquiring perception–goal psychology, children come to master so-called level I perspective-taking[37]: they understand that different viewers can see different things in a given scene. For instance, they understand that one person might be able to see a tree in front of a house, whereas a second person can see only the house

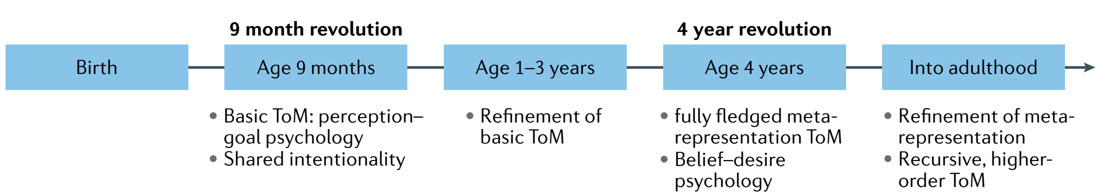

Fig. 1 | **Milestones in the development of theory of mind.** The most basic form of theory of mind (ToM), perception–goal psychology, emerges from around 9 months of age. It allows agents to represent that others may have different perceptual perspectives on the world, and different goals, and act accordingly. Fully fledged meta-representation emerges later, from around 4 years of age, in the form of belief–desire psychology. It involves an appreciation that others subjectively represent the world in fine-grained aspectual ways that may be incompatible with one's own view and that may be false.

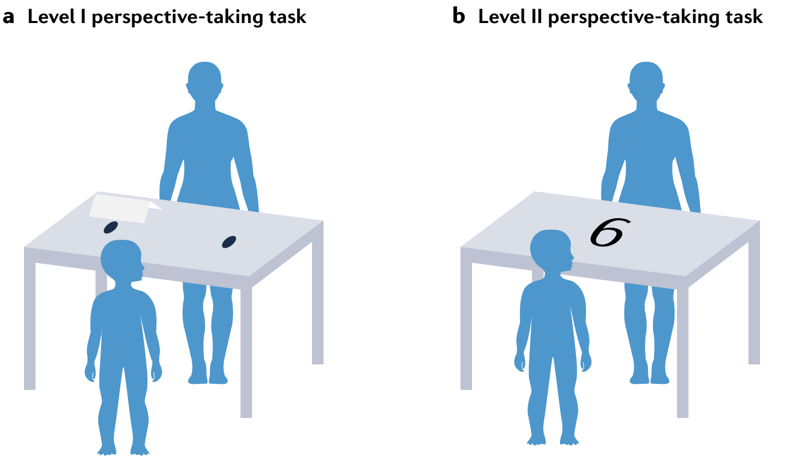

Fig. 2 | **Level I and level II perspective-taking tasks. a** | In a typical level I perspective-taking task, the participant (bottom) sees two objects on the table, while another agent (top) can only see one of the objects because the other one is occluded from their perspective. The crucial question to the participant is which object(s) the other agent can see. This task measures an understanding of what another agent perceives[37,164]. **b** | In a typical level II perspective-taking task, the participant and another agent see the same object (the numeral on the table), but it appears differently from the different perspectives (either as a 6 or as a 9). The crucial question to the participant is what the object looks like to the other agent. This task measures an understanding of how something looks to another agent[37,164].

but not the tree because it is behind the house from their perspective. Only later, when children acquire belief–desire psychology, do they come to master 'level II perspective-taking': they understand that different viewers can see the same thing or scene in different, incompatible, fine-grained ways. For instance, they come to understand that a figure that appears to one viewer as a '6' will appear as a '9' to someone sitting opposite.

The 4 year revolution. In the second and third years of life, perception–goal psychology and basic ToM are enriched in various ways. For example, children acquire a more flexible and nuanced understanding of the ways in which different agents can differ in their goals and preferences[38].

Fundamental developmental changes happen at around age 4, when new conceptual structures emerge. This '4 year revolution' marks the onset of a fully fledged meta-representational ToM. In the course of this major transition, children gain the ability to succeed in a cluster of tasks that all require meta-representation. Among these are false-belief tasks, often considered a litmus test of mature ToM. These tasks explicitly require children to ascribe a subjective mis-representation to another agent and explain/predict their actions accordingly. In the most well known version, the child sees that another agent puts an object into one box and then the object is transferred to another box in the agent's absence (FIG. 3a). The critical test question asked of the child is where the agent believes the object is or where the agent will search for the object[39]. Children younger than four systematically answer incorrectly; they say that the agent thinks the object is where it really is and accordingly say that the agent will search in the second box. However, from around age 4, children succeed at the task by indicating that the agent will search in the first box. That is, children ascribe to the agent a subjective mis-representation: a false belief that deviates from reality and is incompatible with the child's own perspective. In a related task, the child sees a familiar container (such as an egg carton), is asked what is inside (the child answers 'eggs'), and then learns about its unexpected content (a pen). The child is then asked what they initially thought was in the box, and what another naive agent would think is in the box. To answer correctly, the child has to meta-represent how the world (wrongly) appeared to them previously, or would appear to another person. Children younger than four systematically answer incorrectly; they claim that they initially thought a pen was in the box and that another agent would think so also. From around age 4, children succeed by ascribing to both their former self and another agent the mistaken belief that the box contained eggs[40] (FIG. 3c).

Other tasks that require an understanding of incompatible perspectives are also mastered around this age. These tasks include appearance–reality and level II perspective-taking tasks. In appearance–reality tasks, children are confronted with an object that looks different from what it really is, such as a rubber eraser that looks like a walnut, and are asked both what the object really is and what it looks like (FIG. 3d). Before the age of four, children tend to answer in reality-based ways, responding 'it is a rubber eraser and looks like a rubber eraser'. From around the age of 4, children explicitly acknowledge the incompatible perspectives between appearance and reality ('it is a rubber eraser, but it looks like a walnut')[40,41]. Similarly, from the same age, children begin to master level II perspective-taking tasks by acknowledging incompatible visual perspectives (for example, 'it looks like a 6 from here, but like a 9 from there'; FIG. 2b).

Children's emerging meta-representation is also sensitive to the aspectuality of mental representations. From around 4 years of age, children are newly able to track how an agent represents a given object or situation and consequently how they will act. For example, in several studies children were confronted with events involving objects with two aspects or identities, such as an object that is both a pen and a rattle (looks like a pen, can be shaken as a rattle) (FIG. 3b). In the task, the child watched the following scenario: an agent saw the object (as a pen) disappear into one box; then the object was transferred to another box in such a way that it could be heard but not seen (the agent heard a rattle being moved). The child was aware of the two aspects of the object at all times and knew that the pen/rattle had been moved to the second box. However, the agent was not aware that the pen and the rattle are the same object with a dual identity. The critical test question to the child is where the agent would look for the pen. Children from age 4 answered correctly that the agent would look for the pen in the second box if the agent knew about the dual identity, but would look in the first box if unaware of the dual identity[42].

Mastery of these different tasks from around the age of 4 indicates that children grasp various signatures of the subjectivity of mental representation. Representations held by different agents can have mutually incompatible content, can be false with respect to reality, and can represent objects in fine-grained aspectual ways. Crucially, children's understanding of these different characteristics of subjective mental representation emerges in a systematic and coherent fashion. Children come to master all these tasks — despite their dramatic differences in topic, format, and surface structure — at the same time and performance across tasks is highly correlated[40,42,43]. Thus, the 4 year revolution marks a major cognitive transition and the emergence of a meta-representational conceptual capacity.

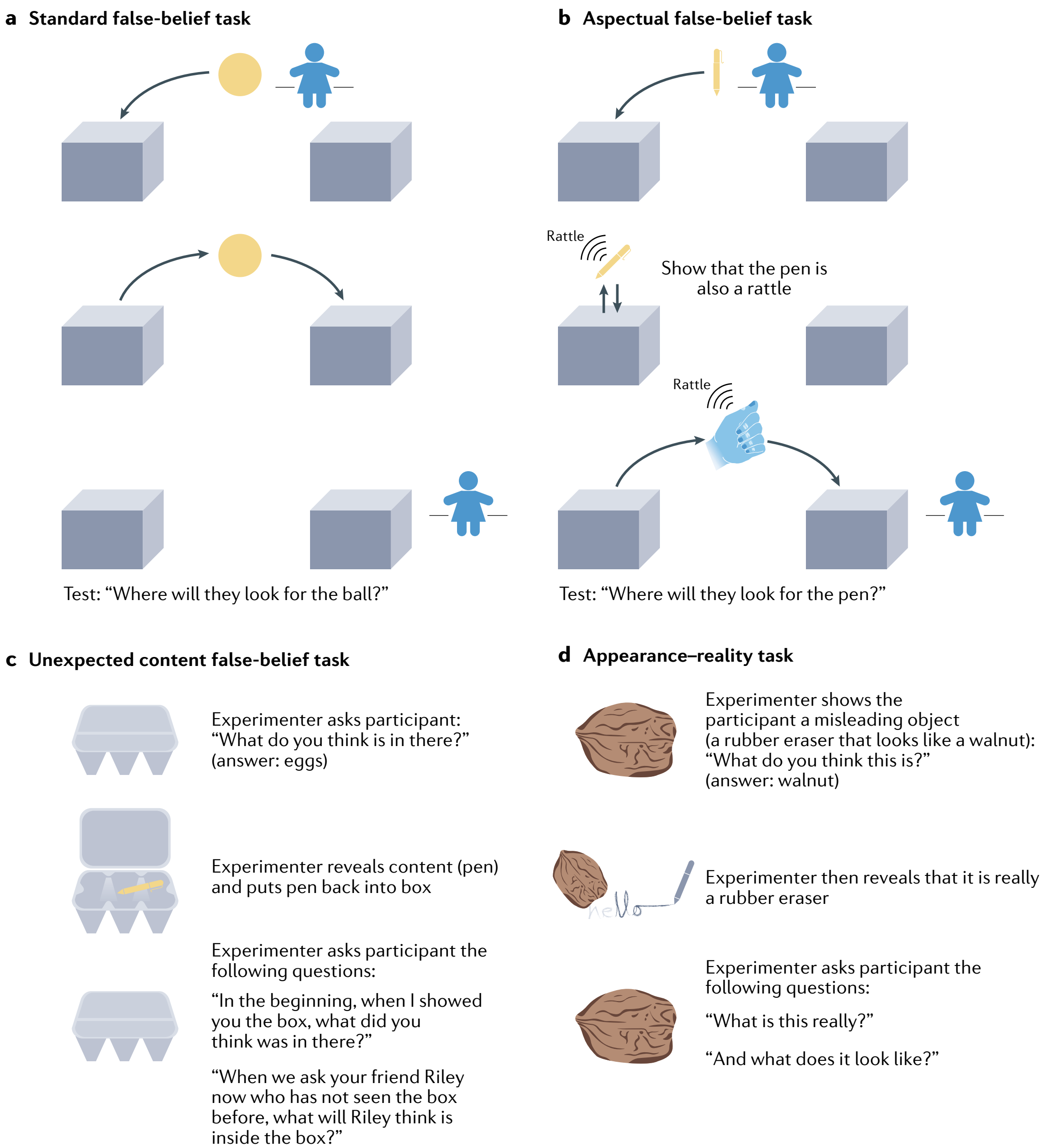

Fig. 3 | **False-belief and related theory of mind tasks. a** | The child (participant) sees another agent put an object into a box. The agent leaves and the object is transferred to another box. The child is then asked where the agent will look for the object. To answer correctly, the child needs to meta-represent how the agent represents the location of the object. **b** | The child watches an object with two identities (a pen that is also a rattle) be put into one box while shown as a pen and then transferred to another box while shaken as a rattle. The child therefore knows that the pen/rattle is now in the second box. Another agent also witnesses the initial placement of the object (as a pen) in the first box and its relocation (as a rattle) to the second box but is unaware of the dual identity of the object. The child is then asked where the agent will search for the pen. To answer correctly ('The first box'), the child needs to meta-represent how the agent represents the placement and relocation of the object. **c** | The child sees a familiar container (such as an egg carton), is asked what is inside (answers 'eggs'), and then learns about the unexpected content (a pen). The child is then asked what they initially thought was in the box, and what another naive agent would think is in the box. To answer correctly, the child has to meta-represent how the world (wrongly) appeared to them previously, or would appear to another person. **d** | The child initially sees a misleading object (such as a rubber eraser that looks like a walnut) and is asked what she thinks the object is (answers 'walnut'). The child then learns that the object is actually something else (a rubber eraser). The child is then asked what the object really is and what it looks like. To answer correctly, the child needs to meta-represent how the object appears, by contrast to what it really is.

With this newly acquired capacity, children understand that others and themselves are rational agents with subjective perspectives on the world — perspectives that can differ from and be incompatible with each other, deviate from reality, and that represent the world in fine-grained, aspectual ways.

These core capacities of meta-representation have numerous associated capacities. For example, once children understand the possibility of subjective, potentially false representation, they make practical use of this in attempting to strategically implant mis-representations in others through lies and deception[44]. In a broader context, children's emerging meta-representation is not restricted to understanding mental representation. In the course of the 4 year revolution, children acquire a broader meta-representational conception of all kinds of representations — including mental, linguistic, or pictorial ones. For example, children acquire meta-linguistic awareness that some words (homonyms) can have multiple meanings, and meta-pictorial awareness that some pictures can have multiple interpretations[45–47]. Meta-representation is also accompanied by new forms of self-consciousness and time consciousness as children develop a diachronic sense of themselves in time, with episodic memory of past experiences, and episodic foresight into potential future ones[48–50]. Finally, children at the age of 4 have a basic meta-representational grasp of emotions such as surprise (finding out that one's belief was false) and frustration (finding out that one's desire was not fulfilled)[51].

Continuing development. Despite the advances in ToM at around age four, children have not achieved adult-like meta-representational capacities at this age. For instance, the understanding of complex and subtle emotions, including fake or concealed ones, continues to develop into middle childhood and adolescence[52,53]. Similarly, higher-order recursive meta-representation (for example, the ability to represent and understand 'A thinks that B thinks that C desires that…') develops gradually over childhood, adolescence and adulthood[54,55]. Complex pragmatic understanding of indirect speech acts such as irony or sarcasm and subtle discourse likewise have a protracted development over the preschool years[56,57]. These different forms of sophisticated ToM all require some form of higher-order, recursive mental state ascription[58]. Thus, their cognitive complexity is plausibly why they develop in more protracted ways than basic ToM.

In the broader lifespan perspective, many of these more complex forms of ToM, such as recognizing subtle emotions or understanding discourse pragmatics, are not an all-or-nothing matter. Rather, differences in degree of ToM abilities persist even in adulthood[59], and they follow U-shaped trajectories across the lifespan: protracted developmental acquisition over childhood and adolescence, mirrored by analogous decline in older age[60].

Developmental determinants

ToM development is influenced by and builds on general cognitive, neural, and environmental determinants. These include capacities for executive function and language, as well as social and demographic factors. Some of these determinants constitute enabling conditions for the emergence of ToM competence. For example, language seems to be necessary to acquire fully fledged meta-representational ToM[61]. Other determinants are not necessary for ToM competence, but influence inter-individual differences in ToM performance in more fine-grained ways. For example, it is certainly not necessary for a child to have siblings for the child to acquire ToM, but having siblings accelerates ToM development[62].

Executive function. Executive function refers to higher-level cognitive processes involved in the control of action, including working memory, inhibition, planning and set shifting. There are close developmental connections between general executive function and ToM over the lifespan[63]. The most conclusive evidence for this link includes longitudinal findings that executive function (such as conflict inhibition) early in development predicts ToM performance in false-belief and related tasks later in development, but not vice versa[64,65]. Furthermore, executive function abilities mediate how fast and proficiently children acquire ToM from training experience over time[66]. Taken together, these individual studies, confirmed by corresponding meta-analyses[67], suggest a crucial role of executive function in the emergence of ToM[68].

These findings raise the crucial question of the underlying proximal mechanisms through which executive function enables the emergence and development of ToM. One possibility is that executive function is part of the cognitive foundation that enables the flexible coordination and confrontation of multiple perspectives and is fundamental for meta-representation. Support for this view comes from studies showing that executive function is specifically related to ToM tasks that require the handling of diverging and incompatible perspectives by different agents, and that 'conflict inhibition' executive function tasks account for these correlations[69–71]. In one typical conflict inhibition task for children, the participant receives commands to perform simple manual actions from two puppet agents, a bear and a dragon, and is supposed to follow the bear only while ignoring the dragon[72]. Such tasks involve working memory (remembering who said what and whom to follow) and inhibition (of the impulse to follow the commands given by one agent), which are also needed for representing and handling conflicting perspectives in ToM tasks. Working memory is required to keep in mind different perspectives and inhibition is required to suppress conflicting representations that interfere with the solution to a given problem (for instance, suppressing the representation of reality when answering a question about the subjective false belief of another agent)[71].

Another potentially complementary possibility is that a more general cognitive capacity for recursive embedding of representations underlies both executive function and ToM development. Development of this capacity for recursion would enable more complex rule-following in executive function tasks and recursive meta-representational embedding in ToM. That is, as children learn to form and follow embedded rules (for example, 'IF [IF a THEN b] THEN [IF c THEN d]') they can also represent embedded beliefs (for example 'She believes [that a]')[58].

It is therefore clear that executive function plays a fundamental role in the emergence and execution of ToM. Plausibly, the proximal mechanisms of this influence include the facilitating role of executive function in flexibly coordinating different perspectives and in recursive reasoning, but more research is needed to delineate these proximal mechanisms in more fine-grained ways.

Language. The crucial role of language in ToM development is documented by a growing body of evidence with diverse measures, methods and populations[61]. First, numerous findings show that the development of general language competence, including syntactic, semantic and pragmatic capacities, is correlated with ToM performance in typical and atypical development[73]. Second — and more informatively, because it is not merely correlational evidence — training studies reveal that ToM development accelerates after linguistic training (confirmed by a corresponding meta-analysis[74]). Children between the ages of three and four who receive even short sessions of linguistic experience subsequently show better performance on ToM tasks. This improvement is relative to children who had an equally long training session of an equal amount of non-specific non-linguistic experience[74]. According to these studies, the specific components that seem to accelerate ToM development are experience with general discourse pragmatics (such as engaging in turn-taking discourse in which each speaker takes into account and responds to the other's speech acts), with semantics of specific words (mental state verbs like 'know' or 'think') and with specific syntactic structures. Of particular relevance are so-called that-complementation syntactic constructions, used to report propositional attitudes (for example, 'Alice believes that [the apple is in the box]'). These constructions enable speakers to embed any kind of proposition (here, 'the apple is in the box') as the content of a propositional attitude (here, Alice's belief)[74,75].

Third, studies with children and adults who are deaf point to a substantial role of language in ToM development. Comparisons of deaf children who grow up in hearing families without native signers and deaf children who grow up in families with native signers reveal fundamental differences: the latter show typical language development, quickly become fluent native signers, and show typical ToM development. By contrast, children in the former group show delayed language development and equally delayed ToM development[76,77]. These different trajectories in ToM development in two populations that differ primarily in their linguistic experience thus suggest a crucial role for language in the emergence of ToM.

Relatedly, studies with Nicaraguan Sign Language provide a quasi-experimental view on language evolution and its consequences for ToM. Nicaraguan Sign Language is a language that emerged in the 1970s, initially as a gesturing and home signing system among deaf children, and developed quickly into the mature linguistic system of a growing deaf community[78]. Signers from different cohorts were compared on their mastery of complex linguistic constructions used to report mental states, and their performance on non-verbal false-belief tasks. In both tasks, signers who acquired Nicaraguan Sign Language when it was semantically and syntactically more complex outperformed signers of a less-complex version of the language from an older cohort. However, over time as both cohorts progressed in their linguistic proficiency and complexity, the performance gap between them narrowed[79]. These findings are another piece of evidence for a substantial role of language in the development of ToM; as language develops so does ToM.

Finally, perhaps language matters not only in the emergence of ToM capacities but also remains the medium for ToM cognition throughout the lifespan. When adults had to engage in a linguistic dual task in which they had to listen to and repeat sentences while solving ToM problems, their performance radically declined relative to when the dual task was a non-linguistic rhythm-tapping task. This decrement in performance brought adults almost to the level of 3 year olds[80] (but see REF.[81]).

The evidence reviewed above is compatible with two potentially complementary possibilities for the proximal mechanisms underlying the relationship between language and ToM development. Language — particularly language use in discourse and conversation — might be a particularly rich basis for learning about mental states and perspectives. Beyond that role, the specific syntactic devices of language could supply the representational structures and the medium to make possible the meta-representation of propositional attitudes[82,83]. Similar views that ascribe a substantial role to language in enabling and shaping higher cognition have recently been put forward in other domains such as numerical, spatial or modal cognition[84–86].

Neural foundations. In addition to cognitive correlates and determinants, there are concomitant neural changes that support the development of ToM. Social-cognitive neuroscience has uncovered characteristic neural correlates of adult ToM functioning. In adults, ToM is subserved by a network that encompasses the frontal and tempo-parietal areas of the brain[87–91]. Electroencephalographic (EEG) work indicates that representing others' beliefs involves specific temporal signatures such as frontal and parietal late slow waves[92]. Finally, neurophysiological studies have found single neurons in the dorsomedial prefrontal cortex that are specifically tuned to meta-representing others' beliefs[93].

The crucial questions in the development of ToM concern the neural correlates and how neural changes support cognitive changes. Recent research suggests that functional and connectivity changes across the ToM network, in particular between the right temporo-parietal junction (rTPJ) and more frontal areas, might be the crucial neural underpinning of the emergence of fully fledged ToM from around 4 years old[94]. EEG studies show that characteristic processes underlying ToM processing (such as the frontal late slow wave) take adult shape only from around the age of 4 (REF.[92]). Diffusion-weighted magnetic resonance imaging (MRI) studies suggest that changes in white matter structure, particularly the connectivity between the frontal and temporo-parietal regions, underpin the cognitive changes in the course of the 4 year revolution[95,96].

Social determinants. Finally, ToM development is also shaped by a number of social determinants[97,98]. Some evidence for the role of the social environment can be seen in the findings of poor ToM capabilities in those deaf children who are not exposed to sign language and therefore do not have access to social communication of the perspectives of other agents. Additional factors in children's immediate and broader environments also affect the development of ToM.

General socio-demographic factors that positively correlate with ToM development include socioeconomic status[99] and the number of siblings in the house[62,100]. Regarding potential proximal mechanisms, it is plausible to assume that both higher socioeconomic status and greater number of siblings increase social, in particular linguistic, learning opportunities and therefore foster ToM development.

Other specific social factors that shape ToM development pertain to parenting style. The more parents talk to their children about mental states, the more proficient the children become at ToM[101,102]. Relatedly, many studies (and a corresponding meta-analysis[103]), reveal that parents who are more competent at identifying and responding appropriately to the thoughts and feelings of their children ('mind-minded' parents), have children with higher ToM proficiency[104]. Parental mental-state discourse and mind-mindedness plausibly supply children with more social-cognitive experiences in which to learn about mental states more generally. Furthermore, they enable children to experience and acquire the requisite linguistic means to represent and talk about mental states.

In addition, a large body of evidence from cross-cultural studies suggests that ToM develops in largely analogous ways across cultures[105,106]. However, minor characteristic differences have been found in the order of acquisition of various sub-components of ToM. For example, whereas children in individualist cultures master some tasks of understanding opinion diversity before they master analogous tasks of understanding knowledge access, children from collectivist cultures reveal the reverse developmental pattern[107,108].

All in all, ToM development is crucially shaped by social factors. Across cultures, given the requisite linguistic and social input, all children acquire ToM in largely analogous ways. But the order and speed of development are influenced by large-scale cultural, and small-scale family factors.

Implicit theory of mind

As described so far, ToM refers to a high-level conscious capacity that is tapped by explicit measures. These measures include direct tasks that involve verbal abilities (that is, verbal abilities form the dependent measure or verbal instructions are used) and tasks that tap high-level strategic decision-making, planning and interaction. However, one fundamental concern with such tasks is that they might mask the ToM competence of young children owing to extraneous performance factors[109].

Measures and results. Over the past 20 years, implicit measures of ToM, with radically reduced linguistic and other cognitive task demands, have been developed (TABLE 2). Such measures include the following four kinds of tasks. In violation-of-expectation looking-time tasks, infants or young children watch a scene in which an agent forms a mental state (such as a true or false belief), as in a standard (explicit) false-belief task (FIG. 3a). However, unlike in the false-belief task, the participant is not asked what the agent will do but rather is shown the agent acting in accordance with their belief (expected) or not acting in that way (unexpected). The child's looking time is measured and compared between the expected and unexpected conditions. If the child looks longer at the unexpected outcome, this suggests that they

Table 2 | **Overview of measures of implicit ToM and the original and replication findings**

Measure	Populations tested	Original findings	Replication findings
Violation-of-expectation looking time	Infants	Children look longer at unexpected (agent acting inconsistently with their belief) than at expected (agent acting consistently with their belief) outcomes[110]	Mixed, including successful[159], partial[128] and non-replications[129–131]
Anticipatory looking	Infants and children (ages 1–10 years) and adults and non-human primates	Infants, children, adults and non-human primates spontaneously anticipate based on the agent's belief[111–113,160,161]	The basic findings could largely not be replicated in direct large-scale replication studies[134,137–139], including a self-replication[140]
Interaction	Children (from age 2)	Children in their communicative or cooperative interaction with another agent adapt their responses to the partner's belief; they respond differently, for example, when the partner has a false belief as compared to when she has a true belief[114,115]	Mixed, including one positive[162] and several failed replications[128,142,144]
Altercentric bias	Adults	Participants are slower and more error-prone in performing a given task when the perspective of another agent who is present but irrelevant to the task is inconsistent with their own (relative to when it is consistent)[116]	Mixed, including successful[163], partial[145] and non-replications[146]

formed implicit expectations that the agent would act in accordance with their belief and are surprised when this expectation is not fulfilled[110].

Anticipatory looking tasks use a similar paradigm in which participants watch videos in which an agent forms a mental state (such as a true or false belief). In the video, the agent then sets out to perform an action and eyetracking is used to measure whether the participant spontaneously anticipates where the agent will act next by looking towards a relevant part of the screen. If the participant systematically anticipates the agent's actions based on their belief, the participant has formed implicit expectations about the agent's mental state[111–113]. This task is suitable for evaluating ToM in infants, older children, adults and even non-human animals.

A third class of implicit ToM tasks is interaction tasks, in which participants are involved in a communicative or cooperative interaction with another agent. This agent forms a mental state (such as a true or false belief regarding the contents of a box) and experimenters measure whether participants spontaneously take the agent's belief into account in their interaction with the agent (for instance by helping or by interpreting the agent's communicative acts accordingly)[114,115].

Finally, in altercentric interference tasks, the participant is required to make a judgement (for example, 'how many dots are there?') in the incidental presence of another agent whose perspective is either congruent or incongruent with the participant's perspective (for example, sees the same or a different number of dots than the participant). Participants' reaction times and error rates are measured to evaluate altercentric interference: slower and more error-prone responses are expected when the agent's perspective is incongruent with the participant's perspective than when it is congruent. If such a pattern is found, it suggests that participants spontaneously and implicitly take into account the other agent's perspective, even though it is irrelevant to the task[116,117].

The first wave of studies with implicit measures, including violation-of-expectation, anticipatory looking and interaction, suggested that even infants have meta-representational ToM capacities. Once the tasks were suitably simplified, infants systematically distinguished between true- and false-belief conditions and formed expectations and acted accordingly[118,119]. Initial evidence with tasks that can also be used with adults — anticipatory looking and altercentric interference — also suggested that capacities of ToM remain in operation throughout the human lifespan in the form of unconscious and spontaneous processes that reveal themselves in implicit tasks[112].

Theoretical implications. The original positive findings with implicit measures have had far-reaching theoretical repercussions. They led to the development of, and have been taken as evidence for, ambitious novel accounts.

According to nativist accounts, meta-representational ToM is largely innate and in operation very early in development, much earlier than suggested by explicit tasks[119–121]. Children fail at standard false-belief and related tasks until the age of 4 merely owing to performance limitations: the tasks have extraneous task demands (such as verbal and inhibitory) beyond the requirement for meta-representation. If such task demands are removed, children should be able to express their existing meta-representational ToM and pass the tests. Implicit tasks yield exactly this pattern of findings and are thus taken as strong evidence for nativism.

By contrast, dual-process theories assume that there are at least two types of ToM that develop independently[95,122–125]. Type I processes develop early, operate implicitly and unconsciously, and remain in operation throughout the lifespan in spontaneous and indirect tasks. These processes involve a simple ToM, a basic way of representing mental states that goes beyond mere perception–goal psychology (though not incorporating belief–desire psychology). Type II processes — fully fledged meta-representational ToM — develop later, are based on language and executive function capacities, and operate in explicit and conscious ways across the lifespan. Findings with infants and adults that show some dissociations between indirect and direct tasks, and therefore supposedly between type I and type II processes, have been taken as evidence for this dual process framework[126,127].

Replicability and validity. Since the publication of the original implicit ToM tasks, multiple studies with infants, children and adults have either failed to replicate the original findings (calling into question their reliability), or have found that the effects are not stable under more stringent conditions (calling into question their validity).

There is a growing body of published studies that fail to replicate the original results of the implicit tasks described above: violation-of-expectation[128–131], anticipatory looking[128,132–140], interaction[128,141–144] and altercentric bias[145,146]. Furthermore, a survey revealed that there are additional unpublished studies that also do not find the same results as the original studies[147]. Crucially, although some of these negative results are in conceptual replication studies with minor procedural differences from the original studies (which could account for the different findings), procedural differences are not present in all of the studies. Many studies that fail to replicate the original results — in particular, studies with anticipatory looking measures that can be fully automated — are direct replications with the original stimuli, methods and procedures, but with much larger samples than the original studies[138,139]. These studies include a direct self-replication (with a larger sample) of one of the first and most influential anticipatory looking studies[111] in which the original effect could not be replicated[140].

The validity of the implicit ToM measures has also been challenged by two sets of findings. Regarding construct validity, some original effects can be replicated but disappear once problematic confounds have been removed and suitable controls have been added[138,146]. For example, in a large-scale replication attempt of several anticipatory looking measures, it was found that only two out of many conditions proved replicable[138]. However, it turned out that there were procedural confounds in these two conditions. For example, the correct answer in training and experimental trials was accidentally made identical; as a result, participants may have simply developed a bias over trials to look at a particular side. The effects vanished once the confounds in these conditions were removed. Similar results have been found with other interactive and altercentric interference tasks[145,148,149].

Similarly, implicit tasks suffer from a lack of convergent validity. Decades of research with explicit ToM tasks have shown that tasks that differ dramatically in surface features but share the same meta-representational deep structure, such as various false-belief and other meta-representational tasks (FIG. 3), systematically converge. Proficiency in the different tasks ontogenetically emerges in tandem, and performance on the tasks is highly inter-correlated[43,150]. By contrast, no systematic correlations have been found between the different types of implicit tasks, nor even within different tasks of the same type, all of which are designed to tap the same underlying construct[128–131,138,139].

Status of implicit theory of mind. Returning to the theoretical implications, early results showing implicit ToM in infants were taken as strong evidence for nativism[120]. Similarly, findings with infants and adults that show some dissociations between implicit (type I) and explicit (type II) tasks have been taken as evidence for the general dual-process account[126,127]. However, in light of the discrepancy between positive original and negative replication findings, there is substantial dispute around the existence of implicit meta-representational ToM. One party to the debate has argued that the positive original findings are solid enough to warrant belief in implicit ToM[151]. They suggest that failed replications presenting false negatives might largely be accounted for by procedural differences. Others have expressed more cautious views and argue that it is currently unclear which results are false positives or false negatives and therefore that it is unclear (without more systematic future experimentation) whether implicit ToM exists[152].

One interesting possibility is that there might be two classes of implicit tasks. On the one hand, some tasks strictly require meta-representational ToM such as false-belief tasks and may not be robustly replicable. On the other hand, tasks that can be potentially be solved in simpler ways, for example by level I perspective-taking, may turn out to be replicable. Empirically, some studies have found preliminary evidence for such a pattern[129,153]. Theoretically, this pattern of findings would be highly compatible with a dual-process account. The replicable implicit ToM tasks tap early-developing more basic type I processes, whereas more complex and sophisticated, truly meta-representational type II processes reveal themselves only later in development in explicit tasks.

Summary and future directions

ToM is the capacity for people to understand others and themselves as rational agents with subjective perspectives on the world, who act on the basis of their thoughts, desires, and feelings. At the heart of this capacity — which is fundamental for virtually all forms of human social life — lies meta-representation: representing that and how agents represent the world.

Even though basic forms of ToM are evolutionarily ancient, develop early in human ontogeny, and are shared with non-human primates (BOX 1), fully fledged meta-representational ToM emerges slowly across human childhood. Major changes in ToM capacity occur at around 9 months and 4 years of age. This trajectory is based on developing general neuro-cognitive capacities such as executive function and on linguistic experience. Developmentally, ToM also has characteristic social determinants and wide-ranging social and educational consequences.

There are many crucial aspects of ToM development and its underpinnings that are still poorly understood. For example, there is solid empirical evidence that executive function and language are crucial for acquiring meta-representation. However, the field lacks detailed insight into the underlying proximal mechanisms that permit executive function and language to enable meta-representation. In the case of language, these questions relate to similar questions concerning the interplay of language and thought in other areas. For example, language plays a key part in the acquisition of spatial[86], numerical[84] or modal thought[85] — but it is unclear how. The current debate concerning the role of language in ToM research should therefore be part of a much broader research programme that aims to understand how language shapes thought during development.

Likewise, the field has a solid understanding of the neural underpinnings of adult ToM and studies are

Box 1 | **Theory of mind across species**

Decades of comparative research with non-human primates strongly suggests that fully fledged meta-representational ToM is a uniquely human capacity not shared by our closest living animal relatives: chimpanzees and the other great apes[165,166]. But it remains controversial whether more basic forms of ToM might be evolutionarily more ancient and shared with other species. Numerous studies suggest that basic ToM (perception–goal psychology) is not uniquely human, but shared at least by great apes, perhaps other non-human primates, and maybe by even more remote species such as some birds[165,166]. But even the perception–goal psychology of non-human primates might be limited in comparison to that of human children, who possess the capability from around 1 year of age. Non-human primates seem to use their perception–goal psychology mainly for competition and manipulation, from a detached third-person perspective in which other agents are seen as potential competitors to be manipulated rather than as potential communication or cooperation partners[6]. In contrast to humans, they seem not to enter into the characteristic second-person and first-person plural forms of shared intentionality on the basis of their perception–goal psychology[5].

Regarding fully fledged meta-representational ToM, decades of studies with non-verbal interactive measures have consistently produced negative results. Even though non-human primates master relevant control tasks that show that they can engage in smart interaction and take into account basic social information about the other agents' perspective, they consistently fail when it comes to false-belief and related tasks[165]. Two eye-tracking studies with apes and monkeys found that subjects showed some anticipatory looks that tracked how other agents in an observed scenario would act on the basis of their beliefs. This evidence could suggest that there is some implicit ToM in non-human primates that reveals itself in looking, but not in more active behaviour[160,161]. However, these findings are very difficult to interpret because they used measures borrowed from human infancy research that turned out to be non-replicable[167].

Fully fledged ToM therefore seems to be uniquely human. One reason is that even though humans and other primates share basic perception–goal psychology, the human form also grounds second-person and first-person plural forms of shared intentionality. This shared intentionality provides the basis for cooperation, culture and communication which, in turn, crucially via language and discourse, ground the emergence of fully fledged meta-representation[5,168].

underway to reveal the characteristic neural underpinnings of developmental change. However, it is still unknown how these neural changes underlie cognitive development. Future research needs to aim at developmental implementation theories, which detail how neural changes implement or realize cognitive transitions. New computational theories of the underpinnings of ToM[154,155] might have a crucial role in this endeavour.

Finally, one of the most pressing questions for future study concerns the status of putative implicit ToM. The current complex empirical situation leaves it highly unclear whether implicit ToM tasks are reliable and valid, whether there is such a thing as a separate form of implicit and unconscious ToM, and whether even infants might have a precocious form of meta-representation. The only way forward in this situation is a concerted, collective effort in which researchers join forces to systematically test the reliability and validity of implicit ToM tasks. Fortunately, under the umbrella of ManyBabies[156], a platform for replicability research in infant cognition, such a consortium has recently been formed[157]. This effort includes authors of original and replication studies and proponents from across the theoretical spectrum, in a true adversarial collaboration[158]. In a collaborative spirit, with a priori planned studies, this consortium will test whether implicit ToM tasks are replicable across laboratories. The results will help determine whether there is such a thing as implicit meta-representational ToM, already present in infancy and in operation throughout the lifespan.

Although these questions are mainly theoretical, the answers might also have important practical implications. The more psychologists understand how this fundamental form of human social cognition works and develops, the better equipped they will be to apply this knowledge, for example, by developing educational and intervention programmes that foster development.

Published online: 07 March 2022

1. Perner, J. *Understanding the Representational Mind* (MIT Press, 1991).
2. Wellman, H. M. *Making Minds: How Theory of Mind Develops* (Oxford Univ. Press, 2014).
3. Rakoczy, H. In defense of a developmental dogma: children acquire propositional attitude folk psychology around age 4. *Synthese* **194**, 689–707 (2017).
4. Schilbach, L. et al. Toward a second-person neuroscience. *Behav. Brain Sci.* **36**, 393–414 (2013).
5. Tomasello, M. & Rakoczy, H. What makes human cognition unique? From individual to shared to collective intentionality. *Mind Lang.* **18**, 121–147 (2003).
6. Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. Understanding and sharing intentions: the ontogeny and phylogeny of cultural cognition. *Behav. Brain Sci.* **28**, 675–735 (2005).
7. Perner, J. & Lang, B. Development of theory of mind and executive control. *Trends Cogn. Sci.* **3**, 337–344 (1999).
8. Frankfurt, H. Freedom of the will and the concept of a person. *J. Phil.* **68**, 5–20 (1971).
9. Peterson, C., Slaughter, V., Moore, C. & Wellman, H. M. Peer social skills and theory of mind in children with autism, deafness, or typical development. *Dev. Psychol.* **52**, 46–57 (2016).
10. Peterson, C. C., Slaughter, V. & Wellman, H. M. Nimble negotiators: how theory of mind (ToM) interconnects with persuasion skills in children with and without ToM delay. *Dev. Psychol.* **54**, 494–509 (2018).
11. Slaughter, V., Peterson, C. C. & Moore, C. I can talk you into it: theory of mind and persuasion behavior in young children. *Dev. Psychol.* **49**, 227–231 (2013).
12. Slaughter, V., Imuta, K., Peterson, C. C. & Henry, J. D. Meta-analysis of theory of mind and peer popularity in the preschool and early school years. *Child Dev.* **86**, 1159–1174 (2015).
13. Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B. & Ruffman, T. Theory of mind and prosocial behavior in childhood: a meta-analytic review. *Dev. Psychol.* **52**, 1192–1205 (2016).
14. Lecce, S., Caputi, M., Pagnin, A. & Banerjee, R. Theory of mind and school achievement: the mediating role of social competence. *Cogn. Dev.* **44**, 85–97 (2017).
15. Ding, X. P., Wellman, H. M., Wang, Y., Fu, G. & Lee, K. Theory-of-mind training causes honest young children to lie. *Psychol. Sci.* **26**, 1812–1821 (2015).
16. Talwar, V. & Lee, K. Social and cognitive correlates of children's lying behavior. *Child Dev.* **79**, 866–881 (2008).
17. Gasser, L. & Keller, M. Are the competent the morally good? Perspective taking and moral motivation of children involved in bullying. *Soc. Dev.* **18**, 798–816 (2009).
18. Johnson, M. in *Oxford Handbook of Face Perception* (eds Rhodes, G., Calder, A., Johnson, M. & Haxby, J. V.) 3–14 (Oxford Univ. Press, 2012).
19. Simion, F., Regolin, L. & Bulf, H. A predisposition for biological motion in the newborn baby. *Proc. Natl Acad. Sci. USA* **105**, 809–813 (2008).
20. Rochat, P., Querido, J. G. & Striano, T. Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Dev. Psychol.* **35**, 950–957 (1999).
21. Meltzoff, A. N. & Moore, M. K. Imitation of facial and manual gestures by human neonates. *Science* **198**, 75–78 (1977).
22. Oostenbroek, J. et al. Re-evaluating the neonatal imitation hypothesis. *Dev.Sci.* **22**, e12720 (2019).
23. Oostenbroek, J. et al. Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Curr. Biol.* **26**, 1334–1338 (2016).
24. Davis, J. et al. Does neonatal imitation exist? Insights from a meta-analysis of 336 effect sizes. *Persp. Psychol. Sci.* **16**, 1373–1397 (2021).
25. Tomasello, M. *The Cultural Origins of Human Cognition* (Harvard Univ. Press, 1999).
26. Brooks, R. & Meltzoff, A. The importance of eyes: how infants interpret adult looking behavior. *Dev. Psychol.* **38**, 958–966 (2002).
27. Woodward, A. L. Infants selectively encode the goal object of an actor's reach. *Cognition* **69**, 1–34 (1998).
28. Gergely, G. & Csibra, G. Teleological reasoning in infancy: the naive theory of rational action. *Trends Cogn. Sci.* **7**, 287–292 (2003).
29. Wellman, H. M. in *The Wiley-Blackwell Handbook of Childhood Cognitive Development* 2nd edn (ed. Goswami, U.) 258–284 (Wiley-Blackwell, 2011).
30. Sodian, B., Thoermer, C. & Metz, U. Now I see it but you don't: 14-month-olds can represent another person's visual perspective. *Dev. Sci.* **10**, 199–204 (2007).
31. Heider, F. & Simmel, M. An experimental study of apparent behavior. *Am. J. Psychol.* **57**, 243–259 (1944).
32. Gergely, G. & Csibra, G. Teleological reasoning in infance: the naive theory of rational action. *Trends Cogn. Sci.* **7**, 287–292 (2003).
33. Warneken, F. & Tomasello, M. Altruistic helping in human infants and young chimpanzees. *Science* **311**, 1301–1303 (2006).
34. Gergely, G., Bekkering, H. & Kiraly, I. Rational imitation in preverbal infants. *Nature* **415**, 755 (2002).
35. Carpenter, M., Nagell, K. & Tomasello, M. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* **63**, 176 (1998).
36. Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* **28**, 675–735 (2005).
37. Flavell, J. H., Everett, B. A., Croft, K. & Flavell, E. R. Young children's knowledge about visual perception: further evidence for the level 1–level 2 distinction. *Dev. Psychol.* **17**, 99–103 (1981).
38. Wellman, H. M. & Liu, D. Scaling of theory-of-mind tasks. *Child Dev.* **75**, 523–541 (2004).
39. Wimmer, H. & Perner, J. Beliefs about beliefs — representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* **13**, 103–128 (1983).
40. Gopnik, A. & Astington, J. W. Children's understanding of representational change and its relation to the understanding of false belief and the appearance–reality distinction. *Child Dev.* **59**, 26–37 (1988).
41. Flavell, J. H., Flavell, E. R. & Green, F. L. Development of the appearance–reality distinction. *Cogn. Psychol.* **15**, 95–120 (1983).
42. Rakoczy, H., Fizke, E., Bergfeld, D. & Schwarz, I. Explicit theory of mind is even more unified than previously assumed: belief ascription and understanding aspectuality emerge together in development. *Child Dev.* **86**, 486–502 (2015).
43. Perner, J. & Roessler, J. From infants' to children's appreciation of belief. *Trends Cogn. Sci.* **16**, 519–525 (2012).
44. Lee, J. Y. S. & Imuta, K. Lying and theory of mind: a meta-analysis. *Child Dev.* **92**, 536–553 (2021).
45. Perner, J., Mauer, M. C. & Hildenbrand, M. Identity: key to children's understanding of belief. *Science* **333**, 474–477 (2011).
46. Perner, J., Huemer, M. & Leahy, B. Mental files and belief: a cognitive theory of how children represent belief and its intentionality. *Cognition* **145**, 77–88 (2015).
47. Leekam, S. R., Perner, J. L., Healey, L. & Sewell, C. False signs and the non-specificity of theory of mind: evidence that preschoolers have general difficulties in understanding representations. *Br. J. Dev. Psychol.* **26**, 485–497 (2008).
48. Redshaw, J. Does metarepresentation make human mental time travel unique? *Wiley Interdisc. Rev. Cogn. Sci.* **5**, 519–531 (2014).
49. Redshaw, J. & Suddendorf, T. Temporal junctures in the mind. *Trends Cogn. Sci.* **24**, 52–64 (2020).
50. Perner, J. in *The Self in Time: Developmental Perspectives* (ed Chris, M. & Karen, L.) 268 (Psychology Press, 2001).
51. Hadwin, J. & Perner, J. Pleased and surprised: children's cognitive theory of emotion. *Br. J. Dev. Psychol.* **9**, 215–234 (1991).
52. Apperly, I. in *Understanding Other Minds: Perspectives from Developmental Social Neuroscience* (eds Baron-Cohen, S., Tager-Flusberg, H. & Lombardo, M. V.) 3rd edn, 72–92 (Oxford Univ. Press, 2013).
53. Blakemore, S.-J. The social brain in adolescence. *Nat. Rev. Neurosci.* **9**, 267–277 (2008).
54. Liddle, B. & Nettle, D. Higher-order theory of mind and social competence in school-age children. *J. Cult. Evol. Psychol.* **4**, 231–244 (2006).
55. O'Grady, C., Kliesch, C., Smith, K. & Scott-Phillips, T. C. The ease and extent of recursive mindreading, across implicit and explicit tasks. *Evol. Hum. Behav.* **36**, 313–322 (2015).
56. Filippova, E. & Astington, J. W. Children's understanding of social-cognitive and social-communicative aspects of discourse irony. *Child Dev.* **81**, 913–928 (2010).
57. Happé, F. 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**, 129–154 (1994).

58. Zelazo, P. D. The development of conscious control in childhood. *Trends Cogn. Sci.* **8**, 12–17 (2004).
59. Conway, J. R., Catmur, C. & Bird, G. Understanding individual differences in theory of mind via representation of minds, not mental states. *Psychon. Bull. Rev.* **26**, 798–812 (2019).
60. Henry, J. D., Phillips, L. H., Ruffman, T. & Bailey, P. E. A meta-analytic review of age differences in theory of mind. *Psychol. Aging* **28**, 826–839 (2013).
61. de Villiers, J. G. & de Villiers, P. A. The role of language in theory of mind development. *Top. Lang. Disord.* **34**, 313–328 (2014).
62. Devine, R. T. & Hughes, C. Family correlates of false belief understanding in early childhood: a meta-analysis. *Child Dev.* **89**, 971–987 (2018).
63. Wade, M. et al. On the relation between theory of mind and executive functioning: a developmental cognitive neuroscience perspective. *Psychon. Bull. Rev.* **25**, 2119–2140 (2018).
64. Carlson, S. M., Mandell, D. J. & Williams, L. Executive function and theory of mind: stability and prediction from ages 2 to 3. *Dev. Psychol.* **40**, 1105–1122 (2004).
65. Marcovitch, S. et al. A longitudinal assessment of the relation between executive function and theory of mind at 3, 4, and 5 years. *Cogn. Dev.* **33**, 40–55 (2015).
66. Benson, J. E., Sabbagh, M. A., Carlson, S. M. & Zelazo, P. D. Individual differences in executive functioning predict preschoolers' improvement from theory-of-mind training. *Dev. Psychol.* **49**, 1615–1627 (2013).
67. Devine, R. T. & Hughes, C. Relations between false belief understanding and executive function in early childhood: a meta-analysis. *Child Dev.* **85**, 1777–1794 (2014).
68. Carlson, S. M. & Moses, L. J. Individual differences in inhibitory control and children's theory of mind. *Child Dev.* **72**, 1032–1053 (2001).
69. Sabbagh, M., Moses, L. & Shiverick, S. Executive functioning and preschoolers' understanding of false beliefs, false photographs, and false signs. *Child Dev.* **77**, 1034–1049 (2006).
70. Fizke, E., Barthel, D., Peters, T. & Rakoczy, H. Executive function plays a role in coordinating different perspectives, particularly when one's own perspective is involved. *Cognition* **130**, 315–334 (2014).
71. Rakoczy, H. Executive function and the development of belief–desire psychology. *Dev. Sci.* **13**, 648–661 (2010).
72. Reed, M., Pien, D. L. & Rothbart, M. K. Inhibitory self-control in preschool children. *Merrill Palmer Q.* **30**, 131–147 (1984).
73. Milligan, K., Astington, J. W. & Dack, L. A. Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding. *Child Dev.* **78**, 622–646 (2007).
74. Hofmann, S. G. et al. Training children's theory-of-mind: a meta-analysis of controlled studies. *Cognition* **150**, 200–212 (2016).
75. Lohmann, H. & Tomasello, M. The role of language in the development of false belief understanding: a training study. *Child Dev.* **74**, 1130–1144 (2003).
76. Yu, C.-L., Stanzione, C. M., Wellman, H. M. & Lederberg, A. R. Theory-of-mind development in young deaf children with early hearing provisions. *Psychol. Sci.* **32**, 109–119 (2021).
77. Peterson, C. C. & Siegal, M. Representing inner worlds: theory of mind in autistic, deaf, and normal hearing children. *Psychol. Sci.* **10**, 126–129 (1999).
78. Senghas, A., Kita, S. & Özyürek, A. Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science* **305**, 1779–1782 (2004).
79. Pyers, J. E. & Senghas, A. Language promotes false-belief understanding: evidence from learners of a new sign language. *Psychol. Sci.* **20**, 805–812 (2009).
80. Newton, A. M. & de Villiers, J. G. Thinking while talking. *Psychol. Sci.* **18**, 574–579 (2007).
81. Dungan, J. & Saxe, R. Matched false-belief performance during verbal and nonverbal interference. *Cogn. Sci.* **36**, 1148–1156 (2012).
82. deVilliers, J. G. & deVilliers, P. A. in *Children's Reasoning and the Mind* (eds Mitchell, P. & Riggs, K. J.) 191–228 (Psychology Press/Taylor & Francis, 2000).
83. Bermudez, J. *Thinking without Words* (Oxford Univ. Press, 2003).
84. Carey, S. *The Origin of Concepts* (Oxford Univ. Press, 2009).
85. Leahy, B. P. & Carey, S. E. The acquisition of modal concepts. *Trends Cogn. Sci.* **24**, 65–78 (2020).
86. Spelke, E. in *Advances in the Study of Language and Thought* (eds Gentner, D. & Goldin-Meadow, S.) 277–311 (MIT Press, 2003).
87. Schurz, M., Radua, J., Aichhorn, M., Richlan, F. & Perner, J. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci. Biobehav. Rev.* **42**, 9–34 (2014).
88. Schurz, M. et al. Toward a hierarchical model of social cognition: a neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychol. Bull.* **147**, 293–327 (2021).
89. Saxe, R. in *Encyclopedia of Consciousness* (ed. Banks, W. P.) 401–409 (Academic Press, 2009).
90. Frith, C. D. & Frith, U. The neural basis of mentalizing. *Neuron* **50**, 531–534 (2006).
91. Van Overwalle, F. Social cognition and the brain: a meta-analysis. *Hum. Brain Mapp.* **30**, 829–858 (2009).
92. Liu, D., Sabbagh, M. A., Gehring, W. J. & Wellman, H. M. Neural correlates of children's theory of mind development. *Child Dev.* **80**, 318–326 (2009).
93. Jamali, M. et al. Single-neuronal predictions of others' beliefs in humans. *Nature* **591**, 610–614 (2021).
94. Xiao, Y., Geng, F., Riggins, T., Chen, G. & Redcay, E. Neural correlates of developing theory of mind competence in early childhood. *NeuroImage* **184**, 707–716 (2019).
95. Wiesmann, C. G., Friederici, A. D., Singer, T. & Steinbeis, N. Two systems for thinking about others' thoughts in the developing brain. *Proc. Natl Acad. Sci. USA* **117**, 6928–6935 (2020).
96. Grosse Wiesmann, C., Schreiber, J., Singer, T., Steinbeis, N. & Friederici, A. D. White matter maturation is associated with the emergence of theory of mind in early childhood. *Nat. Commun.* **8**, 14692 (2017).
97. Hughes, C. & Devine, R. T. Individual differences in theory of mind from preschool to adolescence: achievements and directions. *Child Dev. Persp.* **9**, 149–153 (2015).
98. Hughes, C. et al. Origins of individual differences in theory of mind: from nature to nurture? *Child Dev.* **76**, 356–370 (2005).
99. Ebert, S., Peterson, C., Slaughter, V. & Weinert, S. Links among parents' mental state language, family socioeconomic status, and preschoolers' theory of mind development. *Cogn. Dev.* **44**, 32–48 (2017).
100. McAlister, A. R. & Peterson, C. C. Siblings, theory of mind, and executive functioning in children aged 3–6 years: new longitudinal evidence. *Child Dev.* **84**, 1442–1458 (2013).
101. Dunn, J. et al. Young children's understanding of other people's feelings and beliefs: individual differences and their antecedents. *Child Dev.* **62**, 1352–1366 (1991).
102. Slaughter, V. & Peterson, C. C. in *Access to Language and Cognitive Development* (eds Siegal, M. & Surian, L.) 3–22 (Oxford Univ. Press, 2012).
103. Aldrich, N. J., Chen, J. & Alfieri, L. Evaluating associations between parental mind-mindedness and children's developmental capacities through meta-analysis. *Dev. Rev.* **60**, 100946 (2021).
104. Meins, E. et al. Maternal mind-mindedness and attachment security as predictors of theory of mind understanding. *Child Dev.* **73**, 1715–1726 (2002).
105. Callaghan, T. et al. Synchrony in the onset of mental-state reasoning: evidence from five cultures. *Psychol. Sci.* **16**, 378–384 (2005).
106. Bradford, E. E. et al. Cross-cultural differences in adult theory of mind abilities: a comparison of native-English speakers and native-Chinese speakers on the self/other differentiation task. *Q. J. Exp. Psychol.* **71**, 2665–2676 (2018).
107. Shahaeian, A., Peterson, C. C., Slaughter, V. & Wellman, H. M. Culture and the sequence of steps in theory of mind development. *Dev. Psychol.* **47**, 1239–1247 (2011).
108. Taumoepeau, M., Sadeghi, S. & Nobilo, A. Cross-cultural differences in children's theory of mind in Iran and New Zealand: the role of caregiver mental state talk. *Cogn. Dev.* **51**, 32–45 (2019).
109. Bloom, P. & German, T. P. Two reasons to abandon the false belief task as a test of theory of mind. *Cognition* **77**, B25–B31 (2000).
110. Onishi, K. H. & Baillargeon, R. Do 15-month-old infants understand false beliefs? *Science* **308**, 255–258 (2005).
111. Southgate, V., Senju, A. & Csibra, G. Action anticipation through attribution of false belief by 2-year-olds. *Psychol. Sci.* **18**, 587–592 (2007).
112. Schneider, D., Bayliss, A. P., Becker, S. I. & Dux, P. E. Eye movements reveal sustained implicit processing of others' mental states. *J. Exp. Psychol.* **141**, 433–438 (2012).
113. Surian, L. & Geraci, A. Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *Br. J. Dev. Psychol.* **30**, 30–44 (2012).
114. Buttelmann, D., Carpenter, M. & Tomasello, M. Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition* **112**, 337–342 (2009).
115. Southgate, V., Chevallier, C. & Csibra, G. Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Dev. Sci.* **13**, 907–912 (2010).
116. Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J. & Bodley Scott, S. E. Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J. Exp. Psychol.* **36**, 1255–1266 (2010).
117. Kovács, Á. M., Téglás, E. & Endress, A. D. The social sense: susceptibility to others' beliefs in human infants and adults. *Science* **330**, 1830–1834 (2010).
118. Scott, R. M. & Baillargeon, R. Early false-belief understanding. *Trends Cogn. Sci.* **21**, 237–249 (2017).
119. Baillargeon, R., Scott, R. M. & He, Z. False-belief understanding in infants. *Trends Cogn. Sci.* **14**, 110–118 (2010).
120. Carruthers, P. Mindreading in infancy. *Mind Lang.* **28**, 141–172 (2013).
121. Leslie, A. M. Developmental parallels in understanding minds and bodies. *Trends Cogn. Sci.* **9**, 459–462 (2005).
122. Apperly, I. A. & Butterfill, S. A. Do humans have two systems to track beliefs and belief-like states? *Psychol. Rev.* **116**, 953–970 (2009).
123. Frith, C. D. & Frith, U. Implicit and explicit processes in social cognition. *Neuron* **60**, 503–510 (2008).
124. Low, J., Apperly, I. A., Butterfill, S. A. & Rakoczy, H. Cognitive architecture of belief reasoning in children and adults: a primer on the two-systems account. *Child Dev. Persp.* **10**, 184–189 (2016).
125. Butterfill, S. A. & Apperly, I. A. How to construct a minimal theory of mind. *Mind Lang.* **28**, 606–637 (2013).
126. Edwards, K. & Low, J. Reaction time profiles of adults' action prediction reveal two mindreading systems. *Cognition* **160**, 1–16 (2017).
127. Low, J. & Watts, J. Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychol. Sci.* **24**, 305–311 (2013).
128. Dörrenberg, S., Rakoczy, H. & Liszkowski, U. How (not) to measure infant theory of mind: testing the replicability and validity of four non-verbal measures. *Cogn. Dev.* **46**, 12–30 (2018).
129. Powell, L. J., Hobbs, K., Bardis, A., Carey, S. & Saxe, R. Replications of implicit theory of mind tasks with varying representational demands. *Cogn. Dev.* **46**, 40–50 (2018).
130. Poulin-Dubois, D. & Yott, J. Probing the depth of infants' theory of mind: disunity in performance across paradigms. *Dev. Sci.* **21**, e12600 (2017).
131. Yott, J. & Poulin-Dubois, D. Are infants' theory-of-mind abilities well integrated? Implicit understanding of intentions, desires, and beliefs. *J. Cogn. Dev.* **17**, 683–698 (2016).
132. Burnside, K., Ruel, A., Azar, N. & Poulin-Dubois, D. Implicit false belief across the lifespan: non-replication of an anticipatory looking task. *Cogn. Dev.* **46**, 4–11 (2018).
133. Grosse Wiesmann, C., Friederici, A. D., Disla, D., Steinbeis, N. & Singer, T. Longitudinal evidence for 4-year-olds' but not 2- and 3-year-olds' false belief-related action anticipation. *Cogn. Dev.* **46**, 58–68 (2018).
134. Schuwerk, T., Priewasser, B., Sodian, B. & Perner, J. The robustness and generalizability of findings on spontaneous false belief sensitivity: a replication attempt. *R. Soc. Open. Sci.* **5**, 172273 (2018).
135. Kulke, L. & Rakoczy, H. Testing the role of verbal narration in implicit theory of mind tasks. *J. Cogn. Dev.* **20**, 1–14 (2019).
136. Kulke, L., Johannsen, J. & Rakoczy, H. Why can some implicit theory of mind tasks be replicated and others cannot? A test of mentalizing versus submentalizing accounts. *PLoS ONE* **14**, e0213772 (2019).
137. Kulke, L., Wübker, M. & Rakoczy, H. Is implicit theory of mind real but hard to detect? Testing adults with different stimulus materials. *R. Soc. Open. Sci.* **6**, 190068 (2019).

138. Kulke, L., von Duhn, B., Schneider, D. & Rakoczy, H. Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. *Psychol. Sci.* **29**, 888–900 (2018).
139. Kulke, L., Reiβ, M., Krist, H. & Rakoczy, H. How robust are anticipatory looking measures of theory of mind? Replication attempts across the life span. *Cogn. Dev.* **46**, 97–111 (2018).
140. Kampis, D., Kármán, P., Csibra, G., Southgate, V. & Hernik, M. A two-lab direct replication attempt of Southgate, Senju and Csibra (2007). *R. Soc. Open. Sci.* **8**, 210190 (2021).
141. Crivello, C. & Poulin-Dubois, D. Infants' false belief understanding: a non-replication of the helping task. *Cogn. Dev.* **46**, 51–57 (2018).
142. Wenzel, L., Dörrenberg, S., Proft, M., Liszkowski, U. & Rakoczy, H. Actions do not speak louder than words in an interactive false belief task. *R. Soc. Open Sci.* **7**, 191998 (2020).
143. Dörrenberg, S., Wenzel, L., Proft, M., Rakoczy, H. & Liszkowski, U. Reliability and generalizability of an acted-out false belief task in 3-year-olds. *Infant. Behav. Dev.* **54**, 13–21 (2019).
144. Grosse Wiesmann, C., Friederici, A. D., Singer, T. & Steinbeis, N. Implicit and explicit false belief development in preschool children. *Dev. Sci.* **20**, e12445 (2017).
145. Conway, J. R., Lee, D., Ojaghi, M., Catmur, C. & Bird, G. Submentalizing or mentalizing in a level 1 perspective-taking task: a cloak and goggles test. *J. Exp. Psychol.* **43**, 454–465 (2017).
146. Cole, G. G., Atkinson, M. A., D'Souza, A. D. C. & Smith, D. T. Spontaneous perspective taking in humans? *Vision* **1**, 17 (2017).
147. Kulke, L. & Rakoczy, H. Implicit theory of mind — an overview of current replications and non-replications. *Data Brief.* **16**, 101–104 (2018).
148. Priewasser, B., Rafetseder, E., Gargitter, C. & Perner, J. Helping as an early indicator of a theory of mind: mentalism or teleology? *Cogn. Dev.* **46**, 69–78 (2018).
149. Phillips, J. et al. A second look at automatic theory of mind. *Psychol. Sci.* **26**, 1353–1367 (2015).
150. Wellman, H., Cross, D. & Watson, J. Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev.* **72**, 655–684 (2001).
151. Baillargeon, R., Buttelmann, D. & Southgate, V. Invited Commentary: Interpreting failed replications of early false-belief findings: methodological and theoretical considerations. *Cogn. Dev.* **46**, 112–124 (2018).
152. Poulin-Dubois, D. et al. Do infants understand false beliefs? We don't know yet — a commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cogn. Dev.* **48**, 302–315 (2018).
153. Holland, C. & Phillips, J. S. A theoretically driven meta-analysis of implicit theory of mind studies: the role of factivity. *Proc. 42nd Annu. Conf. Cognitive Science Society* (2020).
154. Baker, C. L., Jara-Ettinger, J., Saxe, R. & Tenenbaum, J. B. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nat. Hum. Behav.* **1**, 0064 (2017).
155. Jara-Ettinger, J. Theory of mind as inverse reinforcement learning. *Curr. Opin. Behav. Sci.* **29**, 105–110 (2019).
156. Frank, M. C. et al. A collaborative approach to infant research: promoting reproducibility, best practices, and theory-building. *Infancy* **22**, 421–435 (2017).
157. Schuwerk, T., et al. (2022). Action anticipation based on an agent's epistemic state in toddlers and adults. Child Development, (In-Principle Acceptance of Registered Report Stage 1: Study Design).
158. Mellers, B., Hertwig, R. & Kahneman, D. Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychol. Sci.* **12**, 269–275 (2001).
159. Yott, J. & Poulin-Dubois, D. Breaking the rules: do infants have a true understanding of false belief? *Br. J. Dev. Psychol.* **30**, 156–171 (2012).
160. Krupenye, C., Kano, F., Hirata, S., Call, J. & Tomasello, M. Great apes anticipate that other individuals will act according to false beliefs. *Science* **354**, 110–114 (2016).
161. Hayashi, T. et al. Macaques exhibit implicit gaze bias anticipating others' false-belief-driven actions via medial prefrontal cortex. *Cell Rep.* **30**, 4433–4444. e5 (2020).
162. Király, I., Oláh, K., Csibra, G. & Kovács, Á. M. Retrospective attribution of false beliefs in 3-year-old children. *Proc. Natl Acad. Sci. USA* **115**, 11477–11482 (2018).
163. Michael, J. et al. Seeing it both ways: using a double-cuing task to investigate the role of spatial cuing in level-1 visual perspective-taking. *J. Exp. Psychol.* **44**, 693–702 (2018).
164. Surtees, A. D. R., Butterfill, S. A. & Apperly, I. A. Direct and indirect measures of level-2 perspective-taking in children and adults. *Br. J. Dev. Psychol.* **30**, 75–86 (2012).
165. Call, J. & Tomasello, M. Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn. Sci.* **12**, 187–192 (2008).
166. Martin, A. & Santos, L. R. What cognitive representations support primate theory of mind? *Trends Cogn. Sci.* **20**, 375–382 (2016).
167. Horschler, D. J., MacLean, E. L. & Santos, L. R. Do non-human primates really represent others' beliefs? *Trends Cogn. Sci.* **24**, 594–605 (2020).
168. Tomasello, M. How children come to understand false beliefs: a shared intentionality account. *Proc. Natl Acad. Sci. USA* **115**, 8491–8498 (2018).

Acknowledgements
The author thanks Natalie Bleijlevens, Isa Garbisch, Feride Nur Haskaraca Kızılay, Lydia Schidelko, Britta Schünemann and Rowan Titchener for helpful comments on a previous version of this paper.

Competing interests
The author declares no competing interests.

Peer review information
Nature Reviews Psychology thanks Martin Doherty, Virginia Slaughter, and the other, anonymous, reviewer for their contribution to the peer review of this work.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.