26

THE DEVELOPMENT OF IMPLICIT THEORY OF MIND

Hannes Rakoczy

The most fundamental aim of developmental cognitive science is to describe and explain trajectories of cognitive ontogeny. What is the starting state with respect to the conceptual repertoire and abilities of a subject? What is the mature state? What are intermediate stages on the way from the initial to the mature state? And what are mechanisms of transition?

A recurring theme in many accounts of developmental cognitive science is the transition from earlier implicit to later explicit forms of cognition. In a given domain, how does development progress from implicit forms of representing matters in that domain to explicit ones? In this chapter, we will focus on Theory of Mind as a case study. At the outset, we start with a relatively vague and pre-theoretic notion of the implicit-explicit distinction – roughly to the effect that explicit capacities are those that figure in flexible ways in inference, rational action planning and linguistic expression, and implicit ones are those that fall short of these characteristics in some way or other. Provisionally, we accept the premises that the implicit-explicit distinction is a unitary and sharp one, and that there is a tight correspondence between capacities and tasks such that there are explicit tasks that tap explicit capacities, and implicit tasks that tap implicit capacities. In the course of the chapter, the pre-theoretic notion will then be sharpened, and more nuanced distinctions will be introduced and developed. The simplified assumption of a 1:1 correspondence of competence and task will finally be discussed and questioned.

The structure of the chapter is as follows: Section 1 gives an overview of the current state of research on Theory of Mind development, reviewing findings from the last four decades on explicit Theory of Mind, and more recent work on earlier capacities from implicit tasks. Section 2 focuses on the main conceptual question: If there was solid evidence for early competence from implicit tasks, what would this show with regard to the underlying cognitive processes and their development? How would the capacities tapped in implicit tasks relate – in terms of cognitive architecture and development – to those tapped in explicit tasks? Section 3 focuses on the main empirical question: Is there such solid evidence? Original findings as well as a more recent replication crisis are reviewed. Section 4, finally, discusses future directions for investigating implicit and explicit Theory of Mind in more fine-grained ways.

DOI: 10.4324/9781003014584-33

1. Implicit and Explicit Theory of Mind: The Current State of Affairs

Theory of Mind (ToM) is the conceptual framework with which we describe and understand others and ourselves as rational agents with a subjective view on the world. At the heart of ToM lies the ascription of propositional attitudes like beliefs and desires and thus a form of meta-representation (Dennett 1978; Perner 1991): We meta-represent how agents subjectively represent the world as it appears to them (beliefs) and as they want it to be (desires); and we predict and explain their rational actions accordingly (typically, rational agents perform actions that, according to their beliefs, would further their ends). Propositional attitudes are subjective in several respects: Different agents can have representational access to different objects and situations (Adam believes that he has apples in his bag, Eve believes that she has pears in hers); agents can misrepresent a given situation (Adam believes the fruit on the tree are apples whereas in fact they are pears); and agents represent a given situation always from some perspectives, under some aspects, and not under others so that different agents may view the same situation in different ways (Adam may believe that he has an apple tree in front of him while failing to represent it as the tree of knowledge; Eve may believe about the same tree that it is the tree of knowledge without representing it as an apple tree . . .).

1.1 The Development of Explicit Theory of Mind: The Standard Picture

When and how does meta-representational ToM develop? Decades of research that addressed this question with explicit tasks have yielded a clear and consistent picture: The capacity for meta-representation and for understanding the subjectivity of propositional attitudes emerges in protracted ways over the preschool years. Even in the first two years of life, children ascribe simple mental states to others and track, for example, what they perceive or intend. Young children are thus sometimes said to operate with a "perception-goal folk psychology". But this early conceptual framework is fundamentally limited in its cognitive sophistication. It allows children to understand only very basic forms of subjectivity (different agents may perceive or aim at different things), but falls short of a truly meta-representational understanding of subjective, aspectual, potentially inaccurate representations. These limitations are overcome in the course of a crucial conceptual transition around age four (Wellman et al. 2001). From around this age, children begin to master what has become the litmus test for operating with a fully-fledged "belief-desire psychology" and understanding misrepresentation: so-called False Belief (FB) tasks (Wimmer and Perner 1983). In FB tasks, a story protagonist holds a belief that turns out to be false (e.g., she puts an object in box 1, which is then transferred in her absence to box 2), and the child is explicitly asked to predict what the protagonist will do (e.g., where she will look for the object). Young children typically answer incorrectly on the basis of the actual state of affairs (e.g., protagonist will go to where the object really is) whereas children from age four typically answer correctly that the agent will act on the basis of her subjective belief, irrespective of whether it is true or false (e.g., go to the location where she believes the object to be even if it is not there). The very same developmental patterns emerge in superficially very diverse tasks that share a conceptual deep structure in that they all tap meta-representation. Furthermore, performance across these tasks is highly consistent and correlated: Competence emerges in a systematic package such that children tend to become competent at solving different tasks at the same time (Perner and Roessler 2012). Finally, this emerging competence is closely linked to the development of domain-general cognitive and linguistic capacities.

The standard interpretation of this rich body of empirical evidence has been that the four-year transition marks a kind of conceptual revolution (Gopnik and Astington 1988; Perner 1991; Rakoczy 2017). Children at this point acquire the conceptual apparatus for meta-representing that agents represent that something is the case; paradigmatically, they come to form meta-representational beliefs that another agent believes that p. Even though meta-representation is much wider and includes, for example, desires about desires, beliefs about desires etc., in the present chapter we will follow standard practice in the field and focus on beliefs about beliefs as the paradigmatic form of meta-representational ToM.

1.2 Challenges to the Standard Picture

New research in the last 15–20 years has challenged this standard interpretation. This research has used nonverbal, implicit measures that do not require subjects to answer questions or engage in high-level action planning. Rather it has tapped looking behaviour, neural signatures, priming or other forms of more or less spontaneous behaviour as dependent measures (for review, see Baillargeon et al. 2016; Baillargeon et al. 2010; Scott and Baillargeon 2017).

In *violation of expectation* (VoE) studies, children first see sequences of events like those in explicit tasks. For example, a protagonist puts an object O in box 1, which is then transferred to box 2 in the presence (true belief (TB) condition) or absence of the protagonist (FB condition). In the test phase, they then see how the protagonist acts either consistently with her beliefs (searches for the object in box 1 in FB/ box 2 in TB) or inconsistently (searches for the object in box 2 in FB/ box 1 in TB). Results of several experiments since the seminal Onishi and Baillargeon (2005) study suggest that children much younger than four (even infants) look longer to, and thus seem surprised by, belief-inconsistent events. In particular, they look longer when the protagonist acts inconsistently with her false belief even if that means that she searches for the object where it *really* is.

Results from *spontaneous interaction* studies suggest that children from one to two years respond differently and appropriately to an interaction partner as a function of her true or false belief (e.g., D. Buttelmann et al. 2009; Knudsen and Liszkowski 2012; Southgate et al. 2010). For example, in one set of studies the protagonist put object A in box 1, and object B in box 2. The objects were then swapped in her presence (TB) or absence (FB), and the protagonist pointed to box 1 and asked the child, ambiguously, "Can you give it to me?" (Southgate et al. 2010). Children tended to give her the object from box 1 in the TB, but the object from box 2 in the FB condition – which indicates that they took into account the agent's belief in order to disambiguate what she meant.

The broadest body of evidence comes from *anticipatory looking* (AL) tasks. In such tasks, like in VoE studies, subjects see standard scenarios in which a protagonist forms a true/false belief (e.g., as to whether an object is in box 1 or box 2). Rather than tapping post-hoc looking at (in-)consistent outcomes, a subject's direction of gaze (towards box 1/box 2) is recorded as an indicator of expectations regarding what the protagonist will do (where she will go to search). These tasks, in contrast to VoE and interaction tasks, are suitable for, and have been used with, a wide variety of populations, across the lifespan, across species and across typical versus atypical conditions. Regarding lifespan development, results from several studies suggest that infants, preschoolers, older children and adults all show belief-consistent AL patterns (e.g., Schneider et al. 2012; Southgate et al. 2007; Surian and Geraci 2012). Two recent comparative studies suggest that perhaps apes (Krupenye et al. 2016) and even monkeys (Hayashi et al. 2020) show analogous AL patterns. And clinical studies suggest that children and adults with Asperger

Syndrome do not spontaneously engage in the same kind of AL patterns (e.g., Schuwerk et al. 2016; Senju et al. 2009).

Complementary findings come from studies indicating that young children, before they can solve explicit FB tasks, show different peripheral physiological responses (e.g., pupil dilation) or different neurophysiological signatures to events in which agents act in belief-incongruent as compared to belief-congruent ways (Hyde et al. 2018; Southgate and Vernetti 2014).

2. What Would Solid Evidence From Implicit Theory of Mind Tasks Really Show?

Now, let us assume for the moment that these findings are robust, reliable and replicable (whether this is really the case will be the topic of section 3). If so, what would they mean? What kind of underlying cognitive processes and capacities would they be evidence for?

Minimally, what such findings would show is that infants (or adults[1]) are representationally sensitive to events that involve agents forming (true or false) beliefs – as indicated in their looking and related responses. Something in the infant, we could say, clearly represents something about a situation in which an agent has a belief. But does that amount to the infant holding a meta-representational belief that this agent believes that p? This question actually breaks down into two sub-questions: First, does the infant hold meta-representational beliefs (rather than some simpler kind of representational states) about the agent? Second, does the infant ascribe to the agent beliefs (rather than some simpler kind of states)?

These two sub-questions help to map the theoretical territory nicely. Strong nativist early competence accounts answer emphatically "yes" to both sub-questions (e.g., Carruthers 2013; Leslie 2005); skeptical sub-*mentalizing* and related accounts answer "not at all" to both (e.g., Heyes 2014); and intermediate conceptual change and two-systems accounts answer "not quite" to both (e.g., Apperly and Butterfill 2009; Perner 1991). While the radical responses (2x "yes" vs. 2x "no") of nativist vs. skeptical accounts seem relatively straightforward, the more nuanced and intermediate 2x "not quite"-responses of conceptual change and two-systems views appear more in need of explanation. Conceptual change accounts assume that children's ToM develops gradually. Basic mental state concepts (e.g., for perception and goal-directed action in the form of a "perception-goal folk psychology") develop first, with subsequent refinement and acquisition of more sophisticated concepts (e.g., for subjective beliefs and desires) (Gopnik and Astington 1988; Perner 1991). Relatedly, two-systems accounts assume that ToM is not necessarily a unitary capacity that develops uniformly. Rather, basic ToM processes may be phylogenetically more ancient and develop earlier ontogenetically; these processes may continue to operate rapidly and more or less automatically, with little need for central cognitive resources, throughout the lifespan. Fully-fledged ToM processes, in contrast, may ontogenetically develop in more protracted ways, on the basis of linguistic experience and central cognitive resources. While the fully-fledged ToM processed do involve the meta-representational ascription of beliefs and other propositional attitudes to other agents, the more basic processes may only involve simpler forms of keeping track of simpler forms of mental states of other agents (Apperly and Butterfill 2009). Both conceptual change and two-systems accounts would thus take intermediate positions regarding the answers to the two types of questions: Implicit tasks may indicate that infants form some kinds of representations of others' representational states – but that does not necessarily mean that they hold fully-fledged meta-representational beliefs about others' fully-fledged beliefs. In the following, we will address the two sub-questions in more detail in turn.

2.1 Would Solid Findings From Implicit Tasks Show That Infants Form Meta-representational Beliefs?

Infants' looking behaviour in the studies described earlier, if solid, would clearly show that the infant, or something in the infant, representationally tracks something about the agent's epistemic condition – to put it as neutrally as possible. But would it show, less neutrally, that the infant forms a meta-representational belief . . . (. . . about the agent's belief)?

The standard nativist response is strongly affirmative and goes like this (Baillargeon et al. 2010; Carruthers 2013; Leslie 2005): Looking time and related "implicit" tasks tap the core competence for ToM that is more or less innate and realized in a more or less modular architecture (perhaps in a "Theory of Mind Module"). The competence involves the use of meta-representation, and generally the very same conceptual resources as those used later in life in explicit tasks. That "implicit" and "explicit" tasks diverge so massively, with success in the latter lagging years behind the former, does not mean that the two types of tasks tap into different kinds of conceptual competence. Rather, both types of tasks tap the very same conceptual competence – meta-representation – but the "explicit" tasks are artificially difficult because of additional extraneous task demands. In addition to ToM, these tasks require sophisticated linguistic, executive and other competencies. Young children's failure in standard explicit tasks before age four does not indicate any form of ToM competence deficit, but merely performance limitations caused by the extraneous task demands. In terms of conceptual development, there thus need not be any form of fundamental conceptual change. Even infants form meta-representational beliefs. What develops is nothing about the meta-representational competence itself, but merely how this competence gets integrated with other processes and thus, as a consequence, translates into performance in verbally or otherwise more taxing tasks. This becomes particularly clear when considering the relation of the core meta-representational ToM competence and executive function. A large body of evidence from the last two decades has documented intimate relations between ToM performance in standard explicit tasks and executive function ((Carlson and Moses 2001; Devine and Hughes 2014). The nativist interpretation of this relation comes as a pure "*expression account*": Executive function develops substantially over the preschool years and is crucial for "expressing" the ToM core competence – for translating it into performance – in tasks that are taxing in terms of inhibition and the like. Standard explicit FB tests are paradigms of such tasks: In predicting what the mistaken agent will do on the basis of her subjective belief, one's own perspective, and what is taken to be objective truth (the two usually coincide, of course) have to be put aside. According to nativist expression accounts, early ToM capacities are thus implicit in the following sense: The infant operates with representational states of exactly the same type and with exactly the same conceptual content as later explicit states; it just so happens that this content cannot yet be expressed verbally or put to use in complex tasks.

Such *expression accounts* contrast with *emergence accounts* according to which executive function is not just an extraneous add-on that helps to express the core meta-representational competence that is already in place, but part of the very meta-representational competence itself that emerges over time. Similar points hold regarding the well-established relation between ToM (tapped in explicit tasks) and language: Nativist accounts merely assign an expressive role to language, whereas other accounts consider language to have a more substantial role in the emergence and constitution of meta-representational competence itself.

Nativist accounts thus move from the premise "infants are representationally sensitive to belief-involving situations" to the conclusion "infants form meta-representational beliefs". Alternative accounts such as conceptual change and dual-process views would point out in response that this move involves a potential inferential gap. What it neglects is that the logical

geography is much more complex (see also Burge 2010, 2018). Yes, robust findings from implicit tasks would show some representational sensitivity in infants (and adults). But there are many ways of being representationally sensitive without forming fully-fledged propositional attitudes (in the present case, meta-representational beliefs). A starting intuition is the following: It is one thing for something to be information in a system; yet another for it to be information to the system (e.g., Clark and Karmiloff-Smith 1993; Cummins 1983). There have been many attempts to explicate this intuition, most of them closely related to one version or another of implicit-explicit distinctions. In the following, I will focus on two of them. The first distinction is between propositional attitudes (beliefs, in particular) and more basic subdoxastic states (Davies 1989; Stich 1978). Cognitive science postulates many types of representational states, but most of them are quite different from the representational states we ascribe at the personal level in our folk psychology. Take, for example, states that represent time in us in some way. Many representational states involved in speech perception track time, for instance in the form of "voice onset time" (VOT) that distinguishes different phonemes from each other. Now, both the state in me that tracks a VOT of, say, 80 ms, and my personal level belief "80 ms is a remarkably short time interval" are states that are about, that represent time in some sense. But they differ in fundamental and crucial ways: Beliefs, in contrast to *subdoxastic states*, are conceptualized, inferentially integrated and potentially conscious.

A second, closely related distinction centers around the *accessibility* of different types of representational states (Block 1995):[2] A given state is *access conscious* if its content is inferentially promiscuous (i.e., potentially available and integrated in open-ended flexible ways typical of conceptual thought), and available for the rational control of action planning and (typically) for the rational control of language. Again, my occurrent belief that "80 ms is a remarkably short time interval" is inferentially integrated and flexibly available for thought and action, and thus access-conscious, in ways that my speech perceptual tracking of 80 ms VOT in principle is not.

From a functionalist point of view, simpler (subdoxastic, non-access conscious) representational states differ from fully-fledged propositional attitudes in their much more restricted functional profiles. That a belief is inferentially promiscuous and available in flexible and open-ended ways for reasoning, planning and action is not an accidental, extraneous fact about it – beliefs are not subdoxastic states that merely happen to be expressible in more varied circumstances. Rather, it is an essential feature of what makes it a belief. From this functionalist perspective, we can make clearer sense of two intuitions: that there is some kind of continuity in the development from simple to more complex states while at the same time there are deep qualitative differences between them. According to functionalism, a given state (in humans, typically neural in nature) realizes or implements a given mental state if it realizes the corresponding functional profile. Now, assume a given (neural) state of kind N finally, after some developmental history, comes to realize the functional profile of mental state of kind M (say, beliefs). Then, in the mature state of the system, N is the *core realizer* of M – given its functional relations to other elements of the system (the former and the latter together are the total realizers of M; (Shoemaker 1981). For the description of cognitive development, this means the following: There possibly is cognitive continuity in the sense that a given state N is in place early in development that eventually will turn out to be the core realizer of M. But in the early stages the right kinds of functional relations have not been established yet to make N realize M. N only comes to be the core realizer of M over time, once the right kinds of functional relations get established.

Turning to the findings from implicit ToM tasks with these distinctions at hand, the following picture emerges: Results from these tasks would indicate simple (subdoxastic) representational states in the infant that track belief-involving situations. But as long as these representational states merely reveal themselves in the guidance of looking behavior, without any evidence that

they embody anything like the functional profile of beliefs (conceptualization, inferential integration, availability for reasoning and rational action control),[3] there is no reason to assume the states in question are the very same kinds of states tapped later in explicit verbal tasks. Rather, the following alternative picture seems more accurate: Over development, there is substantial conceptual change. The cognitive trajectory goes from subdoxastic representations of others' epistemic situation in infancy to fully-fledged meta-representation from around age four.

This also has consequences for describing neurocognitive development in this domain. That a given neural structure that is involved in explicit ToM reasoning later in life (for example, the temporoparietal junction) is already involved in infants' responses to belief-involving events in looking times (e.g., Hyde et al. 2018), does not mean that the very same cognitive processes (meta-representational judgments) are in operation – in contrast to some nativist interpretations of such findings (e.g., Scott, Roby and Baillargeon in press). Rather, this may simply mean that what will turn out to be a core realizer of true meta-representation later in ontogeny is present earlier but still lacks the right kind of functional connections that would turn it into a (meta-representational) belief. What happens over development neuro-cognitively, then, is not so much local change in this or that area, but functional connection between a given area and others (in particular, those related to language, executive function and more central cognition in general) so that the core area plus its functional connections come to constitute the total realizer of the capacity in question. More recent findings on the neurocognitive changes underlying the transition to explicit ToM are highly compatible with such a picture (Grosse Wiesmann et al. 2017).

According to this alternative picture of conceptual change and two-systems-views, early ToM capacities are thus implicit in quite a different sense from that envisaged by nativism. They are not the same capacities as the later ones, merely limited in their expressibility. Rather, they are qualitatively different, subdoxastic states rather than fully-fledged beliefs (see Frankish 2009).

2.2 Would Solid Findings From Implicit Tasks Show That Infants Ascribe Beliefs to Agents?

The looking patterns in violation-of-expectation and anticipatory looking FB tasks, if solid, would minimally indicate that infants (and adults) are sensitive to situations in which an agent acts according to her true/false belief. The nativist early competence interpretation of the data goes further: What the looking data suggest, according to nativism, is that infants are not just sensitive to belief-involving situations, but represent beliefs as such. They operate with a concept of "belief" (and other propositional attitudes); in fact, with the very same concept of belief that older children and adults have. It is just that the infants cannot use this concept as freely and flexibly yet.

In contrast, conceptual change and two-systems accounts claim that this nativist line of argumentation, again, involves a potential inferential gap. It fails to do justice to the fact that there are many ways of keeping track of belief-involving situations without using a full-blown concept of "belief" and thus without ascribing beliefs proper to agents. The notion of "belief" is a very complex one[4] that may simply be beyond the conceptual reach of infants. But simpler proxies can do many of the jobs of belief ascription in more primitive ways. One particular suggestion along these lines goes as follows (Apperly and Butterfill 2009; Butterfill and Apperly 2013): Infants may simply keep track of who registered which kind of information. This can get you quite far. You can engage in level-I perspective-taking (understanding who has seen what; Flavell et al. 1981); understand who has and who hasn't got knowledge (in the sense of

information access) about some states of affairs (Phillips et al. 2020); and even understand simple forms of outdated/inaccurate informational relations (as a proxy of false beliefs): An agent may have registered that an object O was at place 1, but failed to register that O then moved to place 2, and thus acts on her outdated registration of O at place 1.

But what such tracking of simpler representational "belief-like" states lacks is an appreciation not just of *what* agents have represented, but *how* they did so. Beliefs are essentially subjective or aspectual states in the sense that they are individuated in conceptually fine-grained ways: It matters fundamentally *how* the agent represents the object or states of affairs under consideration. Oedipus believes he ought to marry Yocasta, but he certainly does not believe he ought to marry his mother – even though, of course, Yocasta is his mother (unbeknownst to him). The more basic tracking of belief-like states, however, does not admit of such fine-grained, aspectual distinctions that are critical for any fully-fledged propositional attitude concepts. The informational connections of agents to objects and situations that the more basic system can track are purely relational, like non-epistemic seeing, for example (Dretske 1969). If Oedipus registers Yocasta, he thereby registers his mothers. As a consequence, reports about such informational relations are extensional (if agent A registers that O is in location L, and if O is identical to Z, then A registers that Z is in L as well), in contrast to the intensionality of propositional attitude reports (if A believes that O is in L, and O is identical to Z, then it does not necessarily follow that A believes that Z is in L – A may be as unaware about the O=Z identity as Oedipus is about the Yocasta=my mother identity).

This two-system account thus predicts clear signature limits of early, implicit ToM capacities: Infants (and adults), in their looking behaviour, can master level-I perspective taking situations, and some FB tasks (those that can be solved by merely keeping track of who has registered what); but they cannot master level-II perspective taking problems, nor FB tasks that require the ascription of aspectual beliefs proper (regarding *how* an agent represents a situation). Nativism and conceptual change/two-systems views thus make clearly competing predictions with regard to the scope and limits of early ToM capacities: Nativist accounts assume that infants should be able, in principle, to solve all kinds of perspective-taking and FB tasks. In practice, any given limitations that may nevertheless arise in infants' performance should have nothing to do with the content of the tasks, but only with extraneous (e.g., linguistic or inhibitory) tasks demands. In contrast, conceptual change/two-systems views posit characteristic content-related signature limits such that infants fail all tasks to do with the subjectivity and aspectuality of *how* agents represent situations.

Now, what does the empirical evidence say vis-à-vis these competing positions and predictions? Unfortunately, we currently do now know. The pattern of existing evidence that speaks to this question is complex, mixed and inconclusive. There is evidence from looking time and interaction studies both for (e.g., Edwards and Low 2017; Fizke et al. 2017; Low et al. 2014; Low and Watts 2013; Oktay-Gür et al. 2018; Surtees et al. 2012) and against the signature limits predicted by the two-systems view (e.g., F. Buttelmann et al. 2015; Elekes et al. 2017; Kampis and Kovács 2022; Scott and Baillargeon 2009). More systematic and comprehensive research is thus needed in the future to test for such signature limits.

In summary, what evidence from merely implicit ToM tasks would show with regard to the underlying cognitive capacities and processes is very much contested. Strong nativist accounts claim such evidence would show infants have meta-representational beliefs about other agents' beliefs. Conceptual change and two-systems views concede that such evidence would indicate some form of ToM, some representational sensitivity to others' representational states; but would claim that the states in question need not be fully-fledged meta-representational beliefs, nor need they involve the ascription of fully-fledged beliefs.

3. Is There Such Evidence?

All of these foregoing debates arise in response to the question: How should we interpret findings from looking time and related implicit ToM tasks? They presuppose the reliability and validity of these findings. But the premise presupposed – that the findings are reliable and valid – has recently come under empirical attack in the course of a serious replication crisis. For all of the implicit measures reported earlier, many more recent studies with infants, children and adults, often with bigger samples than the initial studies, have either failed to replicate the original findings (thus putting into question their reliability), or have suggested that the effects vanish under more stringent conditions (thus putting into question their validity). A comprehensive and detailed review of this complex empirical situation goes beyond the scope of the present chapter (for such reviews, see Baillargeon et al. 2018; Barone et al. 2019; Kulke and Rakoczy 2018; Poulin-Dubois et al. 2018; Rakoczy 2022). But here is a short summary:

Regarding VoE studies, initial positive findings came from relatively few labs. In addition, it has recently been noted that these findings are difficult to interpret since across these initial studies, no consistent set of methodological parameters (such as inclusion criteria, end-of-trial criteria etc.) has been used (Rubio-Fernández 2019). Independent replication results have been mixed, with some successful, some mixed, and some failed replications of original findings (Dörrenberg et al. 2018; Powell et al. 2018; Yott and Poulin-Dubois 2016).

Similarly, with regard to interaction studies, the robustness of original findings is currently under dispute: Some studies did replicate original findings at least partially (Fizke et al. 2017; Király et al. 2018); other studies failed to replicate original findings and thus question their reliability (Dörrenberg et al. 2018; Dörrenberg et al. 2019; Grosse Wiesmann et al. 2016; but see Rubio-Fernandez et al. 2021; Wenzel et al. 2020); finally, yet other studies replicated some original findings but questioned their validity: They produced additional evidence to suggest that these findings need not indicate what they were taken to indicate, rich ToM (Priewasser et al. 2018).

As noted earlier, the broadest body of original evidence comes from AL studies. Similarly, for this measure we also have the biggest and most systematic corpus of replication data. In addition, replication data from AL methods are most straightforward to interpret since direct (in contrast to merely conceptual) replications are possible given these studies run in completely automated ways (on eye-tracking machines). Several large-scale replication studies with several hundreds of children and adults, many of them as direct replications as possible, with exactly the same original stimuli and methods, yield a relatively coherent yet disappointing picture: Original findings could largely not be replicated (e.g., Burnside et al. 2018; Dörrenberg et al. 2018; Kulke et al. 2019; Kulke and Rakoczy 2019; Kulke, Reiß, et al. 2018; Kulke, von Duhn, et al. 2018; Kulke, Wübker, et al. 2019; Schuwerk et al. 2018). This also includes a failed self-replication attempt by the original authors of Southgate et al. (2007), one of the first and most influential AL studies (Kampis et al. 2021). There were two exceptions to this pattern: One condition ("FB1") from one study (Southgate et al. 2007) could mostly be replicated but is so highly ambiguous that it is impossible to interpret by itself. Another study (Low and Watts 2013; location condition) also stood out in that it could be replicated; but follow-up replication studies showed that this effect vanished once crucial confounds were removed and thus suggest, in terms of validity, that the original task did not measure what it was designed to measure (ToM) (Kulke, von Duhn, et al. 2018).

Finally, another crucial question concerns the relation of different implicit measures to each other. If they all tap the same underlying capacity – implicit ToM – then they should all converge and correlate. Such patterns have been widely observed in the case of explicit ToM tasks:

Despite massive differences in surface features, formats and measures, there is substantial correlation, and thus convergent validation between explicit ToM tasks (for review, see Perner and Roessler 2012). In contrast, several recent studies tested for such correlations between implicit tasks. Neither within a given type of task (such as AL, Kulke, Reiß, et al. 2018; Kulke, von Duhn, et al. 2018) nor between types of tasks were there any systematic correlations (e.g., Dörrenberg et al. 2018; Poulin-Dubois and Yott 2017; Yott and Poulin-Dubois 2016).

All in all, in light of the complex and inconclusive empirical situation it is currently neither clear whether initial evidence from implicit ToM tasks is robustly replicable (reliability) nor whether it actually measures what it is supposed to measure (validity). Clearly, however, interpreting complex patterns of existing original and replication evidence in post-hoc ways is only of limited epistemic value. What is needed is a concerted effort to look and move ahead. Fortunately, exactly this is now happening: A large-scale international consortium has recently constituted itself under the umbrella of the *ManyBabies* initiative (Frank et al. 2017). This consortium involves original authors as well as authors of replication studies and many other experts in ToM research and brings together scientists from all theoretical backgrounds. In a true case of "adversarial collaboration" (Mellers et al. 2001), the group collectively agrees upon competing predictions made from diverging theoretical perspectives and how to test them against each other in large-scale, multi-lab, preregistered replication and validation studies. In due time, hopefully, the findings from these large-scale studies will thus allow us to reach firmer conclusions about the (non-)existence of solid evidence from various measures for various forms of early implicit ToM (Schuwerk et al. 2022).

4. Conclusion

In this chapter, we have discussed the conceptual question how solid evidence from implicit ToM tasks would need to be interpreted theoretically, and the empirical question whether there is such evidence. With regard to the empirical question, the current situation is very much inconclusive, but progress is on the way in the form of the collaborative replication studies of the *ManyBabies* consortium. With regard to the conceptual question, I have tried to argue that the space of theoretical options is bigger and more complex than often assumed. If findings from implicit tasks turn out to be robust, this would present solid evidence that infants are representationally sensitive to others' epistemic situations. But this in itself would neither mean that they form meta-representational beliefs (rather than some simpler representational states) nor that they ascribe fully-fledged beliefs (rather than some simpler representational states). Future work in this area certainly will need to map out this conceptual territory (of representational states about representational states that do not yet amount to beliefs about beliefs) in clearer and more fine-grained ways; and to devise new tasks that allow us to locate the cognitive capacities of a given creature within the space ranging from more basic to fully-fledged forms of (meta-) representational functioning.

Another challenge for future theory building and experimentation in this area will be to overcome simplistic presumptions of a 1:1 correspondence of task and process such that a given type of (implicit) task taps certain types of (implicit) processes whereas other types of (explicit) tasks tap other types of (explicit) processes. Research on implicit vs. explicit processes in other domains of cognitive science has found ways to overcome and move beyond such assumption of the "process purity" of types of tasks. In memory research, in particular, so-called *process-dissociation procedures* (Jacoby 1991) have been devised in order to isolate implicit and explicit memory processes in different kinds of direct and indirect tasks. The background assumption here is that, under suitable design conditions, it can be formally spelled out in precise ways that

and how different kinds of processes play into different kinds of tasks to which degrees. Such an approach has only recently been adopted to the study of some aspects of adult ToM (Todd et al. 2017; Todd et al. 2019). Hopefully, future developmental and cognitive ToM research will adopt and refine such more nuanced approaches to the study of implicit vs. explicit cognition across the board.[5]

Notes

1 Implicit tasks have widely been used with infants before they master explicit ones in principle. But implicit tasks have also been used with adults in situations where they are supposedly not consciously aware of engaging in any ToM reasoning, for example where they are not asked to reason about a protagonist's belief but their eyes, so to speak, engage in AL nonetheless (Schneider et al. 2012). In the following, for reasons of simplicity I will often refer to infants, but similar questions apply for the adult data.

2 Block distinguishes between two notions of consciousness: phenomenal (what it is like) and access consciousness (roughly, whether a given representational state is accessible to the subject for use in reasoning, planning, action and language). According to Block, while the two mostly converge, they are conceptually not wholly overlapping and there are thus rare cases of dissociations in both directions. I will here ignore these complications and merely focus on access consciousness.

3 Actually, when it comes to the interpretation of interactive tasks, things are more complicated. These tasks, intuitively, even if they do not document fully-fledged flexible conceptual thought, at least go beyond mere looking time tasks in that they involve some use of the requisite information for action planning (see Carruthers 2013). In fact, if infants revealed seemingly meta-representational deliberation and planning capacities in interactive tasks in the way older children and adults do, with the only exception that they are not yet able to express these capacities verbally, this would be convincing evidence that they operated with something very close to (access-conscious) beliefs about other agents' beliefs. But so far, the interactive capacities of infants in such tasks have been very limited (not to mention the fact, discussed in the next section, that the reliability and validity even of these very limited interactive tasks has recently come into question). Against this background, and given space limitation, I am here ignoring these complications and focus on looking time as main indicator of early implicit ToM (see, e.g., Newen and Wolf (2020) for further distinctions between looking time and interactive studies on early FB understanding).

4 Beliefs are conceptually structured, holistically related, normatively governed etc., to name just some of those complexities (for details of these arguments, see Burge 2018; Butterfill and Apperly 2013).

5 Thank you for helpful comments on a previous version to Natalie Bleijlevens, Isa Garbisch, Feride Nur Haskaraca Kizilay, Marina Proft, Lydia Schidelko, Britta Schünemann, Rowan Titchener and Lisa Wenzel. I am deeply grateful to Robert Thompson and an anonymous reviewer for incredibly detailed, constructive and helpful feedback on the first version of this chapter.

References

Apperly, I. A., and Butterfill, S. A. 2009. "Do humans have two systems to track beliefs and belief-like states?". *Psychological Review*, *116*: 953–970. https://doi.org/10.1037/a0016923

Baillargeon, R., Buttelmann, D., and Southgate, V. 2018. "Invited commentary: Interpreting failed replications of early false-belief findings: methodological and theoretical considerations". *Cognitive Development, 46:* 112–124. https://doi.org/10.1016/j.cogdev.2018.06.001

Baillargeon, R., Scott, R. M., and Bian, L. 2016. "Psychological reasoning in infancy". *Annual Review of Psychology*, *67*: 159–186.

Baillargeon, R., Scott, R. M., and He, Z. 2010. "False-belief understanding in infants". *Trends in Cognitive Sciences*, *14*: 110–118. https://doi.org/10.1016/j.tics.2009.12.006

Barone, P., Corradi, G., and Gomila, A. 2019. "Infants' performance in spontaneous-response false belief tasks: a review and meta-analysis". *Infant Behavior and Development, 57:* 101350. https://doi.org/10.1016/j.infbeh.2019.101350

Block, N. 1995. "On a confusion about a function of consciousness". *Behavioral and Brain Sciences*, *18*: 227–247.

Burge, T. 2010. *Origins of objectivity*. Oxford: Oxford University Press.

Burge, T. 2018. "Do infants and nonhuman animals attribute mental states?". *Psychological Review*, *125*: 409–434. https://doi.org/10.1037/rev0000091

Burnside, K., Ruel, A., Azar, N., and Poulin-Dubois, D. 2018. "Implicit false belief across the lifespan: non-replication of an anticipatory looking task". *Cognitive Development*, *46*: 4–11. https://doi.org/10.1016/j.cogdev.2017.08.006

Buttelmann, D., Carpenter, M., and Tomasello, M. 2009. "Eighteen-month-old infants show false belief understanding in an active helping paradigm". *Cognition*, *112*: 337–342.

Buttelmann, F., Suhrke, J., and Buttelmann, D. 2015. "What you get is what you believe: eighteen-month-olds demonstrate belief understanding in an unexpected-identity task". *Journal of Experimental Child Psychology*, *131*: 94–103. http://dx.doi.org/10.1016/j.jecp.2014.11.009

Butterfill, S. A., and Apperly, I. A. 2013. "How to construct a minimal Theory of Mind". *Mind and Language, 28:* 606–637. https://doi.org/10.1111/mila.12036

Carlson, S. M., and Moses, L. J. 2001. "Individual differences in inhibitory control and children's theory of mind". *Child Development*, *72*: 1032–1053.

Carruthers, P. 2013. "Mindreading in infancy". *Mind and Language*, *28*: 141–172. https://doi.org/10.1111/mila.12014

Clark, A., and Karmiloff-Smith, A. 1993. "The cognizer's innards: a philosophical and psychological perspective on the development of thought". *Mind and Language*, *8*: 487–519.

Cummins, R. 1983. *The nature of psychological explanation*. Cambridge, MA: MIT Press.

Davies, M. 1989. "Tacit knowledge and subdoxastic states". In A. George, ed., *Reflections on Chomsky*. Oxford: Blackwell: 131–152.

Dennett, D. 1978. "Beliefs about beliefs". *Behavioral and Brain Sciences, 4:* 568–570.

Devine, R. T., and Hughes, C. 2014. "Relations between false belief understanding and executive function in early childhood: a meta-analysis". *Child Development*, *85*: 1777–1794.

Dörrenberg, S., Rakoczy, H., and Liszkowski, U. 2018. "How (not) to measure infant Theory of Mind: testing the replicability and validity of four non-verbal measures". *Cognitive Development, 46:* 12–30. https://doi.org/10.1016/j.cogdev.2018.01.001

Dörrenberg, S., Wenzel, L., Proft, M., Rakoczy, H., and Liszkowski, U. 2019. "Reliability and generalizability of an acted-out false belief task in 3-year-olds". *Infant Behavior and Development*, *54*: 13–21. https://doi.org/10.1016/j.infbeh.2018.11.005

Dretske, F. 1969. *Seeing and knowing*. Chicago: University of Chicago Press.

Edwards, K., and Low, J. 2017. "Reaction time profiles of adults' action prediction reveal two mindreading systems". *Cognition*, *160*: 1–16. http://dx.doi.org/10.1016/j.cognition.2016.12.004

Elekes, F., Varga, M., and Király, I. 2017. "Level-2 perspectives computed quickly and spontaneously: evidence from eight- to 9.5-year-old children". *British Journal of Developmental Psychology*, *35*: 609–622. https://doi.org/10.1111/bjdp.12201

Fizke, E., Butterfill, S., van de Loo, L., Reindl, E., and Rakoczy, H. 2017. "Are there signature limits in early theory of mind?". *Journal of Experimental Child Psychology, 162:* 209–224. https://doi.org/10.1016/j.jecp.2017.05.005

Flavell, J. H., Everett, B. A., Croft, K., and Flavell, E. R. 1981. "Young children's knowledge about visual perception: further evidence for the Level 1-Level 2 distinction". *Developmental Psychology*, *17*: 99–103.

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., . . . Yurovsky, D. 2017. "A collaborative approach to infant research: promoting reproducibility, best practices, and theory-building". *Infancy*, *22*: 421–435. https://doi.org/10.1111/infa.12182

Frankish, K. 2009. "Systems and levels: dual-system-theories and the personal-subpersonal distinction". In J. S. B. T. Evans and K. Frankish, eds., *In two minds: dual processes and beyond*. New York: Oxford University Press: 89–107.

Gopnik, A., and Astington, J. W. 1988. "Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction". *Child Development*, *59*: 26–37.

Grosse Wiesmann, C., Friederici, A. D., Singer, T., and Steinbeis, N. 2016. "Implicit and explicit false belief development in preschool children". *Developmental Science*, *20*: e12445. https://doi.org/10.1111/desc.12445

Grosse Wiesmann, C., Schreiber, J., Singer, T., Steinbeis, N., and Friederici, A. D. 2017. "White matter maturation is associated with the emergence of Theory of Mind in early childhood". *Nature Communications*, *8*: 14692. https://doi.org/10.1038/ncomms14692

Hayashi, T., Akikawa, R., Kawasaki, K., Egawa, J., Minamimoto, T., Kobayashi, K., . . . Hasegawa, I. 2020. "Macaques exhibit implicit gaze bias anticipating others' false-belief-driven actions via medial prefrontal cortex". *Cell Reports*, *30*: 4433–4444. https://doi.org/10.1016/j.celrep.2020.03.013

Heyes, C. 2014. "Submentalizing: I am not really reading your mind". *Perspectives on Psychological Science*, *9*: 131–143.

Hyde, D. C., Simon, C. E., Ting, F., and Nikolaeva, J. I. 2018. "Functional organization of the temporal-parietal junction for theory of mind in preverbal infants: a near-infrared spectroscopy study". *The Journal of Neuroscience*, *38*: 4264–4274. https://doi.org/10.1523/jneurosci.0264-17.2018

Jacoby, L. L. 1991. "A process dissociation framework: separating automatic from intentional uses of memory". *Journal of Memory and Language*, *30*: 513–541. https://doi.org/10.1016/0749-596X(91)90025-F

Kampis, D., Kármán, P., Csibra, G., Southgate, V., and Hernik, M. 2021. "A two-lab direct replication attempt of Southgate, Senju and Csibra (2007)". *Royal Society Open Science*, *8*: 210190. https://doi.org/10.1098/rsos.210190

Kampis, D., and Kovács, Á. M. 2022. "Seeing the world from others' perspective: 14-month-olds show altercentric modulation effects by others' beliefs". *Open Mind*: 1–19. https://doi.org/10.1162/opmi_a_00050

Király, I., Oláh, K., Csibra, G., and Kovács, Á. M. 2018. "Retrospective attribution of false beliefs in 3-year-old children". *Proceedings of the National Academy of Sciences*, *115*: 11477–11482. https://doi.org/10.1073/pnas.1803505115

Knudsen, B., and Liszkowski, U. 2012. "18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly". *Infancy, 17:* 672–691.

Krupenye, C., Kano, F., Hirata, S., Call, J., and Tomasello, M. 2016. "Great apes anticipate that other individuals will act according to false beliefs". *Science*, *354*: 110–114. https://doi.org/10.1126/science.aaf8110

Kulke, L., Johannsen, J., and Rakoczy, H. 2019. "Why can some implicit Theory of Mind tasks be replicated and others cannot? A test of mentalizing versus submentalizing accounts". *PLoS ONE*, *14*: e0213772. https://doi.org/10.1371/journal.pone.0213772

Kulke, L., and Rakoczy, H. 2018. "Implicit Theory of Mind – an overview of current replications and non-replications". *Data in Brief*, *16*(Supplement C): 101–104. https://doi.org/10.1016/j.dib.2017.11.016

Kulke, L., and Rakoczy, H. 2019. "Testing the role of verbal narration in implicit theory of mind tasks". *Journal of Cognition and Development*, *20*: 1–14. https://doi.org/10.1080/15248372.2018.1544140

Kulke, L., Reiß, M., Krist, H., and Rakoczy, H. 2018. "How robust are anticipatory looking measures of Theory of Mind? Replication attempts across the life span". *Cognitive Development*, *46*: 97–111. https://doi.org/10.1016/j.cogdev.2017.09.001

Kulke, L., von Duhn, B., Schneider, D., and Rakoczy, H. 2018. "Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study". *Psychological Science, 29:* 888–900. https://doi.org/10.1177/0956797617747090

Kulke, L., Wübker, M., and Rakoczy, H. 2019. "Is implicit Theory of Mind real but hard to detect? Testing adults with different stimulus materials". *Royal Society Open Science, 6:* 190068. https://doi.org/10.1098/rsos.190068

Leslie, A. M. 2005. "Developmental parallels in understanding minds and bodies". *Trends in Cognitive Science*, *9*: 459–462. https://doi.org/10.1016/j.tics.2005.08.002

Low, J., Drummond, W., Walmsley, A., and Wang, B. 2014. "Representing how rabbits quack and competitors act: limits on preschoolers' efficient ability to track perspective". *Child Development*, *85*: 1519–1534. https://doi.org/10.1111/cdev.12224

Low, J., and Watts, J. 2013. "Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system". *Psychological Science*, *24*: 305–311. https://doi.org/10.1177/0956797612451469

Mellers, B., Hertwig, R., and Kahneman, D. 2001. "Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration". *Psychological Science*, *12*: 269–275.

Newen, A., and Wolf, J. 2020. "The situational mental file account of the false belief tasks: a new solution of the paradox of false belief understanding". *Review of Philosophy and Psychology, 11:* 717–744. https://doi.org/10.1007/s13164-020-00466-w

Oktay-Gür, N., Schulz, A., and Rakoczy, H. 2018. "Children exhibit different performance patterns in explicit and implicit theory of mind tasks". *Cognition*, *173*: 60–74. https://doi.org/10.1016/j.cognition.2018.01.001

Onishi, K. H., and Baillargeon, R. 2005. "Do 15-month-old infants understand false beliefs?". *Science*, *308*: 255–258. https://doi.org/10.1126/science.1107621

Perner, J. 1991. *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perner, J., and Roessler, J. 2012. "From infants' to children's appreciation of belief". *Trends in Cognitive Sciences*, *16*: 519–525. https://doi.org/10.1016/j.tics.2012.08.004

Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., . . . Knobe, J. 2020. "Knowledge before belief". *Behavioral and Brain Sciences*, *44*: 1–37. https://doi.org/10.1017/s0140525x20000618

Poulin-Dubois, D., Rakoczy, H., Burnside, K., Crivello, C., Dörrenberg, S., Edwards, K., . . . Ruffman, T. 2018. "Do infants understand false beliefs? We don't know yet – a commentary on Baillargeon, Buttelmann and Southgate's commentary". *Cognitive Development, 48:* 302–315. https://doi.org/10.1016/j.cogdev.2018.09.005

Poulin-Dubois, D., and Yott, J. 2017. "Probing the depth of infants' theory of mind: disunity in performance across paradigms". *Developmental Science*, *21*: e12600. https://doi.org/10.1111/desc.12600

Powell, L. J., Hobbs, K., Bardis, A., Carey, S., and Saxe, R. 2018. "Replications of implicit theory of mind tasks with varying representational demands". *Cognitive Development, 46:* 40–50. https://doi.org/10.1016/j.cogdev.2017.10.004

Priewasser, B., Rafetseder, E., Gargitter, C., and Perner, J. 2018. "Helping as an early indicator of a theory of mind: mentalism or teleology?". *Cognitive Development*, *46*: 69–78. https://doi.org/10.1016/j.cogdev.2017.08.002

Rakoczy, H. 2017. "In defense of a developmental dogma: children acquire propositional attitude folk psychology around age 4". *Synthese*, *194*: 689–707. https://doi.org/10.1007/s11229-015-0860-8

Rakoczy, H. 2022. "Foundations of theory of mind and its development in early childhood". *Nature Reviews Psychology.* https://doi.org/10.1038/s44159-022-00037-z

Rubio-Fernández, P. 2019. "Publication standards in infancy research: three ways to make violation-of-expectation studies more reliable". *Infant Behavior and Development*, *54*: 177–188. https://doi.org/10.1016/j.infbeh.2018.09.009

Rubio-Fernandez, P., Southgate, V., and Király, I. 2021. "Pragmatics for infants: commentary on Wenzel et al. (2020)". *Royal Society Open Science*, *8*: 210247. https://doi.org/10.1098/rsos.210247

Schneider, D., Bayliss, A. P., Becker, S. I., and Dux, P. E. 2012. "Eye movements reveal sustained implicit processing of others' mental states". *Journal of Experimental Psychology: General*, *141*: 433–438.

Schuwerk, T., Jarvers, I., Vuori, M., and Sodian, B. 2016. "Implicit mentalizing persists beyond early childhood and is profoundly impaired in children with autism spectrum condition". *Frontiers in Psychology,* 7: 1696. https://doi.org/10.3389/fpsyg.2016.01696

Schuwerk, T., Kampis, D., Baillargeon, R., Biro, S., Bohn, M., Byers-Heinlein, K., . . . Rakoczy, H. 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).

Schuwerk, T., Priewasser, B., Sodian, B., and Perner, J. 2018. "The robustness and generalizability of findings on spontaneous false belief sensitivity: a replication attempt". *Royal Society Open Science*, *5*: 172273. https://doi.org/10.1098/rsos.172273

Scott, R. M., and Baillargeon, R. 2009. "Which penguin is this? Attributing false beliefs about object identity at 18 months". *Child Development*, *80*: 1172–1196.

Scott, R. M., and Baillargeon, R. 2017. "Early false-belief understanding". *Trends in Cognitive Sciences*, *21*: 237–249. https://doi.org/10.1016/j.tics.2017.01.012

Scott, R. M., Roby, E., and Baillargeon, R. (in press). "How sophisticated is infants' theory of mind?". In O. Houdé and G. Borst, eds., *Cambridge handbook of cognitive development*. Cambridge: Cambridge University Press.

Senju, A., Southgate, V., White, S., and Frith, U. 2009. "Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome". *Science*, *325*: 883–885.

Shoemaker, S. 1981. "Some varieties of functionalism". *Philosophical Topics*, *12*: 93–119.

Southgate, V., Chevallier, C., and Csibra, G. 2010. "Seventeen-month-olds appeal to false beliefs to interpret others referential communication". *Developmental Science, 13:* 907–912.

Southgate, V., Senju, A., and Csibra, G. 2007. "Action anticipation through attribution of false belief by 2-year-olds". *Psychological Science*, *18*: 587–592.

Southgate, V., and Vernetti, A. 2014. "Belief-based action prediction in preverbal infants". *Cognition*, *130*: 1–10. http://dx.doi.org/10.1016/j.cognition.2013.08.008

Stich, S. 1978. "Beliefs and subdoxastic states". *Philosophy of Science, 45:* 499–518.

Surian, L., and Geraci, A. 2012. "Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months". *British Journal of Developmental Psychology*, *30*: 30–44. https://doi.org/10.1111/j.2044-835X.2011.02046.x

Surtees, A. D. R., Butterfill, S. A., and Apperly, I. A. 2012. "Direct and indirect measures of level-2 perspective-taking in children and adults". *British Journal of Developmental Psychology, 30:* 75–86. https://doi.org/10.1111/j.2044-835X.2011.02063.x

Todd, A. R., Cameron, C. D., and Simpson, A. J. 2017. "Dissociating processes underlying level-1 visual perspective taking in adults". *Cognition*, *159*: 97–101. https://doi.org/10.1016/j.cognition.2016.11.010

Todd, A. R., Simpson, A. J., and Cameron, C. D. 2019. "Time pressure disrupts level-2, but not level-1, visual perspective calculation: a process-dissociation analysis". *Cognition*, *189*: 41–54. https://doi.org/10.1016/j.cognition.2019.03.002

Wellman, H., Cross, D., and Watson, J. 2001, "Meta-analysis of theory-of-mind development: the truth about false belief". *Child Development, 72:* 655–684.

Wenzel, L., Dörrenberg, S., Proft, M., Liszkowski, U., and Rakoczy, H. 2020. "Actions do not speak louder than words in an interactive false belief task". *Royal Society Open Science*, 7: 191998. https://doi.org/10.1098/rsos.191998

Wimmer, H., and Perner, J. 1983. "Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception". *Cognition*, *13*: 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Yott, J., and Poulin-Dubois, D. 2016. "Are infants' theory-of-mind abilities well integrated? Implicit understanding of intentions, desires, and beliefs". *Journal of Cognition and Development*, *17*: 683–698. https://doi.org/10.1080/15248372.2015.1086771