Journal of Experimental Child Psychology 249 (2025) 106103

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Thinking about thinking: A longitudinal investigation linking developments in metacognition, inhibitory control, and theory of mind

Mariel Symeonidou [a], Martin J. Doherty [b], Josephine Ross [a,*]

[a] Psychology, School of Humanities, Social Sciences, and Law, University of Dundee, Dundee DD1 4HN, Scotland, UK
[b] School of Psychology, University of East Anglia, Norwich NR4 7TJ, UK

ARTICLE INFO

Article history:
Received 11 April 2024
Revised 13 September 2024

Keywords:
Metacognition
Introspection
Inhibitory Control
Theory of Mind
Longitudinal Development
Child Development

ABSTRACT

This longitudinal study tracked the developmental relations linking metacognition, theory of mind, and inhibitory control in 52 children across a 1.5-year interval, beginning at 3 or 4 years of age. Metacognition and inhibitory control emerged before theory of mind and predicted subsequent theory of mind competence. Moreover, there was evidence of developmental mediation, whereby metacognition predicted inhibitory control, which predicted theory of mind. We suggest that metacognitive self-reflection may provide the "developmental enrichment" necessary to think about thinking, and when inhibitory control is sufficiently developed this thinking can be extended to complex reasoning about own and other minds.

Introduction

Metacognition can be defined as the capacity to reflect on our own mental states, our ability to think about our own thinking (Flavell, 2002). Deceptively simple to describe, the term metacognition has grown to encompass an umbrella of definitions and skills, including metacognitive knowledge,

* Corresponding author.
E-mail address: juross@dundee.ac.uk (J. Ross).

https://doi.org/10.1016/j.jecp.2024.106103
0022-0965/Crown Copyright © 2024 Published by Elsevier Inc.

experience, goals, and strategies, and extending to related cognitive developments, such as inhibitory control (the ability to control our own thoughts and behavior) and theory of mind (ToM; the ability to reason flexibly about our own and others' mental states) (Veenman et al., 2006). Reflecting on this complexity, Kuhn (2000) proposed that metacognitive monitoring (the ability to reflect and report on mental processes by introspection) may serve as a developmental foundation for higher-order aspects of metacognitive function, such as inhibitory control and ToM. This is in keeping with our understanding of development as a process through which higher-order skills build on early foundations. However, careful operationalization and definition of these three skill sets, alongside longitudinal measurement, is needed to separate their developmental trajectories and interrelations. The current article offers the first longitudinal investigation of 3- to 5-year-olds' explicit metacognitive monitoring, inhibitory control, and ToM, comparing developmental progression in these cognitive skill sets and addressing their interrelationships to elucidate the developmental mechanism by which young children come to understand other minds.

Cross-sectional research suggests that explicit forms of metacognition, inhibitory control, and ToM emerge across the third to fifth years of childhood (see Goupil & Kouider, 2019, Petersen et al., 2016, and Rakoczy, 2022, for reviews). In keeping with the overlapping development and definition of metacognition and ToM, there is ongoing philosophical debate concerning whether explicit metacognition and ToM are part of the same metarepresentational system or best viewed as independent processes (e.g., see Carruthers, 2009; Proust, 2010). Moreover, proponents of the independent processes differ in their views concerning whether we need to have awareness of our own mental states in order to open reflection on others' mental states (Goldman, 2006) or need to first develop a theory of others' mental states to take a metarepresentational stance on our own thinking (Carruthers, 2011). These stances can be related to overarching theories of the mechanisms underpinning ToM (see Alcalá-López et al., 2019, for review).

Historically, simulation theorists (Harris, 1992; Gallese & Goldman, 1998; Goldman, 2006) have argued that we come to understand others' mental states by reasoning by analogy from our own minds, suggesting that metacognitive monitoring is separate from, and a prerequisite to, ToM. On the contrary, proponents of theory-theory (Gopnik & Meltzoff, 1997; Gopnik & Wellman, 1992) argue that ToM develops through experience of how others act given certain situational factors and mental states. On this reading, our thinking about thinking is based on our experience of trying to predict and explain others' behavior rather than reflection on our internal states. Depending on the formulation, theory-theory is compatible with a single system view, where representational understanding of minds does not distinguish between first- and second-person experience, or an independent process view, in which metarepresentational reflection on our own mind is derived from representation of other minds. The weight of the available longitudinal evidence speaks to the latter, whereby explicit representation of others' minds leads to deeper understanding of our own mind. Several studies have reported that ToM at 4 years of age predicts children's ability to understand the factors that influence memory (e.g., item number, prompts) and effective memory strategies (e.g., use of prompts) later in childhood (Lecce et al., 2015; Lockl & Schneider, 2006; Ebert, 2015). Lecce et al. (2014) also found that ToM at 5 years of age predicted children‘s understanding of their own learning processes at 8 years. However, researchers focused on more basic forms of metacognitive monitoring (reporting level of certainty in a memory) find that early metacognition (at 6 years of age) predicts later ToM (at 7 years) (Feurer et al., 2015; cf. Kloo et al., 2021, but metacognition is measured only at the later time point, biasing directionality). However, these studies involve older children, considering the consequences of developing explicit ToM and metacognition rather than their emergence across 3 to 5 years of age.

Far more research has focused on the co-development of inhibitory control and ToM at the end of infancy. Although the definitions of inhibitory control and ToM show little overlap, there is robust evidence of a strong positive relationship between these skill sets as they emerge (see Devine & Hughes, 2014, for review). Originally, proponents of theory-theory argued that ToM may help us to develop inhibitory control, arguing that it may be necessary to recognize the causal role of mental states in order to reflect on and control our current behavior (e.g., Perner, 1998). However, inhibitory control is thought to be a requirement of ToM tasks, which often require the participants to avoid responding on the basis of their own knowledge (Carlson et al., 2015; Perner & Lang, 1999; Wellman et al., 2001). Moreover, it has since become clear that when measured explicitly, inhibitory control emerges before, and longitudinally predicts, ToM (Flynn, 2007; Flynn et al., 2004; Hughes & Ensor, 2007; Marcovitch et al., 2015). As a result, the dominant view of the relationship between inhibitory control and ToM has switched to an ''emergence'' or ''developmental enrichment'' account, capturing the idea that inhibitory control and other executive functions may be necessary for us to reason flexibly about other minds (Kloo et al., 2020).

As set out above, inhibitory control might be thought of as a higher-order form of metacognition (see Roebers, 2017, for review). This definitional overlap, together with a lack of longitudinal evidence, makes it difficult to unpick the developmental trajectories of metacognitive monitoring and inhibitory control (Kälin & Roebers, 2022). The first longitudinal study of these skill sets found some evidence to suggest that performance on an inhibitory control task at 7 years of age was predictive of metacognitive control (as measured by the ability to withdraw incorrectly spelled words in a spelling test) at 8 years (Roebers et al., 2012). Moreover, a recent longitudinal study following the development of 5- and 6-year-old children replicated this result for metacognitive monitoring, finding that children's performance on an inhibitory control task was predictive of their ability to accurately report their level of certainty during a recall task 1 year later (Kälin & Roebers, 2022). Thus, it is possible that inhibitory control is involved in the development of children's ability to explicitly take a metarepresentational stance on their own thoughts. However, given that these studies focused on older children, it remains possible that becoming aware of our *own* thoughts at around 3 years of age is the first step to cognitive inhibitory control.

The current study

To explore potential causal relations linking metacognitive monitoring, inhibitory control, and ToM, we provide longitudinal data tracking the emergence of explicit forms of these skill sets across 3 to 5 years of age (see Goupil & Kouider, 2019, Hendry et al., 2022, and Rakoczy, 2022, for discussion of implicit forms of metacognition, inhibitory control, and ToM predating this developmental stage). Metacognitive monitoring—defined as the ability to reflect and report on mental processes by introspection—was operationalized in an established perceptual certainty monitoring task first developed by Lyons and Ghetti (2011). Inhibitory control was operationalized across two well-established tasks for 3- to 5-year-olds, focused on following rules to inhibit prepotent behavioral responses (Reed et al.'s (1984) Bear/Dragon task) and verbal responses (Gerstadt et al.'s (1994) Day/Night Stroop task). Finally, ToM was measured using an adaptation of an established five-task battery, first developed by Wellman and Liu (2004), to capture development in children's understanding of beliefs, knowledge, and emotions across 3 to 6 years of age. To establish the developmental relations among metacognitive monitoring, inhibitory control, and self- and other-referent ToM, we considered the emergence of each skill set based on chance performance, the developmental trajectories of task performance, and the predictive relations across tasks and time. Children also completed the receptive short form of the British Picture Vocabulary Scale (BPVS) (Dunn et al., 1997) to allow us to control for common variance in performance associated with language development given that this has been identified as an important variable in ToM development (Ebert, 2015; Milligan et al., 2007).

To our knowledge, no extant longitudinal research has explored the developmental relations linking explicit metacognitive monitoring, inhibitory control, and ToM during the emergence of these skill sets at 3 or 4 years of age. This is theoretically important because simulation theory, but not theory-theory, predicts that metacognition and inhibitory control provide the developmental building blocks of ToM. If we are to reason by analogy from our own mental states to others, to the extent that our current perspective is salient, we may require inhibitory control to set our own perspective aside to ''simulate'' other (Carlson et al., 2015). Employing a novel developmental perspective on simulation theory, we proposed that the capacity to reflect on current mental states might be the first step to understanding mental states separately from actions. When inhibitory control is sufficiently developed to allow us to pause reflection on our own mental states and turn this lens outward, this thinking may be extended to more complex reasoning about other minds. On this reading, children might not exclusively or routinely solve ToM tasks by simulating how they would think in another person's shoes, but the capacity to reflect on their own mental states may be a necessary developmental step to open representation of the mental states of other people. Our novel developmental simulation account would predict a specific order of emergence of metacognition > inhibitory control > ToM and a specific pattern of interrelation whereby the time lagged links between metacognition and ToM are dependent on the emergence of inhibitory control. On the contrary, a single process theory-theory would predict the shared emergence of ToM and metacognition (both of which may or may not relate to inhibitory control), whereas an independent process theory-theory might cast ToM as the developmental foundation to a mentalistic understanding of self (and inhibitory control, depending on the formulation).

Method

Participants

A total of 71 children were originally recruited for the study; of this original sample, 16 children were lost to attrition and 3 had incomplete data. This left 52 children (58% male) with a full dataset for all three waves of the study. The sample size was comparable to previous studies (Sodian et al., 2020) and was determined by convenience sampling. Sensitivity for moderate main effects of time were calculated before analysis using G*Power predictions for a repeated-measures, three-phase within-participants design (power = 80%, p = .05, η^2s = .04–.06) (Faul et al., 2007), which suggested a minimum sample size of 27 to 41 participants. Sensitivity for correlations was also calculated a priori, based on the robust correlation ($r \leq .35$) between ToM and inhibitory control, G*Power suggested a minimum sample size of 49 participants to achieve 80% power. Children were recruited from three nurseries in Dundee, Scotland. Race and socioeconomic data were not collected, but the children were recruited from predominantly White, lower- to middle-class Scottish areas. All of the children started at the study aged 3 or 4 years and exited the study aged 4 or 5 years. Children were tested in a quiet corner of their usual education environment. After baseline (Phase 1), children were followed up ~ 6 months later (Phase 2) and ~ 12 months later (Phase 3), with an error margin of 6 weeks. For the final sample, children were aged 39 to 54 months at Phase 1 (M = 46.81 months, SD = 3.77), 45 to 61 months at Phase 2 (M = 53.23 months, SD = 4.17), and 51 to 66 months at Phase 3 (M = 58.98 months, SD = 3.87). Ethical review and approval was provided by the University of Dundee non clinical research ethics committee. Data (https://osf.io/cqt72/) and supplementary material are available. https://osf.io/cqt72/?view_only=c797f80fc51b4607bf3dfab0177bc73c.

Materials and procedure

Tasks were completed across two sessions. In the first session, children completed inhibitory control and ToM tasks in a partially counterbalanced order. Children began and ended the session with an inhibitory control task (order counterbalanced) and completed the ToM battery in the interim; to minimize order effects, one group of children completed the ToM tasks in the order described below supplementary material, and the other group did so in reverse order. In the second session (1 or 2 days later), children completed the certainty monitoring task used to assess metacognition and the BPVS.

Inhibitory control

Children completed two classic inhibitory control tasks: an adaptation of Reed et al.'s (1984) Bear/Dragon task and Gerstadt et al.'s (1994) Day/Night Stroop task. The proportion correct in each task was averaged to create a total inhibitory control score.

Modified Bear/Dragon task. In this task, children were introduced to two hand puppets representing a "nice" knight and a "mean" dragon. The task was inspired by the popular children's game "Simon Says," where children are provided with instructions that they either need to follow or ignore depending on a rule. In this case, the rule was to follow the instruction of the nice knight and ignore the instructions of the mean dragon. The puppets then took turns instructing the children to perform simple actions (e.g., "Touch your nose"). Practice trials were administered until the children responded correctly (reacting to the knight and ignoring the dragon). Ten test trials were then presented, with the puppets issuing commands in a pseudorandomized order; after five trials, children were reminded of the rules. On knight trials, children were awarded 3 points for each trial that they fully imitated the action, 2 points if they partially imitated, 1 point if they performed a wrong movement, and no points if they did not follow the instruction. On dragon trials, children were awarded 3 points for not making any movement, 2 points if they performed the movement partially, 1 point if they made the wrong movement, and no points if they fully imitated the action. This resulted in a total score out of a possible 30 points.

Day–Night Stroop task. In this task, children were told that they would play a "silly game" that involved saying the opposite of what they might want to say. Children were instructed to say "night" when shown a picture of a sun and to say "day" when shown a picture of a moon. Practice trials were administered until the children responded correctly. Sixteen test trials were then administered in a pseudorandomized order, and 1 point was assigned for each correctly answered trial, building to a possible maximum of 16 points.

ToM battery

Children's understanding of others' mental states was assessed using a modified version of Wellman and Liu's (2004) five-item ToM battery. Like Wellman and Liu's battery, our scale contained an assessment of the children's understanding of diverse beliefs, knowledge access, false beliefs (as measured by the unexpected contents task), and hidden emotions. However, we did not include the diverse desires task given that this has been observed to be passed by 18 months of age (Repacholi & Gopnik, 1997). Instead, we supplemented the existing tasks by asking the children to track the protagonist's feelings during the unexpected contents task when the actual contents of the container were revealed to be disappointing (adapted from Bender et al., 2011). We also made an addition to the knowledge access task, asking children whether the protagonist was aware of the contents of the box (as in the original scale) *and* how the protagonist could find out what was in the box. Finally, we extended the scale by including an additional task, Russell et al.'s (2001) transparent intentions task, involving keeping track of the protagonist's intentions and thoughts in a false belief context. These established ToM tasks were chosen not only to ensure sensitivity to developmental change in ToM but also to allow for the inclusion of a matched number of self- and other-focused ToM questions, achieved by asking the children to reflect on their own mental states or the mental states of the "other" (represented by a hand puppet named "Teddy" or the experimenter, depending on the task). Previous research suggests that reasoning about own past mental states shares a developmental timeline with reasoning about others' mental states (neither of which can be solved by introspection on own current belief) (Bender et al., 2011; Fabricius & Khalil, 2003; Grosso et al., 2019; Müller et al., 2005; Razza & Blair, 2009; Russell et al., 2001; Wellman et al., 2001; Williams & Happé, 2010). However, we nonetheless considered it important to consider these question types separately, creating an equivalent scale for each, given that self-referent ToM questions might be considered a higher-level form of metacognition due to overlap in the definition of these key variables. There were five tasks in the battery, as described below. Two tasks had one question for self and one for other, and the remaining three tasks had two questions for self and two for other. This resulted in a maximum of 8 ToM points for self and other, summing to a total of 16 points.

Diverse belief task (1 point self, 1 point other). In this task, children were shown two pictures representing hiding places (e.g., a bush and a tree) and were told that an animal (e.g., a cat) was hiding behind one of them (the hiding locations and animals were varied across phases to avoid practice effects). Children were then asked to point to the picture they thought the cat was hiding behind. Once they did this, a puppet called Teddy joined the game. When the experimenter asked Teddy where he thought the cat was, Teddy would always respond with the opposite location to the one selected by the children. The experimenter then asked a self-focused test question (e.g., "You think the cat is hiding in the bush, so where would you look for the cat?") and an other-focused test question (e.g., "Teddy thinks the cat is hiding in the tree, so where will Teddy look for the cat?"). The order of test questions was counterbalanced between children. This task assesses children's ability to predict how they and others will act in accordance with their beliefs.

Knowledge access task (2 points self, 2 points other). In this task, children were presented with a small plastic box with a small toy figure hidden inside (the appearance of the box and its contents varied across phases to avoid practice effects). The experimenter asked the children if they knew what was in the box (target answer: no) and then how they might find out (target answer: look inside). The box was then opened and the toy figure was revealed. The figure was then put inside the box and the lid placed back to hide the box's contents once again. At this point, Teddy returned to play the game. The experimenter explained that the children had seen inside the box and so knew its contents, but Teddy had not seen inside the box. Children were then asked whether Teddy knew what was inside the box (target answer: no) and then how he might find out (target answer: look inside/ask). The self- and other-focused questions in this task assessed the children's understanding of knowledge–ignorance and acting on this basis. Self and other questions were not directly matched given that the children knew what was inside the box when asked the test questions for other. This sequence was used in the original scale to ensure that the children's answer for other was based on an understanding of other's mental states rather than representing a report of their own.

Unexpected contents task (2 points self, 2 points other). To instill confidence in their guesses (especially given repeated testing across phases), children started the task with a control condition in which they were shown canonical containers (e.g., egg boxes, DVD cases, branded sweets) and were asked what they thought might be inside. The contents were then revealed to be as expected. Where treats were revealed and parental permission was in place, children were given a treat. Mirroring test trials, the contents were then replaced in the containers and the children were asked what they had thought was inside. However, in test trials children were shown a branded packet of treats (treats varied across phases) and were asked what they thought might be inside the box. The experimenter then excitedly opened the packet to find out if there was really a treat inside, only to express disappointment on revealing that the packet contained something unattractive ("Those aren't sweets! Those are pencil shavings. Boo!"). After placing the shavings back inside the packet, children were asked what they thought was inside the packet before it was opened. Children who answered on the basis of their past belief passed this self-focused question. Children were also asked how they felt before and after finding out what was inside the packet, with a happy face and a sad face for reference. Children who reported that their emotions changed from positive to negative/less positive passed this question. Once the self-focused questions were completed, Teddy joined the game, indicating excitement upon seeing the treat container. Children were asked what Teddy thought was in the container. Teddy then excitedly opened the box, only to express disappointment upon finding the unattractive contents. Children were then asked how Teddy felt before and after opening the packet. Again, self- and other-focused questions were not matched or counterbalanced here given that it is necessary for Teddy's beliefs and the children's beliefs to differ at the time of the test questions. Nonetheless, both self- and other-focused questions assessed an understanding of false beliefs and "false" feelings.

Transparent intentions task (2 points self, 2 points other). This task involved a set of line drawings presented on transparent acetate slides. When the slides overlapped, they appeared to create one image (e.g., a mouse, a robot), with a missing feature (drawings were varied across phases to avoid practice effects). The experimenter pointed out the missing feature (e.g., the mouse had no tail, the robot had no ear). On self-focused trials, children were invited to add the missing feature to the drawing ("Can you draw the mouse a tail?", "Can you draw the robot an ear?"); on other-focused trials, the experimenter stated this intention (e.g., "I will draw the mouse a tail," "I will draw the robot an ear") and proceeded to add the feature to the drawing. However, when the top acetate was lifted from the bottom, the picture was transformed (e.g., the mouse's body, removed from context, became the shape of a balloon; the robot's head, removed from context became the shape of a mug). This meant that the added feature changed interpretation (e.g., the "tail" looked like the string of the balloon, the "ear" looked like the handle of a mug). On self-focused trials, children were asked "Did you think you were drawing a ['tail on a mouse'] or ['a balloon string']?" and "Did you mean to draw a ['tail on a mouse'] or a ['balloon string']?" (order counterbalanced). On other-focused trials, the questions were phrased to reference the experimenter, that is, "Did I mean . . .?" and "Did I think . . ."? This task allowed children to demonstrate an understanding of false intentions (what they or the experimenter meant to draw) and false thoughts (what they or the experimenter thought they were drawing). In this task, self- and other-focused questions were closely matched in context and could be counterbalanced for order and drawing content.

Hidden emotions task (1 point self, 1 point other). Children listened to a story accompanied by simple drawings about a protagonist who was excited to receive a gift, but who opened the gift to discover it was disappointing. Rather than upset the gift giver, the protagonist pretended to like the gift. Children were asked what gift the protagonist received (control). Children were asked how the protagonist felt when he opened the gift (target: sad) and how he looked (target: happy), with a happy face and a sad face for reference. Children were was also asked to imagine how they would feel if they got a present they did not like (target: sad) and how they would try to look if they wanted to hide that from the gift giver (target: happy). The order of self and other questions was counterbalanced, and children received 1 point for noting the mismatch of emotions for self and 1 point for noting the mismatch of emotions for other.

Metacognition task

To measure children's capacity to monitor their certainty, participants took part in a perceptual discrimination task with a certainty monitoring component adapted from Lyons and Ghetti (2013). In this task, children needed to identify a target object, animal or fauna, from two digitally pixelated line drawings (taken from Cycowicz et al., 1997) presented on a laptop using DMDX coding software (Forster & Forster, 2003). The pictures were standardized for familiarity and pixelated to 80% degradation. At the start of each test trial, the children would see instructions to find the target item (e.g., "Point to the duck") that were also vocalized by the researcher. A pair of images were then shown (see Fig. 1A), and the children were encouraged to guess which picture showed the target (location counterbalanced). Accuracy was recorded by the software. Each test trial was followed by a confidence judgment screen. On this screen, children saw two pictures of a girl or boy (whichever matched their gender; see Fig. 1B for female stimuli) depicting sure and unsure facial reactions (location counterbalanced). Children were asked to indicate, by pointing to the model, whether they were "sure" or "not so sure" they had found the target.

Children completed four practice trials before beginning to ensure that they understood the task and response set required. Unlike Lyons and Ghetti (2013), we did not use extensive training trials focusing the children on the speed of making their guess because we wanted to assess the children's natural capacities to monitor certainty. There were 10 test trials in total. Although half the length of Lyons and Ghetti's (2011, 2013) battery, piloting indicated that this was sufficient to elicit metacognitive judgments in our age range, and we were keen to minimize total session length to ensure the comfort of our young sample. Following Lyons and Ghetti (2011, 2013) and Feurer et al. (2015), we calculated the proportion of sure trials on which the children had accurately selected the target and subtracted from this the proportion of sure trials on which they were inaccurate to provide a measure of metacognition. Successful certainty monitors should be more accurate when sure, and thus higher certainty difference scores indicate superior metacognition. Note that uncertainty scores were not considered in this calculation given that the chances of selecting the correct foil by chance, even if uncertain, were high. Children were not given feedback on their performance, minimizing the chance that lucky guesses might inflate their metacognitive judgments. An advantage of this "easy" task is that it is engaging and fun for children and that baseline performance can be held developmentally stable across 3 to 6 years of age (Lyons & Ghetti, 2013), meaning that the main task requirement was focused on the ability to introspect and report on one's own certainty. This contrasts with some previous studies that have asked children to report on their confidence in free recall, potentially confounding developments in metacognition and memory. See Roebers et al. (2021) for full review of the advantages and disadvantages of available developmental measures of metacognition.

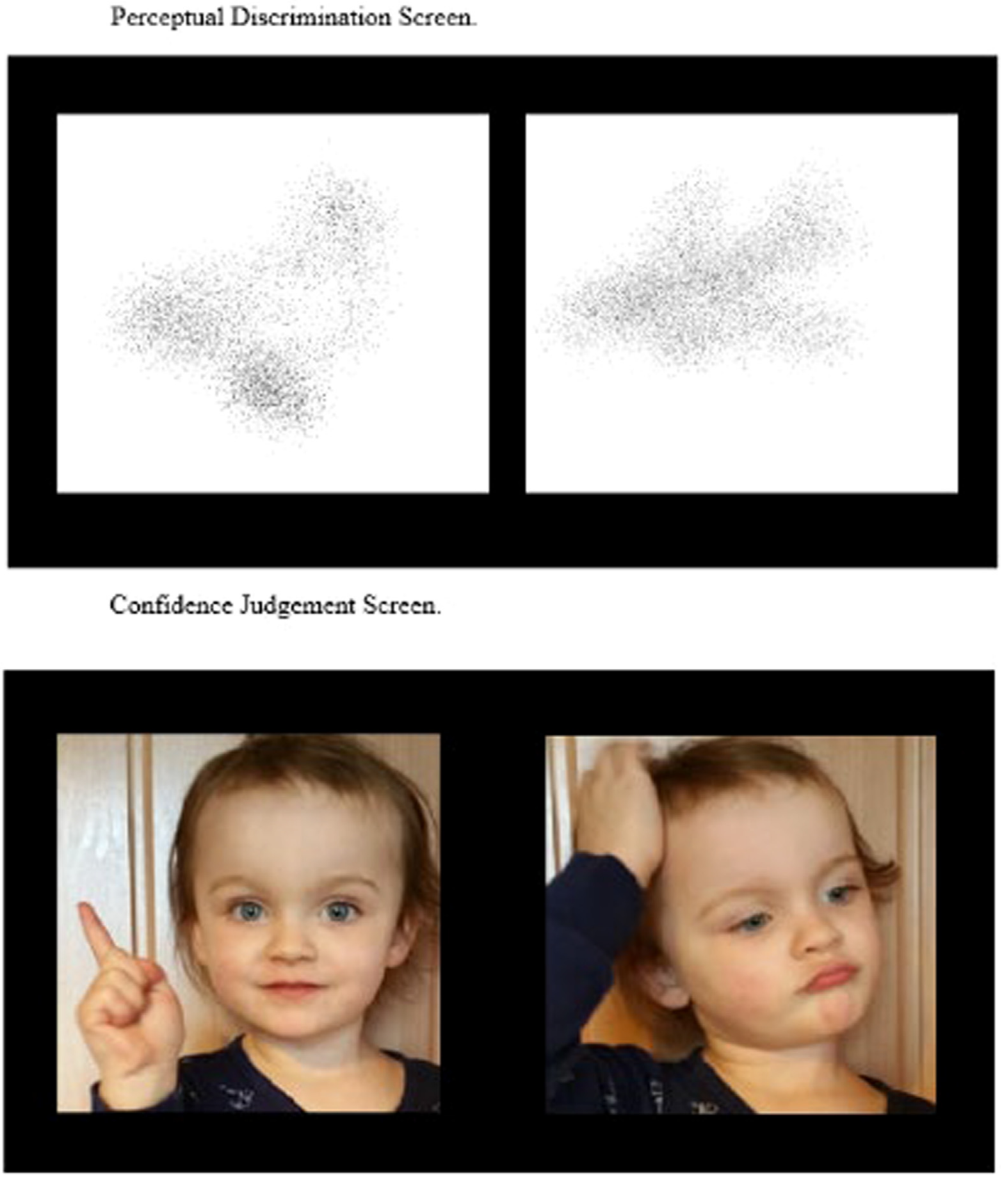

Fig. 1. Metacognition task stimuli. (A) Perceptual discrimination screen: An example of pictures presented during the test trial (left: duck; right: airplane). (B) Confidence judgment screen (left: sure; right: unsure).

Results

Developmental improvements in metacognition, inhibitory control, and ToM

Children's baseline task performance on the perceptual certainty monitoring task remained consistent across phases [number of trials correct: M_{Phase1} = 6.7, M_{Phase2} = 7.2, M_{Phase3} = 7.0, $F(2, 102)$ = 2.64,

$p = .076$, $\eta^2 = .05$; use of sure response: $M_{\text{Phase1}} = 7.4$, $M_{\text{Phase2}} = 8.0$, $M_{\text{Phase3}} = 7.8$, $F(2, 102) = 1.306$, $p = .275$, $\eta^2 = .025$], confirming that baseline task accuracy and confidence levels remained developmentally stable. Fig. 2 shows children's metacognition, ToM, and inhibitory control scores across the three phases. ToM and inhibitory control scores are corrected (proportion correct − proportion incorrect) to allow performance to be mapped comparably with the metacognition task (proportion sure and correct − proportion sure and incorrect). As shown in Fig. 2, there was substantial variation in children's performance. It is difficult to determine a "true" measure of chance given the possibility of systematic biases in ToM tasks and partially correct responses on the Bear/Dragon task. However, the corrected proportions at least allow for the children's performance across tasks to be judged comparably against the standard of expecting more correct responses than incorrect responses. As reflected in Fig. 2, one-sample *t* tests (with chance set at 0) suggested that as a group children were first above chance in metacognition in Phase 1, $t(51) = 7.72$, $p < .001$, inhibitory control in Phase 2, $t(51) = 8.73$, $p < .001$, and ToM in Phase 3 [self: $t(51) = 2.08$, $p = .042$; other: $t(51) = 3.30$, $p = .002$]. As shown in Fig. 2, rather than beginning at chance for self and other ToM, children moved from below chance performance in Phase 1 to above chance performance in Phase 3; this likely reflects a tendency to systematically err by answering on the basis of current reality in false belief tasks. Paired-samples *t* tests indicated no significant differences between self- and other-related ToM scores within phases [Phase 1: $t(51) = 0.28$, $p = .78$; Phase 2: $t(51) = 0.605$, $p = .55$; Phase 3: $t(51) = -1.79$, $p = .080$] (see Part A in online supplementary material for comparison of pass rates for self and other split by task). This suggests that self- and other-related ToM questions may measure the same developmental skill set.

In Phase 1, 80% of children passed the metacognition task, 52% passed the inhibitory control task, and 19% passed the ToM tasks. Considering individual developmental trajectories, Table 1 shows that the majority of children first exceeded the chance threshold of zero for metacognition or for metacognition and inhibitory control in the same phase. Children tended to pass the metacognition and inhibitory control tasks simultaneously (61%). However, a substantial minority of children (33%) achieved metacognition before inhibitory control, and only 6% achieved inhibitory control before metacognition.

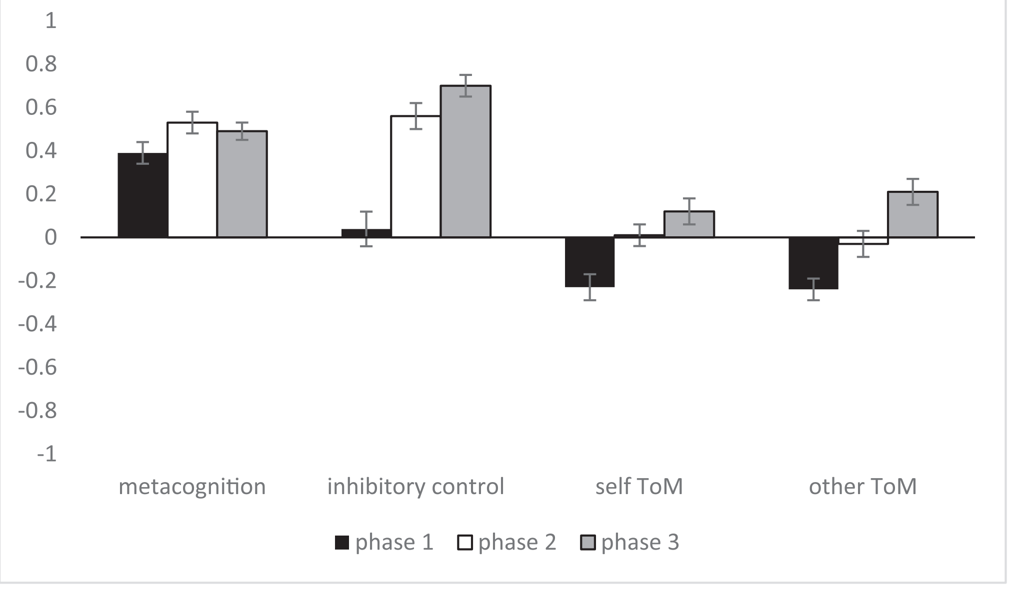

Fig. 2. Corrected scores (and standard errors) for metacognition, theory of mind (ToM), and inhibitory control scores across Phases 1 to 3.

Table 1
First pass sequence for metacognition, inhibitory control, and ToM.

Metacognition	–	+	+	+	+	+	–	–	–
Inhibitory control	–	–	+	+	+	+	+	+	+
Self ToM	–	–	–	+	–	+	–	+	–
Other ToM	–	–	–	+	+	–	–	–	+
Total	0	17	20	8	2	2	1	1	1

Note. TOM, theory of mind; + denotes a pass; – denotes a fail.

No child exceeded the pass threshold for ToM without also exceeding the chance threshold for inhibitory control, and only 6% achieved ToM before metacognition.

Overall, 94% of children had metacognition in their first pass profile, 62% had inhibitory control, and 25% had ToM. Fully 94% of children fit a sequence scaling metacognition to inhibitory control to ToM (first six columns of Table 1), and Green's (1956) coefficient of replicability suggested that this scaling provided a very good fit for the data (*Replicability* = .99). When children did pass the ToM battery, there was no evidence of primacy for passing self- or other-related ToM questions, with 27% of children passing other first, 19% passing for self first, 33% passing for both in the same phase, and 21% passing neither, McNemar test $\chi^2(1) = 0.375$, $p = .54$.

These data suggest that metacognitive monitoring and inhibitory control are likely to precede ToM in development regardless of whether questions are self- or other-focused. However, because our ToM battery was scaled for difficulty, we also checked the developmental trajectory for individual ToM tasks (see Part B in supplementary material). Like Wellman and Liu (2004), we found that children developed an understanding of diverse beliefs before more complex aspects of ToM for both self- and other-focused questions. Although 67% of children passed the Diverse Belief task for self and other in Phase 1, children did not perform above chance in any of the other tasks until later phases. This confirms that robust evidence of ToM, as measured by an understanding of false belief, comes after the development of basic metacognition.

A repeated-measures analysis of variance (ANOVA) confirmed that whereas there was no main effect of phase for metacognition, $F(2, 102) = 2.78$, $p = .067$, $\eta^2 = .052$ (see Fig. 2 for means), there was a main effect of phase for inhibitory control, $F(2, 102) = 50.73$, $p < .001$, $\eta^2 = .60$, and ToM [other: $F(2, 102) = 27.09$, $p < .001$, $\eta^2 = .35$; self: $F(2, 102) = 15.12$, $p < .001$, $\eta^2 = .23$] (see Fig. 3 for means of uncorrected scores, used here and in all subsequent analyses to aid comparability with extant literature). Post hoc Bonferroni comparisons confirmed that children made significant improvement across Phase 1 to Phase 2 for inhibitory control ($p < .001$), stabilizing across Phase 2 to Phase 3 ($p = .110$). For other-related ToM questions, there was improvement across Phase 1 to Phases 2 and 3 ($p < .002$) and across Phase 2 to Phase 3 ($p < .001$), whereas for self-related ToM questions, the improvement was significant only across Phase 1 to Phases 2 and 3 ($p < .001$). These analyses mirror the emergence pattern, confirming that metacognition is established early (Phase 1), whereas developments in inhibitory control (peaking in Phase 2) precede developments in ToM (peaking in Phase 3).

To model individual developmental growth in metacognition, inhibitory control, and self- and other-focused ToM, we used multilevel linear analyses to assess random effects of individual and fixed effects of phase. The best-fitting growth curve was linear (with phase coded as linear time 0, 1, or 2). Repeated covariance parameters were set to scaled identity. Metacognition did not significantly vary at intercept, $\mathrm{Var}(\mu_{0i}) = .021p = .059$, suggesting that all of the children had similar metacognitive abilities at baseline, and there was no fixed effect of phase on metacognition, $F(1, 103) = 2.70$, $p = .103$. These analyses confirm that metacognition was stable across phases. On the contrary, inhibitory control did significantly vary at intercept, $\mathrm{Var}(\mu_{0i}) = .025$, $p = .001$, and by phase, $F(1, 103) = 83.22$, $p < .001$. Each increment in phase was associated with a 16% increase in inhibitory control. Likewise, children's baseline ToM significantly varied at intercept [self: $\mathrm{Var}(\mu_{0i}) = .013$, $p = .004$; other: $\mathrm{Var}(\mu_{0i}) = .024$; $p < .001$] and by phase [self: $F(1, 103) = 28.64$, $p < .001$; other: $F(1, 103) = 54.61$, $p < .001$]. Each increment in phase was associated with ~ 10% improvement in ToM (self: 9%; other: 11%). These analyses confirm that whereas metacognition leveled off across phases, inhibitory control grew relatively

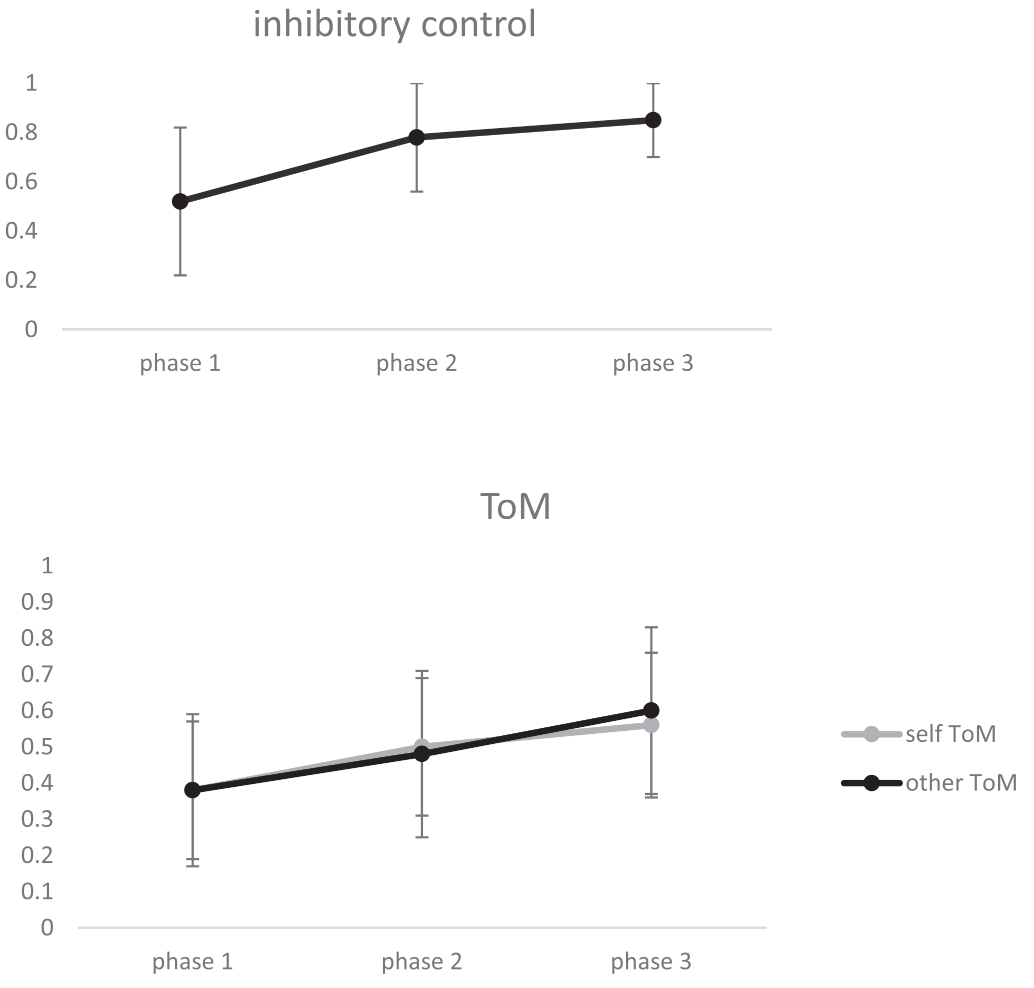

Fig. 3. Average proportions correct (and standard deviations) in inhibitory control and theory of mind (ToM) across Phases 1 to 3.

rapidly and ToM grew relatively slowly over the period of study. Rates of change for self- and other-focused ToM were similar. None of the models was improved by adding the random effects of time or the fixed effects of age.

Developmental interrelations among metacognition, inhibitory control, and ToM

Table 2 shows Pearson's correlations among metacognition, inhibitory control, and ToM within and across the three time points and with age in months. Benjamini–Hochberg corrections were applied to account for multiple comparisons. For brevity, age in months is shown for Phase 1, but the same pattern of significance held if substituted for age at Phase 2 or 3. Only inhibitory control was positively related to age, only reaching significance in Phase 2, suggesting that children who were older at the outset of the study were likely to have higher inhibitory control scores when this skill set first became established. Individual differences in metacognition and ToM were not related to age. This might be expected for metacognition given that this skill was already established at Phase 1. ToM did undergo

Table 2
Zero-order correlations (*r*) for metacognition, inhibitory control, theory of mind, age in months, and receptive language.

	Meta 2	Meta 3	Inh-Ctrl 1	Inh-Ctrl 2	Inh-Ctrl 3	Self ToM 1	Self ToM 2	Self ToM 3	Other ToM 1	Other ToM 2	Other ToM 3	Age 1 (mths)	BPVS 1	BPVS 2	BPVS 3
Meta 1	.08	.42*	.39*	.22	.08	.38*	.15	.28*,†	.32*	.06	.38*	.11	.07	.27	.17
Meta 2		.12	.19	.10	.33*	.21	.18	−.01	−.03	.07	.05	.10	.05	.13	.02
Meta 3			.39*	.13	.01	.36*	.12	.18	.05	.04	.12	−.01	.27	.09	.05
Inh-Ctrl 1				.59**	.43**	.49**	.46**	.53**	.49**	.43**	.48**	.26	.45**	.36*	.45**
Inh-Ctrl 2					.40*	.27	.51**	.44**	.28*,†	.43**	.44**	.37*	.34*	.43*	.42*
Inh-Ctrl 3						.22	.46**	.47**	.23	.32*	.29*,†	.26	.36*	.24	.34*
Self ToM 1							.33*	.17	.68**	.35*	.12	−.001	.43*	.06	.33*
Self ToM 2								.49**	.41*	.37*	.39*	.27	.38*	.36*	.37*
Self ToM 3									.31*	.46**	.62**	.20	.26	.44**	.49**
Other ToM 1										.52**	.41*	.05	.40*	.10	.50**
Other ToM 2											.55**	.002	.26*	.29*	.51**
Other ToM 3												−.02	.16	.45**	.44**
Age 1 (mths)													.22	.15	.17

Note. Meta, metacognition; Inh-Ctrl, inhibitory control; ToM, theory of mind; BPVS, British Picture Vocabulary Scale; mths, months. Phases 1, 2, 3.
* $p < .05$, †$p > .05$ when the Benjamini–Hochberg correction was applied.
** $p < .001$.

developmental change across the period of study. However, ToM did not exceed chance until Phase 3, at which point all children were 4 years of age or over and the amount of variance explained by age may have been low. Given the longitudinal structure of the dataset, we consider phase to be a clearer indicator of developmental growth.

Focusing on the developmental relationship between our key variables, metacognition, inhibitory control, and ToM (both self and other) all were positively related within Phase 1. Thereafter, only inhibitory control and ToM consistently were related within phases. Self- and other-related ToM questions were related within Phases 1, 2, and 3. Looking across phases, metacognition in Phase 1 and inhibitory control in Phases 1 and 2 were positively related to both self and other ToM in Phase 3. There was a small, nonsignificant positive correlation between metacognition in Phase 1 and inhibitory control in Phase 2. However, there was a significant moderate correlation between metacognition in Phase 2 and inhibitory control in Phase 3. Alongside the pattern of developmental emergence, the correlations in Table 2 support the hypothesis that interrelated early developments in metacognition and inhibitory control may be related to later developments in ToM.

Note that although children's metacognition scores were related across Phases 1 and 3, metacognition at Phase 2 was not related to previous and subsequent performance. This might suggest fluctuations in metacognition performance at an individual level. However, given that our multilevel analyses suggest that all children performed similarly well in metacognition at the outset, and there was no significant impact of phase on performance, this lack of correlation is most likely to arise from noise within variable resulting in relatively small fluctuations in performance across phases. This lack of variance in the metacognition measure may also have compromised our ability to capture the links between metacognition and the other variables consistently despite evidence of relationships among all three variables in Phase 1 and a time lagged effect of Phase 1 metacognition on Phase 3 ToM. Interpretation of these zero-order correlations is further complicated by the possibility that the connection between metacognition and ToM depends on (a) a sufficient level of inhibitory control to have developed at the individual level to set one's own perspective aside when answering explicit ToM questions ("traditional" simulation accounts) and/or (b) the development of explicit metacognition opening the possibility of reflection on mental states, but not necessarily being involved in solving individual ToM problems thereafter (our novel developmental simulation account). To capture this complexity, the next step in our analyses moved beyond zero-order correlations to employ linear mixed-model and mediation analysis to predict development of ToM.

Children's receptive language, considered as a potential control variable, was related to inhibitory control and self and other ToM but not to metacognition or age. Relationships were generally strongest for current measurements and were not always significant when relating earlier language levels to later cognitive skills.

Predicting developmental change in ToM

To predict developmental change in ToM while properly accounting for interclass correlation (ICC), we used a linear mixed model to assess random effects of individual and time and fixed effects of phase, inhibitory control, receptive language, and metacognition, using the restricted maximum likelihood ratio. Repeated covariance parameters were set to scaled identity, and random covariance parameters were set to autoregressive heterogeneous. To minimize degrees of freedom, maximize power, and provide a measure of ToM encompassing its traditional definition (reasoning about own and other minds), self and other ToM scores were averaged here to create an overall score (Phase 1: $M = .38$, $SD = .18$; Phase 2: $M = .49$, $SD = .18$; Phase 3: $M = .58$, $SD = .19$). However, separate analyses for self and other ToM found a similar pattern of predictive variance and can be found in Part C of the supplementary material.

The best-fitting growth curve was linear, with a predicted 4% increment in individual ToM scores across each phase independent of the measured predictors, $F(1, 71.86) = 5.727$, $p = .019$. However, ToM varied significantly at intercept, $\text{Var}(\mu_{0i}) = .014$, confirming that individual children had different ToM abilities at baseline ($p = .006$). There was also significant variance, $\text{Var}(\mu_{1i}) = .007$, in the slope showing the rate of change in ToM over time ($p = .014$). Finally, the covariance between the slopes and intercepts, $\text{Cov}(\mu_{0i}, \mu_{1i}) = -.64$, was significant ($p < .001$), suggesting that children with lower baseline ToM

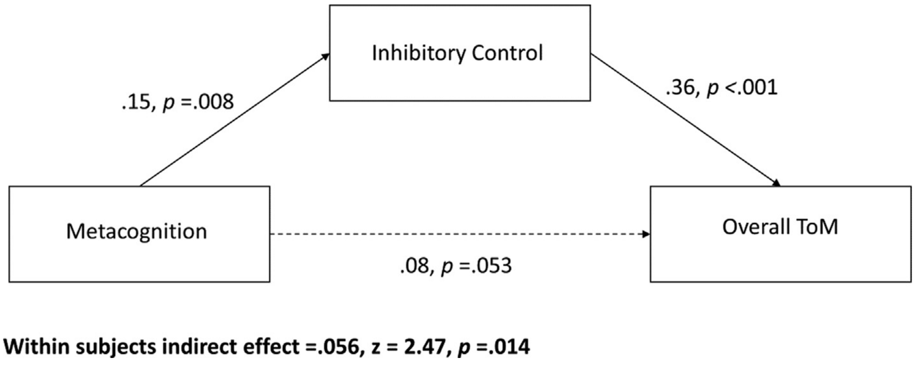

Fig. 4. Within-participants indirect effects of metacognition on theory of mind (ToM) through inhibitory control as assessed by multilevel mediation.

scores showed greater improvement over time. Inhibitory control and metacognition also made significant contributions to the model. Each individual increment in inhibitory control was associated with a 23% increase in ToM ability, $F(1, 119.731) = 17.96$, $p < .001$, whereas each individual increment in metacognition was associated with a 10% improvement in ToM, $F(1, 101.04) = 7.81$, $p = .006$. Receptive language also made a significant contribution to the model, although accounting for only a 2% increase in ToM ability, $F(1, 105.934) = 7.78$, $p = .006$. This analysis confirms that although inhibitory control, metacognition, and language can be used to independently predict developmental growth in ToM, developmental accelerations in inhibitory control appear to be particularly strongly associated with developments in ToM.

Finally, guided by the emergence patterns observed and the knowledge that inhibitory control may be required for the expression of ToM, we tested the possibility that the development of inhibitory control might statistically mediate the link between metacognition and ToM at a group level, consistent with our novel developmental perspective on simulation theory. At an individual level, we had observed zero-order correlations between metacognition at Phase 1 and inhibition at Phase 2, both of which related to self and other ToM at Phase 3. Indeed, inhibition was related to ToM within and between all chronologically subsequent phases, implying a strong developmental connection. We used Hayes and Rockwood's (2020) multilevel mediation macro to test a 1-1-1 within-participants mediation model (metacognition > inhibitory control > ToM) with phase as the Level 2 cluster variable, including random intercepts. With an ICC of.30, power to detect a medium effect size was at 80% (Pan et al., 2018). Confidence intervals were based on the product of 10,000 Monte Carlo simulations. Results are shown in Fig. 4. We found clear evidence of developmental mediation at a group level, such that metacognition affected ToM development only through the mediation of inhibitory control.

Discussion

This longitudinal study tracked the developmental relations linking metacognition, ToM, and inhibitory control across 3 to 5 years of age. As operationalized here, metacognition and inhibitory control emerged before ToM, and these skills were independently predictive of growth in ToM competence. However, there was evidence of developmental mediation over time, whereby metacognition predicted inhibitory control, which predicted ToM. Thus, we provided evidence to support our hypotheses that metacognitive self-reflection may provide the "developmental enrichment" necessary to think about thinking, and when inhibitory control is sufficiently developed, this thinking can be extended to complex reasoning about own and other minds.

Emergence of metacognition, ToM, and inhibitory control

Children's ability to monitor their own certainty was above chance at the outset (M_{age} = 3.9 years) and did not show individual or time-based variation, suggesting that explicit metacognitive monitoring was well-established in our focal age range. On the contrary, inhibitory control performance for our sample first exceeded chance in the second testing phase (M_{age} = 4.4 years) and increased rapidly across the 12-month period sampled. Although the majority (61%) of children passed metacognition and inhibitory control tasks within the same phase, a substantial minority (33%) achieved metacognition before inhibitory control, and very few (6%) showed the opposite pattern. This pattern of results suggests that simpler forms of introspection may precede inhibitory control in development. There was significant individual variance in ToM at the outset, and ToM appeared to grow relatively slowly over the 12-month period sampled, as might be expected in a comprehensive task battery scaled for developmental difficulty. Both metacognition and inhibitory control clearly preceded ToM on this scale; very few children exceeded chance in ToM before metacognition (6%) or inhibitory control (0%), and the majority did not exceed chance performance in ToM until the final testing phase (M_{age} = 4.9 years). Although a simpler form of ToM indicating an understanding of diverse desires was passed by 67% of children in the first phase of testing, pass rates in the simplest task still lagged those of the metacognition task, passed by 80% of children at the outset. Moreover, pass rates for the false belief task—typically considered the litmus test for the onset of flexible reasoning about other minds—did not exceed chance until Phase 3 of testing. This pattern strongly supports the idea that reliably passing false belief tasks is not possible until children have developed sufficient introspection and inhibitory control. This is important given that false belief tasks are considered the litmus test for flexible reasoning about mental states.

Explicit metacognition and ToM appear to emerge as independent systems. The order of emergence supports an independent process developmental simulation view, where a mentalistic understanding of self and inhibitory control are required to reason flexibly about other minds. This emergence pattern undermines theory-theory views, which predict that ToM will emerge simultaneously with metacognition or act as the developmental foundation to a mentalistic understanding of self. However, it should be noted that the ability to reason about own past mental states was confirmed to share a developmental timeline with reasoning about others' mental states. If focused on a more complex form of metacognition, not reliant on introspection of current mental state, one could therefore argue for a single process view. Again, this highlights the complexity inherent in developmental research, whereby definition and measurement of focal skill sets influences conclusions concerning their trajectory. Indeed, there are active debates concerning the extent to which alternative measurements of metacognition, inhibitory control, and ToM are valid in infancy (see Goupil & Kouider, 2019, Hendry et al., 2022, and Rakoczy, 2022, for discussion). Thus, although we can provide some evidence for the primacy of explicit metacognitive monitoring over explicit forms of inhibitory control and ToM in early childhood, we recognize that different methods and manifestations might offer alternative conclusions about developmental emergence.

Developmental relations among metacognition, inhibitory control, and ToM

Early metacognition (M_{age} = 3.9 years) and inhibitory control (M_{age} = 4.4 years) were related to the later development of ToM (M_{age} = 4.9 years). There were strong within-phase relations between inhibitory control and ToM. On the contrary, the concurrent relations between metacognition and inhibitory control were not consistent. This pattern might undermine the idea that children routinely solve ToM tasks by simulating how they would think in another person's shoes. However, the requirement for inhibitory control to be sufficiently developed to express ToM may have compromised this relation. Inhibition was a very strong predictor of ToM, supporting combined emergence and expression accounts (Kloo et al., 2020). However, although metacognition could also be used as an independent predictor of ToM, we found evidence of developmental mediation, whereby inhibitory control mediated the link between metacognitive monitoring and ToM for individual children. Alongside the emergence patterns discussed above, this pattern supports our developmental proposal that the capacity to reflect on current mental states might be the first step to understanding mental states separately from actions. When inhibitory control is sufficiently developed to allow one to pause reflection on one's own mental states and turn this lens outward, this thinking may be extended to more complex reasoning about other minds.

These results are relevant to the historical debate between simulation and theory-theory (Gallese & Goldman, 1998; Goldman, 2006; Gopnik & Melzoff, 1997; Gopnik & Wellman, 1992; Harris, 1992), to developmental enrichment accounts of ToM (Kloo et al., 2020), and to emerging theories of the development of ToM. For example, Kampis and Southgate (2020) proposed that infants are born focused on others and that, rather than aiding social cognition, the emergence of self-reflection may interfere with our ability to understand other minds, at least until inhibitory control is sufficiently developed to set egocentrism aside (see also Yeung et al., 2022). This theory may explain why despite evidence of ToM in infancy, children routinely fail explicit ToM tests in early childhood. The mediational role for inhibitory control found here is consistent with this novel account. However, where present, the relations between metacognitive monitoring (self-reflection) and ToM were positive, complicating the idea that self-reflection may cause interference with social cognition.

Further longitudinal data are needed to explain the developmental relations uncovered here. There is ongoing debate concerning the extent to which implicit forms of metacognition, inhibitory control, and ToM lay the foundations for explicit forms of these skill sets (Goupil & Kouider, 2019; Hendry et al., 2022; Rakoczy, 2022) and for higher-order forms later in development (e.g., Ebert, 2015; Feurer et al., 2015; Kälin & Roebers, 2022; Lecce et al., 2014, 2015; Lockl & Schneider, 2006; Roebers et al., 2012). ToM skills are critical for social success, allowing children to predict and appropriately respond to others' behavior (Slaughter et al., 2015). Inhibitory control skills are similarly critical, allowing children to inhibit prepotent responses to follow socially and educationally salient rules (e.g., Best et al., 2011). Likewise, higher-order metacognition skills have been identified as pivotal to social and academic success, allowing children to express and reflect on their thought processes and to employ strategic thinking (e.g., rehearsing to-be-remembered information) (Whitebread & Neale, 2020; Roebers et al., 2012). However, if we are to fully understand the positive social and academic consequences of this powerful developmental triad, and effectively design interventions to support positive developmental outcomes, we must ultimately understand both their foundations and their intersections. For example, if children's ToM skills depend on their metacognitive and self-regulation capacities, then interventions aimed at these ''sister'' skill sets might ultimately be more effective than those directly aimed to support ToM. We found that receptive language was related to inhibitory control and ToM, although not to our measure of metacognition (cf. Ebert, 2015). Although receptive language predicted a relatively small proportion of variance in ToM relative to inhibitory control and metacognition, alternative measures of language development, including expressive language, may show stronger associations (see Milligan et al., 2007). Thus, language is likely to be an important part of the jigsaw when considering the optimal development of this cognitive triad (see Milligan et al., 2007).

Furthermore, it is crucial to consider the universality of developmental foundations and interrelations linking metacognition, inhibitory control, and ToM. It is well-established that developmental science is biased toward Western, educated, industrialized, rich, and democratic (WEIRD) populations (Nielsen et al., 2017), and this article is no exception to that bias. Our theoretical starting point and empirical data derive from a WEIRD perspective, where self-focus is salient in early childhood and beyond (Markus & Kitayama, 1991). Reflecting on this bias, Symeonidou et al. (2024) found that although metacognition and inhibitory control predict the development of ToM in a cross-sectional Scottish sample, these skills do not predict ToM in Japanese children. In this population, a focus on other is likely to be more salient than a focus on self, and as a result others' behavior is not likely to be predicted from an egocentric starting point. This example highlights how the relationships between developmental skill sets can be theoretically informative and supports the importance of socialization in driving cognitive development.

Conclusion

Despite the long history of simulation theory and the sustained drive to uncover the developmental factors that underpin ToM development, the role of metacognitive self-reflection in the ''developmental enrichment" of ToM has been little explored. We have provided novel data to support the proposal that, for Scottish children, the capacity to reflect on and control current mental states (as operationalized by metacognitive monitoring and inhibitory control, respectively) might be necessary first steps to understand others' and own past mental states.

CRediT authorship contribution statement

Mariel Symeonidou: Writing – review & editing, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Martin J. Doherty:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Josephine Ross:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Acknowledgments

This project was approved by the University of Dundee nonclinical research ethics committee and was made possible by funding received from the Experimental Psychology Society (560 808009) and a University of Dundee PhD studentship held by the first author. We thank the children and families who took part.

Data availability

Data are freely available on the Open Science Framework (https://osf.io/cqt72).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2024.106103.

References

Alcalá-López, D., Vogeley, K., Binkofski, F., & Bzdok, D. (2019). Building blocks of social cognition: Mirror, mentalize, share? *Cortex, 118*, 4–18. https://doi.org/10.1016/j.cortex.2018.05.006.

Bender, P. K., Pons, F., Harris, P. L., & de Rosnay, M. (2011). Do young children misunderstand their own emotions? *European Journal of Developmental Psychology, 8*(3), 331–348. https://doi.org/10.1080/17405629.2010.495615.

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences, 21*, 327–336. https://doi.org/10.1016/j.lindif.2011.01.007.

Carlson, S. M., Claxton, L. J., & Moses, L. J. (2015). The relation between executive function and theory of mind is more than skin deep. *Journal of Cognition and Development, 16*(1), 186–197. https://doi.org/10.1080/15248372.2013.824883.

Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences, 32*(2), 121–138. https://doi.org/10.1017/S0140525X09000545.

Carruthers, P. (2011). *The opacity of mind: An integrative theory of self-knowledge*. Oxford University Press.

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology, 65*(2), 171–237. https://doi.org/10.1006/jecp.1996.2356.

Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development, 85*(5), 1777–1794. https://doi.org/10.1111/cdev.12237.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale–Second Edition (BPVS-II)*. NFER–Nelson.

Ebert, S. (2015). Longitudinal relations between theory of mind and metacognition and the impact of language. *Journal of Cognition and Development, 16*(4), 559–586. https://doi.org/10.1080/15248372.2014.926272.

Fabricius, W. V., & Khalil, S. L. (2003). False beliefs or false positives? Limits on children's understanding of mental representation. *Journal of Cognition and Development, 4*(3), 239–262. https://doi.org/10.1207/S15327647JCD0403_01.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146.

Feurer, E., Sassu, R., Cimeli, P., & Roebers, C. (2015). Development of meta-representations: Procedural metacognition and the relationship to theory of mind. *Journal of Educational and Developmental Psychology, 5*(1), 6–18. https://doi.org/10.5539/jedp.v5n1p6.

Flavell, J. (2002). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology, 50*, 21–45. https://doi.org/10.1146/annurev.psych.50.1.21.

Flynn, E. (2007). The role of inhibitory control in false belief understanding. *Infant and Child Development, 16*, 53–69. https://doi.org/10.1002/icd.500.

Flynn, E., O'Malley, C., & Wood, D. (2004). A longitudinal, microgenetic study of the emergence of false belief understanding and inhibition skills. *Developmental Science, 7*(1), 103–115. https://doi.org/10.1111/j.1467-7687.2004.00326.x.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116–124. https://doi.org/10.3758/BF03195503.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(12), 493–501. https://doi.org/10.1016/S1364-6613(98)01262-5.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½ –7 years old on a Stroop-like day–night test. *Cognition, 53*(2), 129–153. https://doi.org/10.1016/0010-0277(94)90068-X.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford University Press.

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. MIT Press.

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language, 7*(1–2), 145–171. https://doi.org/10.1111/j.1468-0017.1992.tb00202.x.

Goupil, L., & Kouider, S. (2019). Developing a reflective mind: From core metacognition to explicit self-reflection. *Current Directions in Psychological Science, 28*(4), 403–408. https://doi.org/10.1177/0963721419848672.

Green, B. F. (1956). A method of scalogram analysis using summary statistics. *Psychometrika, 21*(1), 79–88. https://doi.org/10.1007/BF02289088.

Grosso, S. S., Schuwerk, T., Kaltefleiter, L. J., & Sodian, B. (2019). 33-month-old children succeed in a false belief task with reduced processing demands: A replication of Setoh et al. (2016). *Infant Behavior and Development, 54*, 151–155. https://doi.org/10.1016/j.infbeh.2018.09.012.

Harris, P. L. (1992). From simulation to folk psychology: The case for development. *Mind & Language, 7*, 120–144. https://doi.org/10.1111/j.1468-0017.1992.tb00201.x.

Hayes, A. F., & Rockwood, N. J. (2020). Conditional process analysis: Concepts, computation, and advances in the modeling of the contingencies of mechanisms. *American Behavioral Scientist, 64*(1), 19–54. https://doi.org/10.1177/0002764219859633.

Hendry, A., Greenhalgh, I., Bailey, R., Fiske, A., Dvergsdal, H., & Holmboe, K. (2022). Development of directed global inhibition, competitive inhibition and behavioural inhibition during the transition between infancy and toddlerhood. *Developmental Science, 25*(5), e13193. https://doi.org/10.1111/desc.13193.

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology, 43*(6), 1447–1459. https://doi.org/10.1037/0012-1649.43.6.1447.

Kälin, S., & Roebers, C. M. (2022). Longitudinal associations between executive functions and metacognitive monitoring in 5- to 8-year-olds. *Metacognition and Learning, 17*, 1079–1095. https://doi.org/10.1007/s11409-022-09306-x.

Kampis, D., & Southgate, V. (2020). Altercentric cognition: How others influence our cognitive processing. *Trends in Cognitive Sciences, 24*(11), 945–959. https://doi.org/10.1016/j.tics.2020.09.003.

Kloo, D., Kristen-Antonow, S., & Sodian, B. (2020). Progressing from an implicit to an explicit false belief understanding: A matter of executive control? *International Journal of Behavioral Development, 44*(2), 107–115. https://doi.org/10.1177/0165025419850901.

Kloo, D., Sodian, B., Kristen-Antonow, S., Kim, S., & Paulus, M. (2021). Knowing minds: Linking early perspective taking and later metacognitive insight. *British Journal of Developmental Psychology, 39*, 39–53. https://doi.org/10.1111/bjdp.12359.

Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science, 9*(5), 178–181. https://doi.org/10.1111/1467-8721.00088.

Lecce, S., Caputi, M., & Pagnin, A. (2014). False-belief understanding at age 5 predicts beliefs about learning in Year 3 of primary school. *European Journal of Developmental Psychology, 12*(1), 40–53. https://doi.org/10.1080/17405629.2014.949665.

Lecce, S., Demicheli, P., Zocchi, S., & Palladino, P. (2015). The origins of children's metamemory: The role of theory of mind. *Journal of Experimental Child Psychology, 131*, 56–72. https://doi.org/10.1016/j.jecp.2014.11.005.

Lockl, K., & Schneider, W. (2006). Precursors of metamemory in young children: The role of theory of mind and metacognitive vocabulary. *Metacognition and Learning, 1*(1), 15–31. https://doi.org/10.1007/s11409-006-6585-9.

Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development, 82*(6), 1778–1787. https://doi.org/10.1111/j.1467-8624.2011.01649.x.

Lyons, K. E., & Ghetti, S. (2013). I don't want to pick! Introspection on uncertainty supports early strategic behavior. *Child Development, 84*(2), 726–736. https://doi.org/10.1111/cdev.12004.

Marcovitch, S., O'Brien, M., Calkins, S. D., Leerkes, E. M., Weaver, J. M., & Levine, D. W. (2015). A longitudinal assessment of the relation between executive function and theory of mind at 3, 4, and 5 years. *Cognitive Development, 33*, 40–55. https://doi.org/10.1016/j.cogdev.2014.07.001.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*(2), 622–646. https://doi.org/10.1111/j.1467-8624.2007.01018.x.

Müller, U., Zelazo, P. D., & Imrisek, S. (2005). Executive function and children's understanding of false belief: How specific is the relation? *Cognitive Development, 20*(2), 173–189. https://doi.org/10.1016/j.cogdev.2004.12.004.

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology, 162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017.

Pan, H., Liu, S., Miao, D., & Yuan, Y. (2018). Sample size determination for mediation analysis of longitudinal data. *BMC Medical Research Methodology, 18*, 32. https://doi.org/10.1186/s12874-018-0473-2.

Perner, J. (1998). The meta-intentional nature of executive functions and theory of mind. In P. Carruthers & J. Boucher (Eds.), *Language and thought* (pp. 270–283). Cambridge University Press. https://doi.org/10.1017/CBO9780511597909.017.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 3*(9), 337–344. https://doi.org/10.1016/S1364-6613(99)01362-5.

Petersen, I. T., Hoyniak, C. P., McQuillan, M. E., Bates, J. E., & Staples, A. D. (2016). Measuring the development of inhibitory control: The challenge of heterotypic continuity. *Developmental Review, 40*, 25–71. https://doi.org/10.1016/j.dr.2016.02.001.
Proust, J. (2010). Metacognition. *Philosophy Compass, 5*(11), 989–998. https://doi.org/10.1111/j.1747-9991.2010.00340.x.
Rakoczy, H. (2022). Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology, 1*(4), 223–235. https://doi.org/10.1038/s44159-022-00037-z.
Razza, R. A., & Blair, C. (2009). Associations among false-belief understanding, executive function, and social competence: A longitudinal analysis. *Journal of Applied Developmental Psychology, 30*(3), 332–343.
Reed, M. A., Pien, D. L., & Rothbart, M. K. (1984). Inhibitory control in preschool children. *Merrill-Palmer Quarterly, 30*(2), 131–147 https://www.jstor.org/stable/23086229.
Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology, 33*(1), 12–21. https://doi.org/10.1037/0012-1649.33.1.12.
Roebers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review, 45*, 31–51. https://doi.org/10.1016/j.dr.2017.04.001.
Roebers, C. M., Cimeli, P., Röthlisberger, M., & Neuenschwander, R. (2012). Executive functioning, metacognition, and self-perceived competence in elementary school children: An explorative study on their interrelations and their role for school achievement. *Metacognition and Learning, 7*(3), 151–173. https://doi.org/10.1007/s11409-012-9089-9.
Roebers, C. M., van Loon, M. H., Buehler, F. J., Bayard, N. S., Steiner, M., & Aeschlimann, E. A. (2021). Exploring psychometric properties of children's metacognitive monitoring. *Acta Psychologica, 220*, 103399. https://doi.org/10.1016/j.actpsy.2021.103399.
Russell, J., Hill, E. L., & Franco, F. (2001). The role of belief veracity in understanding intentions-in-action: Preschool children's performance on the transparent intentions task. *Cognitive Development, 16*(3), 775–792. https://doi.org/10.1016/S0885-2014(01)00057-0.
Slaughter, V., Imuta, K., Peterson, C. C., & Henry, J. D. (2015). Meta-analysis of theory of mind and peer popularity in the preschool and early school years. *Child Development, 86*(4), 1159–1174. https://doi.org/10.1111/cdev.12372.
Sodian, B., Kristen-Antonow, S., & Kloo, D. (2020). How does children's theory of mind become explicit? A review of longitudinal findings. *Child Development Perspectives, 14*(3), 171–177. https://doi.org/10.1111/cdep.12381.
Symeonidou, M., Mizokawa, A., Kabaya, S., Doherty, M., & Ross, J. (2024). Contrasting one's share of the shared life space: Comparing the roles of metacognition and inhibitory control in the development of theory of mind among Scottish and Japanese children. *Developmental Science, 27*(5), e13417. https://doi.org/10.1111/desc.13417.
Veenman, M., Van Hout-Wolters, B., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*, 3–14. https://doi.org/10.1007/s11409-006-6893-0.
Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development, 72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304.
Wellman, H. M., & Liu, D. (2004). Scaling of theory of mind tasks. *Child Development, 75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x.
Whitebread, D., & Neale, D. (2020). Metacognition in early child development. *Translational Issues in Psychological Science, 6*(1), 8–14. https://doi.org/10.1037/tps0000223.
Williams, D., & Happé, F. (2010). Representing intentions in self and other: Studies of autism and typical development. *Developmental Science, 13*(2), 307–319. https://doi.org/10.1111/j.1467-7687.2009.00885.x.
Yeung, E., Askitis, D., Manea, V., & Southgate, V. (2022). Emerging self-representation presents a challenge when perspectives conflict. *Open Mind, 6*, 232–249. https://doi.org/10.1162/opmi_a_00065.