UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Knowledge First, Belief Later? Equalizing Processing Demands in Knowledge and Belief Attribution Tasks

Permalink

https://escholarship.org/uc/item/0t4712nk

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Kispál, Anna
Kovacs, Agnes

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

Knowledge First, Belief Later? Equalizing Processing Demands in Knowledge and Belief Attribution Tasks

Anna Kispál (kispal_anna@phd.ceu.edu)[1], Ágnes Melinda Kovács[1]
[1]Department of Cognitive Science, Central European University

Abstract

Studies show that children succeed in perceptual access-based knowledge attribution before false belief attribution. Recent theories suggest that this may be due to the primacy of knowledge representations. However, to more accurately contrast knowledge and belief attributions, closely matched test scenarios are needed that are comparable in processing demands. In the current study, we tested 4- to 5-year-old children (n = 50) on a novel paradigm requiring more complex knowledge attributions that involve more comparable processing demands — such as egocentric ignorance and self and other retrospective updates — in relation to false belief attributions. Results showed that children performed well on this more complex knowledge attribution, suggesting that knowledge ascription is robust even under increased cognitive demands. However, we found that complex knowledge and false belief attributions were negatively related, suggesting a surprising pattern in which false belief inferences may be overgeneralized early in development.

Keywords: knowledge; belief; mentalization; theory of mind; social cognition

Introduction

A great body of literature shows that children begin to pass explicit false belief tasks around the age of four. Less extensive evidence suggests that attributions of knowledge, traditionally often termed true belief, which usually serve as controls for false belief attribution, appear earlier in development (e.g., Wellman, Cross & Watson, 2001; Wellman & Liu, 2004; Wimmer & Perner, 1983; Yu & Wellman, 2024).

Recent knowledge-first proposals (Holland & Phillips, 2020; Nagel, 2017; Phillips et al., 2021; Phillips & Norby, 2021; Westra & Nagel, 2021) claim that, despite the long-standing tradition of treating knowledge as part of belief, knowledge representations do not depend on previous belief representations. Instead, they argue that knowledge is the most fundamental mental state and should be considered an equally important counterpart to belief attributions. The proposal draws on different sources of evidence, claiming that knowledge attributions (i) appear sooner in development than belief attributions (e.g., Luo & Johnson, 2009; Vouloumanos, Martin, & Onishi, 2014); (ii) happen in a more automatic manner compared to belief attributions (e.g. Surtees et al., 2016, but see Kovács, Téglás & Endress, 2010); (iii) are present in animals unlike belief attributions (e.g. Kaminski, Call, & Tomasello, 2008; Martin & Santos, 2016, but see Krupenye et al., 2016); (iv) and are preserved in patient populations even when belief attributions are not (Baron-Cohen & Goodhart, 1994; Schneider et al., 2013). The knowledge-first framework states that the capacity to attribute knowledge is not the same as attributing a true belief, as many cases of true belief, such as Gettier-cases (Gettier, 1963), fall short of knowledge attribution. The theory also suggests that knowledge is not modality-specific, meaning that one can gain knowledge not just via perceptual access but also through verbal reports or even logical inferences (but see Fogd, Téglás & Kovács, 2025 for belief attribution based on inference). Moreover, knowledge attributions could also appear in special cases, such as egocentric ignorance understanding, when another person knows something that one does not (Nagel, 2017; Phillips & Norby, 2021; Phillips et al., 2021). Despite the suggested variability in knowledge ascriptions, previous tasks assessing children's knowledge attributions have tended to focus on a special case of knowledge: knowledge based on joint perceptual access. Such studies have shown that children start to pass explicit knowledge attribution tasks around the age of 3 (e.g., Wellman, Cross, & Watson, 2001; Wellman & Liu, 2004), although some studies have found a U-shaped developmental curve, suggesting that between 3 and 10 years, children fail certain kinds of knowledge attribution tasks (Fabricius et al., 2010; Fabricius et al., 2021; Oktay-Gür & Rakoczy, 2017; Rakoczy & Oktay-Gür, 2020; Schidelko et al., 2022; Schidelko & Rakoczy, 2025).

Perceptual access-based knowledge and belief attributions differ from false belief attributions in (i) the conflict monitoring they involve and (ii) the nature of the inferences one must make. False belief attribution requires prospectively updating one's own knowledge, but also inferring that, because the other person did not witness a crucial event, they would maintain a representation of an earlier state of affairs. While tracking one's own and the other person's representations of the event, these two representations must be kept separate, and they also happen to be conflicting (I believe the object is in A, she believes the object is in B). In contrast, perceptual access-based knowledge attribution does not require such inferences or conflict monitoring; instead, one simply has to track whether another person has direct visual access to a shared event. Although different accounts suggest different proposals about the underlying representational structure of knowledge and false belief attributions (e.g., Martin & Santos, 2016; Phillips & Norby, 2018; Westra & Nagel, 2021), they would all agree that knowledge attributions are simpler and cognitively less demanding than false belief attributions because different and conflicting perspectives do not have to be maintained. This suggests that while previous research shows that knowledge attributions are easier, faster, and appear sooner in

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* ©2026 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

development, such results could be due to the low cognitive demands and lack of mental conflict monitoring that knowledge attribution situations introduce, and not due to knowledge representations being simpler *per se*.

Let's take a traditional change of location task as an example. In such tasks, children are introduced to a target object that is hidden in one out of two locations, for instance, in box A (and not in B). They witness this event together with an Agent. Then, in false belief conditions, the Agent leaves the room, and the location of the object is changed. The target is now in box B. At this point, children have to change two things in the way they represent the scenario. Firstly, they have to track the situation and prospectively update their own primary representation of reality (the target is in box B). Then, they need to separately track that the Agent is not aware of this (she still thinks it's in A). In case of a matching knowledge condition, the same swapping of the object would happen, but with the Agent being present. This would require only tracking the situation and prospectively updating the shared reality they have together with the Agent.

Knowledge attributions may be easier only under certain, low-demand scenarios, such as the ones applied in previous experiments, where knowledge depends on mere joint perceptual access. One way to determine whether knowledge ascriptions are easier than false belief attributions due either to the primacy of knowledge representations or to the comparatively lower cognitive demands of knowledge attributions would be to eliminate the additional processing demands and conflicts inherent in belief representations, thereby bringing knowledge and belief ascriptions to a comparable level. This would, however, defeat the nature of false belief scenarios themselves, as a mismatch between representations is a defining element of a false belief representation. If we cannot make knowledge and false belief scenarios equally demanding, we cannot study whether the differences in their ease and developmental primacy are rooted in the differing scenarios introducing differing cognitive demands or in the representations themselves. An alternative approach would be to introduce scenarios in which knowledge attributions unfold in a more complex manner, thereby equalizing the cognitive demands and conflict monitoring required across knowledge and false belief tasks. Such scenarios would also better reflect real-world situations, which frequently require the integration of more complex information.

In the present study, we aim to equalize the conflict demands, the inferential demands, and the update demands of knowledge and false belief mental state attributions (Table 1). We use scenarios in which knowledge is not gained through perceptual access, but instead through inference, and also involves egocentric ignorance. With the current study, we would like to test whether knowledge attributions that include multiple factors, which may resemble real-life situations more closely, are still easier for children to solve than false belief tasks.

We examine children aged 4-5 years in an interactive detective task, a population that certainly passes explicit knowledge attribution tasks and should, in great majority, pass explicit false belief tasks as well. We test whether four factors, (i) differing self and other representational contents, (ii) retrospectively updating one's own mental state, (iii) retrospectively updating the content of a partner's mental state, and (iv) conflicting mental states in the end, result in slower computation, regardless of the type of mental state attributed to the partner. In prospective updates, an already computed representation has to be abandoned (i.e., that the object is in A) when a new event occurs, such as an object being moved from box A to B, and a new one has to be formed (that the object is now in B). In retrospective updates, the content of one's own or the Agent's already attributed mental state has to be updated in light of new information (Király et al., 2018; Király, Oláh & Kovács, 2023). For instance, if one infers that an object should be in box A based on a clue, but then it turns out to be in box B (as the clue turned out to be misleading), a retrospective update of one's own (and the other's) previous mental state is needed. Different self and other representational contents mean that one's own representation and that of the Agent will differ, but this does not necessarily mean that they will conflict with each other. This could happen at any point during the mentalization process, and in our experiment, it will occur in scenarios involving egocentric ignorance (I don't know where the object is, but the Agent knows that it's in A or B) and false belief. Conflicting mental states in the end will also include differing representational contents, but these must be incompatible with each other (an object cannot be in both A and B), forming conflicting representations.

Table 1: Summary of the different cognitive demands present in the conditions.

	K-baseline	K-complex	FB-baseline	FB-complex
Different self & other representational contents	×	✓	✓	✓
Self retrospective update	×	✓	×	✓
Other retrospective update	×	✓	×	×
Conflicting mental states in the end	×	×	✓	✓

It is important to note that, although such demands, given that they are different (self retrospective updates, other retrospective updates, conflict, different representational contents), may not place an equal load on cognitive processing in all four conditions. However, at a minimum, they present more equilibrated situations than classical tasks. Traditional tasks involving a change of location will be used as a baseline. Two new conditions will be introduced in the following experiment. In the new K-complex condition, updating one's own mental state and retrospectively updating the content of a partner's mental state will be required in order to arrive at a correct mental state attribution. Moreover, because this trial will begin with an egocentric ignorance (child doesn't know, but Agent knows), the representation of different self and other representational contents will be needed. In the FB-complex condition, updating one's own mental state retrospectively and the conflict between one's own mental state and that of one's partner will be required. The effects of complexity and the type of mental state attribution can be studied by comparing matching conditions.

Methods

Participants

The sample consisted of 50 participants between the ages of 4 years 0 days to 5 years 365 days (28 boys, 22 girls; Mean age = 59 months). Children were divided into two age groups: "4-year-olds" (n = 26; 16 boys, 10 girls; Mean age = 52 months) and "5-year-olds" (n = 24; 12 boys, 12 girls; Mean age = 66 months). An additional 7 children were tested, 1 of them was excluded due to not finishing the game, 2 children were excluded due to parental interference, and 4 children were excluded due to inattentiveness. Participants were recruited through the database of the Central European University Babylab in Vienna. Participating children received a toy upon finishing the study. Parents gave written consent, while children gave verbal consent before participating in the study.

Apparatus and materials

The experiment took 10-15 minutes to complete, and it was played in a room used for behavioral testing. Each testing session was recorded by three cameras. The children sat at a table while the Experimenter stood on the opposite side, and the Agent stood next to the Experimenter.

The game consisted of four rounds corresponding to the four conditions, during which animals were hidden inside two boxes on the table. The boxes had a top part with a small opening through which a body part of the animal could be pulled out to reveal a clue. Otherwise, the children could not see inside the boxes. Occasionally, the Experimenter placed a grey occluder in front of the child to prevent them from seeing the animals being hidden. There were two decks of cards on the table, each consisting of four cards. One deck showed the animals that would be hidden in the current round and indicated which animal the child had to find (the target animal). The other deck showed either a closed or an open door, indicating whether the Agent would have to leave the room at a given point. On each round, the cards were chosen by the Experimenter but were placed in a predefined order based on the counterbalance list that the child was participating in. Six stuffed animals, each approximately 20 cm tall, were used in the experiment: a black octopus, a white octopus, a caterpillar, a bird, a rabbit, and an antelope. The animals were always hidden in pairs. Two pairs had unique properties that were used to trick the children in the two complex conditions. As part of the *hidden identity trick*, the white octopus could be turned inside out, turning into a black octopus. In the *shared property trick*, the caterpillar-bird pair shared the same body part. The green antennae of the caterpillar and the green wings of the bird were identical. Although the body parts were potentially visible during the whole trial, children never reported noticing the shared traits before the trick was revealed.

Procedure

After entering the laboratory, the Experimenter explained the rules of the detective game to the children. They were told that there would be four rounds of the game and that, in each round, the Experimenter would hide two animals in the two boxes, one in each box. The goal of the game was to find the animals together. On each round, the Experimenter would give the children a clue to help them find the animals.

There were four conditions in the experiment. FB-baseline and K-baseline, which were similar to traditional change of location scenarios, and two new conditions, FB-complex and K-complex. Children participated in all four conditions in a within-subjects design. The two complex conditions involved multiple steps, introducing more cognitive demands compared to the baseline conditions. The type of the condition determined whether the Agent and the child saw the hiding, whether the Agent had to leave the room, and whether the animals with the trick properties were used. There were four counterbalance lists, determining the order of the conditions, the animals used, and the side of the hiding. The experiment always started and ended with a complex condition, and two matching mental state attributions never followed each other.

K-baseline and FB-baseline always used the rabbit-antelope pair for the hiding. These two conditions started in the same way, with both the child and Agent seeing the hiding of the animals. For instance, the target animal was hidden in box A, while the other animal was hidden in box B. After the hiding, the Experimenter asked the child the see-control questions; *Did you see the hiding of [target animal]? Did Agent see the hiding of [target animal]?* The Experimenter showed them a clue, a body part hanging out of the box. In such conditions, the clue was not needed to infer the location of the animals, as both participants saw the hiding. Then, in the K-baseline condition, a closed-door card was picked, indicating that the Agent could stay in the room, while in the FB-baseline, she had to leave. In both conditions, the Experimenter playfully changed the location of the toys. In the K-baseline condition, both the child and the Agent saw the change. The child had to prospectively track their shared reality representation, namely that the target animal is not in A anymore, but in B, and that the Agent is aware of this change as well. After swapping the animals, the Experimenter asked the test-questions: *Where will Agent look for [target animal]? And where will you look for [target animal]?*. The correct answer in this case was that both of them would look for it in box B. In the FB-baseline condition, the Agent did not see the change of the animals. Here, the child had to prospectively update their own representation about the whereabouts of the target animal (from A to B) but maintain that the Agent did not see the change, thus has a false belief about the location of the target animal (still in box A).

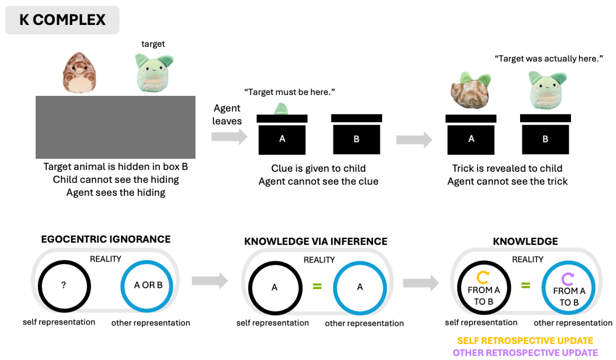

Figure 1: Structure of a K-complex trial.

In the K-complex condition (Figure 1), the Experimenter put up the occluder in front of the child, so that only the Agent saw the hiding of the animals. In this example, the caterpillar and bird were hidden in the boxes, with the caterpillar being the target animal. At this point, the child had an egocentric ignorance; while they did not know the location of the target, they had to represent that the Agent knew. Then, the Experimenter asked the see-control-questions, and a card was picked indicating that the Agent had to leave the room. After this, the Experimenter gave a clue to the child, pulling a green body part out of box A through a small opening. This way, children were able to gain knowledge via inference that the caterpillar was in box A. They also had to attribute the same knowledge to the Agent, who, despite not being present when the clue was shown, saw the hiding of the animals and must know the same. To test whether children correctly made this inference, we used where-control-questions *(Where will Agent look for [target animal]? And where will you look for [target animal]?)*. In case children correctly inferred the location of the caterpillar and also extended this knowledge to the Agent, they should answer that both of them would look for it in box A. Then, the Experimenter revealed the true location of the animals. It turned out that the previously shown green body part did not belong to the caterpillar but to the bird. The Experimenter explained that the antennas of the caterpillar and the wings of the bird are the same, so it is easy to confuse them. The bird was actually in box A, while the caterpillar was in box B. Here, the child had to realize that they made the wrong inference about the location of the caterpillar, who was in fact in box B. Moreover, they also had to realize that the Agent knew the correct location of the caterpillar all along, since she saw the hiding. Children had to perform a self retrospective update and an other retrospective update. Then, the Experimenter asked the test-questions. In case children correctly updated their knowledge and realized that they also had to retrospectively update what the Agent knew, they should respond that both of them would look for caterpillar in box B, at its true location.

In the FB-complex condition (Figure 2), the Experimenter put up the occluder in front of the child and asked the Agent to turn around, so that neither of them could see the hiding. In this example, the hidden animals were the black and white octopuses, with the target animal being the black octopus. After the hiding, the Experimenter asked the see-control-questions, then showed a clue, a black body part hanging out of box A. The clue was witnessed both by the child and the Agent. A card with an open door was chosen, and the Agent left the room. The Experimenter asked the where-control-questions, in order to make sure the children made the respective inferences that the black octopus was in box A, for both themselves and the Agent. Then, the Experimenter revealed the true location of the animals, the white octopus being in box A and the black being in box B (instead of the previously thought A). The Experimenter explained that the white octopus can turn black by being turned inside out. Then, she put the white octopus back into box A, now with its identity visible. At this point, children have to do a self retrospective update and realize that what they thought was true before was actually false. Furthermore, they have to realize that the Agent is not aware of the double identity of the white octopus, so they have a false belief that the black octopus is in box A (instead of its actual location in box B). This way, the children had to maintain two conflicting representations. After this, the Experimenter asked the test questions. In case the children correctly updated their own knowledge but also realized that the Agent had a false belief, they should reply that they would look for the black octopus in box B, but the Agent would look for it in box A.

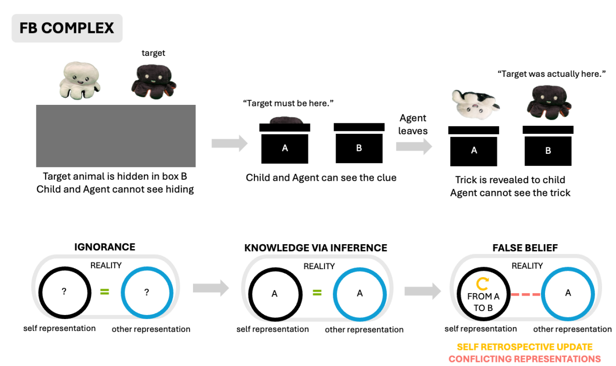

Figure 2: Structure of a FB-complex trial.

While the baseline conditions involved a two-step process regarding representational changes, the complex conditions involved three steps. It is also important to note that the baseline conditions contain a prospective update when children see the animals change location and have to update this information. However, the update in the complex conditions is always retrospective, where children realize that what they knew was false. All conditions involved a clue-showing step, but this was not necessary in the baseline conditions for inferring the location of the animals. The two complex conditions also included extra control questions to test whether children could infer the location of the target animal based on the clue. Children were never suspicious of the tricks beforehand. Some children later mentioned that they were familiar with the octopus toy with two identities; however, they did not expect the trick.

Results

Coding

Data from 50 children were included in the analysis. Children's responses to the control and test questions were coded as 0 (incorrect) and 1 (correct) based on the video recordings. Both pointing and verbal answers were accepted as valid responses. In case a child was alternating between answers, the experimenter asked them to choose, and their final response was included in the data. The data was double coded, with 97.5% agreement between the two coders.

Analysis

A generalized linear mixed model (GLMM) was fitted using the GLMMadaptive package (Pinheiro & Bates, 1995) in R (version 2024.12.1+563) with a binomial distribution and logit link to examine the effects of fixed factors "gender" "order" and random factors "participant id" on the dependent variable "answer" to detect possible gender and order effects but they were not present. A model to test for age differences was also used, but differences were not found. Because of this, the two age groups were analyzed together and will be discussed together. The answers to the control questions significantly differed from chance in all four conditions and will not be discussed further in this paper due to a lack of space.

Binomial tests were conducted to examine whether the observed proportions of each condition differed significantly from chance (Figure 3). Results in the K-complex ($p = 0.003$) and K-baseline ($p < 0.001$) both significantly differed from chance. However, the two FB conditions did not differ from chance (FB-baseline: $p = 0.672$, FB-complex: $p = 0.888$)

A similar GLMM as before was used to examine the effects of "mental state" (FB or K), "type" (baseline or complex), and their interactions, with random slopes and random intercepts (ans ~ mentalstate + type + mentalstate:type, random = ~ mentalstate | ID) on the dependent binary variable "answer". The main effect of mental state was significant, showing that participants had higher log-odds of success in the K conditions compared to FB ($\beta = 2.53$, $SE = 0.78$, $z = 3.23$, $p = 0.001$). The main effect of type was not significant ($\beta = 0.38$, $SE = 0.51$, $z = 0.75$, $p = 0.451$). The interaction between mental state and type approached significance ($\beta = -1.45$, $SE = 0.77$, $z = -1.88$, $p = 0.061$), suggesting a potential differential moderating effect of complexity depending on mental state. While greater complexity was associated with somewhat lower performance in the K conditions, this was not the case in the two FB conditions.

The four conditions were compared using McNemar tests with Yates correction. A significant difference was found between the K-baseline and the FB-baseline conditions ($\chi^2 = 12.9$, $p = 0.001$), the K-baseline and the FB-complex conditions ($\chi^2 = 10.2$, $p = 0.001$), and the K-complex and the FB-baseline ($\chi^2 = 3.89$, $p = 0.049$) conditions. The two K conditions differed numerically, but the difference did not reach statistical significance ($\chi^2 = 2.12$, $p = 0.146$). The two FB conditions did not differ from each other ($\chi^2 = 0.235$, $p = 0.628$). These results indicate that overall, children performed better on knowledge conditions, and that the condition type (baseline or complex) did not affect children's performance. An interesting finding is that despite the K-baseline and FB-baseline conditions differing from each other, we did not find a difference between the K-complex and FB-complex conditions ($\chi^2 = 2.38$, $p = 0.123$), despite the fact that children overall performed worse in the FB-complex condition (52% passed) compared to the K-complex (72% passed).

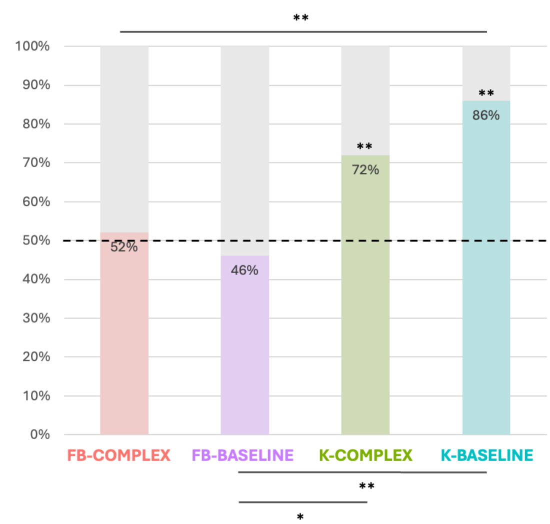

Figure 3: Results of the study with McNemar comparisons.

In order to find out what was driving this effect, we used permutation tests in R to examine associations between binary variables (Pesarin & Salmaso, 2021) in the FB-complex and the K-complex conditions. For each test, the observed difference between a conditional and marginal probability was compared to a null distribution generated by 10,000 random permutations of the conditioning variable. With this, we could test whether the probability of passing K-complex, given that FB-complex was passed, was significantly lower than the overall probability of passing K-complex ($P(Kcomplex=1 | FBcomplex=1) - P(Kcomplex=1)$). Results of the permutation test showed that the probability of passing K-complex is significantly lower when FB-complex is passed, suggesting a negative association between performance in the two conditions ($p = 0.028$). A similar negative association was found between K-complex and FB-baseline ($p = 0.0003$), but not between other conditions. The negative association between K-complex and the two types of FB conditions was further confirmed by phi correlations, showing a significant negative correlation between K-complex and FB-complex ($\chi^2(1) = 7.10$, $\varphi = -0.50$, $p = 0.008$) and between K-complex and FB-baseline ($\chi^2(1) = 10.00$, $\varphi = -0.50$, $p = 0.001$). This finding was rather surprising, and we highlight possible explanations below.

Discussion

The current study aimed to investigate whether knowledge attributions that involve more complex processes and are not based on mere joint perceptual access would be easier to understand for children aged 4-5 than false belief attributions. To test this, we created a new paradigm in which the participating children had to make knowledge and false belief attributions that were more complex than previously used test scenarios. In the K-complex condition, children were tricked into believing something one way, only to discover later that their own knowledge (self retrospective update) and the knowledge they had attributed to the agent (other retrospective update) were both false. In the FB-complex condition, the children discovered that their previous knowledge was incorrect (self retrospective update), but they also had to recognize that their partner still held the now outdated information (conflicting mental states). The results show that, overall, children performed better in the knowledge conditions than in the false belief conditions. Furthermore, they appeared to be resilient to the increased complexity of the new conditions. Their performance did not differ between the two knowledge conditions or between the two false belief conditions. This suggests that, despite the increased cognitive demands of the complex conditions, children were able to overcome them and performed similarly to the matched baseline conditions. However, note that performance on the FB-baseline trial was generally rather low for this age group, which may be due to the general complexity of the procedure that involved four different conditions and many control questions.

Follow-up permutation tests showed that the probability of passing the K-complex condition was significantly lower when the FB-complex or the FB-baseline was passed. This suggests a negative relation between performance on FB and K-complex trials, but not K-baseline trials. In other words, children who passed the FB conditions were not necessarily passing the K-complex condition. This suggests that, although both K trials were easier than FB trials, the K-baseline was robust and not influenced by the children's individual explicit false belief understanding. Conversely, children who passed the FB conditions performed worse in the K-complex condition. To understand why this occurred, we must consider what it means to fail a K-complex trial but pass the two FB trials. To pass the two FB conditions but fail the K-complex condition, children should answer that the agent would look for the target in a different location from them. In the K-complex condition, the Agent sees the hiding while the child does not, causing the child to have egocentric ignorance. Then, the Agent leaves the room, and the clue and the trick are only revealed to the child. Despite the agent not seeing the trick, the children should realize that, as she had seen where the target was originally hidden, she knew its location all along. If children fail to make this inference, they may think that the agent has a false belief and answer that she would look for the target in the wrong place.

This may reflect a unique developmental pattern whereby children who have mastered explicit false belief tasks overgeneralize such false belief inferences, applying them to any situation in which two perspectives differ, even temporarily. Lagattuta and colleagues (2010) found a similar overgeneralization in children. Their study investigated how children reason about when people should share the same versus different interpretations of a situation. Their findings showed that 6- to 7-year-olds display an overgeneralization in theory of mind: after learning that people's interpretations can differ based on prior experience, they apply this rule too broadly and predict differences even when common ground is obvious. This "overinterpretive" pattern suggests a transitional developmental stage in which children's growing insight into mental diversity temporarily leads them to overlook shared understanding. In our study, the negative relation between FB- and K-complex trials might reflect such an overgeneralizing strategy at a younger age. This highlights the need for investigating further how children at different ages represent mismatches in mental states and whether special cases, such as egocentric ignorance understanding, hinder the way children interpret knowledge states.

The negative association between the K-complex and two FB trials may also reflect a mature mentalization strategy. If an agent's perceptual access to an event stops at any point and the event changes, it is easier to consider this person as uninformed about the event than to maintain their original representation. After all, one cannot be sure whether she will return to the scene, nor when. This suggests that testing adult participants would provide a better understanding of when it is worthwhile to perform retrospective updates about another person, and when it is easier to simply form a new representation. Moreover, individual differences could be observed. It is possible that some adults would also tend to use the same strategy, representing the agent as uninformed instead of maintaining their decoupled representation and performing retrospective updates, while others would actively maintain the agent's representation of the event.

As exhibited in the current experiment, overall, children succeeded in managing egocentric ignorance and extended knowledge gained by inference to others, demonstrating great flexibility in how knowledge can be attributed. This suggests that knowledge attributions are resilient in the face of additional cognitive demands. Although the novel knowledge complex condition included a similar amount of cognitive demands to the two false belief scenarios (though of a different kind), children performed better than in the false belief trials. This highlights the importance, and perhaps primary nature of knowledge attributions in mentalization (Phillips et al., 2021).

Acknowledgments

We are grateful to the families who participated in this research. We thank Christina Regorosa, Dorottya Mészégető and Ágnes Volein for research assistance. Anna Kispál is recipient of a DOC Fellowship of the Austrian Academy of Sciences at Central European Univeristy.

References

Baron-Cohen, S., & Goodhart, F. (1994). The 'seeing-leads-to-knowing'deficit in autism: The Pratt and Bryant probe. *British journal of developmental psychology*, 12(3), 397-401.

Fabricius, W. V., Boyer, T. W., Weimer, A. A., & Carroll, K. (2010). True or false: Do 5-year-olds understand belief?. *Developmental Psychology*, 46(6), 1402.

Fabricius, W. V., Gonzales, C. R., Pesch, A., Weimer, A. A., Pugliese, J., Carroll, K., ... & Spinrad, T. L. (2021). Perceptual access reasoning (PAR) in developing a representational theory of mind. *Monographs of the Society for Research in Child Development*, 86(3), 7-154.

Fogd, D., Téglás, E., & Kovács, Á. M. (2025). Adults Represent Others' Logical Inferences Even When It Is Unnecessary. *Cognitive Science*, 49(6), e70076.

Gettier, E. L. (1963). Is justified true belief knowledge?.

Holland, C., & Phillips, J. (2020). A theoretically driven meta-analysis of implicit theory of mind studies: The role offactivity. *In Proceedings of the Annual Meeting of The Cognitive Science Society* (Vol. 42).

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition,* 109(2), 224-234.

Király, I., Oláh, K., Csibra, G., & Kovács, Á. M. (2018). Retrospective attribution of false beliefs in 3-year-old children. *Proceedings of the National Academy of Sciences*, 115(45), 11477-11482.

Király, I., Oláh, K., & Kovács, Á. M. (2023). Can 18-month-olds revise attributed beliefs?. *Open Mind*, 7, 435-444.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330(6012), 1830-1834.

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*(6308), 110-114.

Lagattuta, K. H., Sayfan, L., & Blattman, A. J. (2010). Forgetting common ground: six-to seven-year-olds have an overinterpretive theory of mind. *Developmental Psychology*, 46(6), 1417.

Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental science*, 12(1), 142-149.

Martin, A., & Santos, L. R. (2016). What cognitive representations support primate theory of mind?. *Trends in cognitive sciences*, 20(5), 375-382.

Nagel, J. (2017). Factive and nonfactive mental state attribution. *Mind & Language*, 32(5), 525-544.

Oktay-Gür, N., & Rakoczy, H. (2017). Children's difficulty with true belief tasks: Competence deficit or performance problem?. *Cognition*, 166, 28-41.

Rakoczy, H., & Oktay-Gür, N. (2020). Why do young children look so smart and older children look so dumb on true belief control tasks? An investigation of pragmatic performance factors. *Journal of Cognition and Development*, 21(2), 213-239.

Pesarin, F., & Salmaso, L. (2010). *Permutation tests for complex data: theory, applications and software*. John Wiley & Sons.

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics*, 4(1), 12-35.

Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., ... & Knobe, J. (2021). Knowledge before belief. *Behavioral and Brain Sciences*, 44, e140.

Phillips, J., & Norby, A. (2021). Factive theory of mind. *Mind & Language*, 36(1), 3-26.

Schidelko, L. P., Huemer, M., Schröder, L. M., Lueb, A. S., Perner, J., & Rakoczy, H. (2022). Why do children who solve false belief tasks begin to find true belief control tasks difficult? A test of pragmatic performance factors in theory of mind tasks. *Frontiers in Psychology*, 12, 797246.

Schidelko, L. P., & Rakoczy, H. (2025). Why do children from age 4 fail true belief tasks? A decision experiment testing competence versus performance limitation accounts. *Cognitive Science*, 49(6), e70069.

Schneider, D., Slaughter, V. P., Bayliss, A. P., & Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition*, 129(2), 410-417.

Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, 150, 43-52.

Vouloumanos, A., Martin, A., & Onishi, K. H. (2014). Do 6-month-olds understand that speech can communicate?. *Developmental Science*, 17(6), 872-879.

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.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child development*, 75(2), 523-541.

Westra, E., & Nagel, J. (2021). Mindreading in conversation. *Cognition,* 210, 104618.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103-128.

Yu, C. L., & Wellman, H. M. (2024). A meta-analysis of sequences in theory-of-mind understandings: Theory of mind scale findings across different cultural contexts. *Developmental review,* 74, 101162.