Children's Expectations About Epistemic Change

Mackenzie Briscoe (mbriscoe@g.harvard.edu)*[1], Rui Zhang (rui.zhang.rz378@yale.edu)*[2], Julian Jara-Ettinger (julian.jara-ettinger@yale.edu)[2]

[1]Department of Psychology, Harvard University, [2]Department of Psychology, Yale University; *Equal contribution

Abstract

People's mental states constantly change as they navigate and interact with their environment. Accordingly, social reasoning requires us not only to represent mental states but also to understand the ways in which mental states tend to change. Despite their importance, relatively little is known about children's understanding of the dynamics of mental states. To explore this question, we studied a common type of mental state change: knowledge gain. Specifically, we studied whether five- and six-year-olds distinguish between agents who gain knowledge from those who lose knowledge. In one condition, children saw an agent answer a two-alternative choice question incorrectly, followed by an identical-looking agent who answered the same question correctly (i.e., gaining knowledge). In another condition, children saw the reverse pattern (i.e., losing knowledge). Children were more likely to infer they had seen two different agents in the knowledge loss condition relative to the knowledge gain condition. These results suggest that children have intuitions about how epistemic states change and open new questions about children's naive theories of mental state dynamics.

Keywords: Theory of Mind; Social Inference; Cognitive Development

Introduction

The capacity to think about other people's knowledge and beliefs is critical for navigating the social world. From early in childhood, these representations help us understand what other people say (e.g., Bohn et al., 2018; Saylor & Ganea, 2018), predict how they might act (Jara-Ettinger et al., 2017), and represent what they think about us (Asaba & Gweon, 2022). They also guide our decisions about who to trust (Aboody et al., 2022; Einav & Robinson, 2011; Koenig & Harris, 2005) and who to share knowledge with (Bridgers et al., 2020). Moreover, in adults, inferences about other people's epistemic states can be nuanced (Baker et al., 2017; Aboody et al., 2021; Croom et al., 2023), and in some cases even automatic (Rubio-Fernandez et al., 2019).

Because of their importance, understanding the nature of how people represent each other's knowledge and beliefs has been a major research area, including how we represent these states (e.g., Phillips et al., 2021), how they develop in childhood (e.g., Wellman et al., 2021; Wellman & Liu, 2004), and how they vary across species (e.g., Krupenye et al., 2016; Martin & Santos, 2016). At the same time, research into this question has generally focused on how we infer what an agent may or may not know from their behavior at a particular point in time (e.g., inferring that someone is ignorant if they give a wrong answer or make a bad decision; and that they are knowledgeable if they are consistently accurate, or if they navigate an environment in an efficient and competent manner). In more realistic situations, however, people's knowledge and beliefs are dynamic and in flux as agents interact with the world.

To illustrate this, imagine running into a friend and asking whether they've listened to a new album that just came out. Imagine that your friend tells you they've never heard of the artist, but two weeks later, you overhear them raving about the album. Detecting this knowledge change in your friend might come as a pleasant surprise. You might infer that they likely took note of your comment and listened to the artist since you last saw them. Consider, by contrast, your intuitions if the opposite happened: imagine that your friend agreed with you about the new album being great, but two weeks later, you overheard them claiming to have never heard of that artist before. This epistemic change would no longer appear natural, and you might wonder if your friend was confused or dishonest when they first talked to you.

This example points to a basic intuition about how we expect other people's knowledge to change. Intuitively, gaining knowledge within a short period of time is more common than suddenly losing knowledge. Indeed, a growing body of research has recently argued that these types of dynamic expectations about how minds change are more central to Theory of Mind than previously thought. That is, real-world social interactions require that we represent the dynamic life in other people's minds. Consistent with this, recent computational work has found that intuitions about how mental states change over time support a variety of complex social inferences such as attributions of preferences (Gates et al., 2021), goals (Zhang et al., 2023), and dynamic processes like memory and distraction (Berke & Jara-Ettinger, 2021; Berke et al., 2023).

Despite the growing literature pointing to the importance of these intuitions in adult social cognition, relatively little is known about their developmental origins. While some work has found that children can represent dynamic processes like thinking (Richardson & Keil, 2020, 2022), little is known about children's expectations about what types of changes in mental states are more plausible than others.

In this paper, we sought to investigate these capacities with a focus on knowledge change. Specifically, we sought to test whether preschoolers already have expectations about which types of mental-state changes are more likely (e.g., from ignorance to knowledge) than others (e.g., from knowledge to ignorance).

As a first test, we focused here on five- and six-year-olds because research showing children's understanding of thinking appears to develop at around age five (Richardson

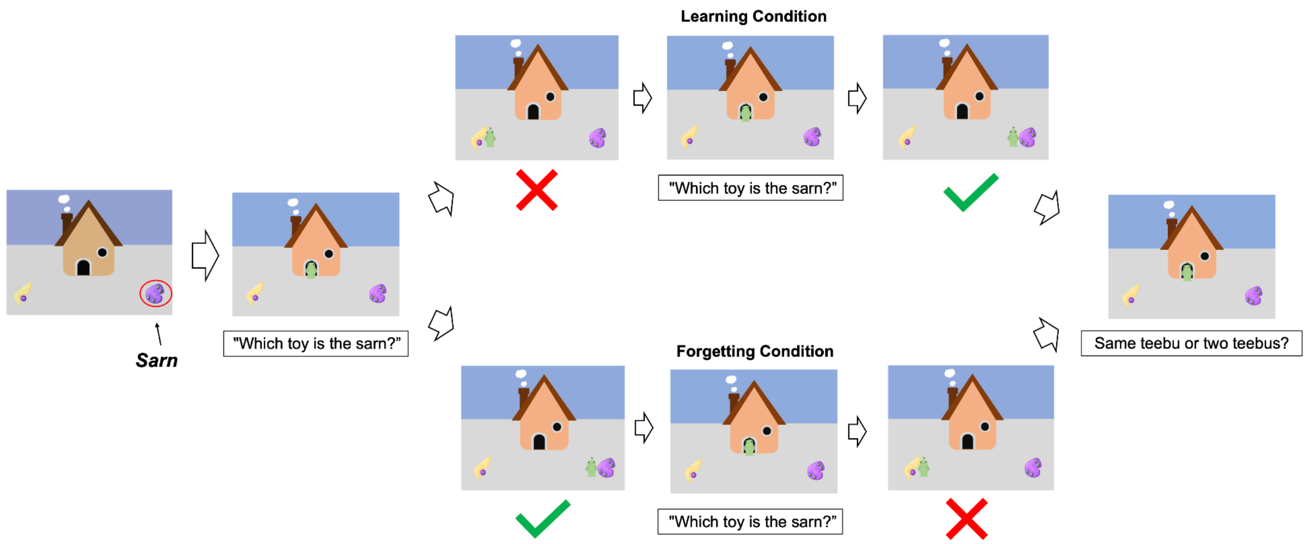

Figure 1. Experiment procedure. Children saw two objects equidistant from a house and were taught which one was the *Sarn* (object randomized across participants). A novel agent (called a *Teebu*) next exited the house and was asked to identify the Sarn. After doing so, the Teebu walked back inside the house. An identical-looking Teebu then emerged from the house and was also asked to identify the Sarn. In the *Learning Condition*, the first Teebu chose the wrong object and the second Teebu chose the right object. In the *Forgetting Condition*, the first Teebu chose the right object and the second Teebu chose the wrong object. At the end, participants were asked whether they had seen the same Teebu or different Teebus.

& Keil, 2020, 2022). At the same time, these specific intuitions might emerge much earlier, and our task was also designed to be easily adaptable for younger ages—even infants—in follow-up work.

Experiment

Our experimental paradigm was inspired by paradigms developed for infant research (e.g., Bonatti et al., 2002; Rivera & Zawaydeh, 2007; Xu & Carey, 1996; Xu, 2002). In our task, children watched two identical agents try to identify which of two novel objects was a "Sarn" and receive feedback on whether they were correct. One agent answered correctly and the other incorrectly, and we varied the order in which they appeared between conditions. Children were then asked to infer whether they had seen a single agent twice, or two different agents. If children have adult-like expectations about knowledge change, they should infer that there were two different agents when the first agent answered correctly and the second agent answered incorrectly. By contrast, children should feel uncertain when the first agent answers incorrectly and the second agent answers correctly (as it is consistent with knowledge gain). All aspects of the experiment were pre-registered unless explicitly noted: OSF link:
https://osf.io/56stx/?view_only=766194a1909448b68ac156506af3d5f0

Methods

Participants 52 five- to six-year-olds (M_{Age}= 6.09, range = 5.05 - 7.00; Male = 21, Female = 17, Prefer not to say = 14; White = 18, Black = 2, Asian = 14, mixed = 9, other = 1, no response = 8) were recruited and tested online. Participants were split between two conditions (N = 26 per condition). Eight additional participants were recruited but excluded from the task (see Results for details).

Stimuli Stimuli consisted of simple animations implemented in Powerpoint (see Fig. 1). Each animation showed a house at the center of the screen with two novel objects placed equidistant from it. An agent (called a *Teebu*) then appeared at the doorway of the house and was asked to identify which object was called a *Sarn*. The Teebu moved towards one of the objects, received feedback on whether they were right or not, and then returned to the house and disappeared. This process was then repeated, with the only difference being that the Teebu in the second trial always moved towards the object that the first Teebu did not choose.

Procedure Children were first introduced to a group of identical novel characters called *Teebus*. Children were told that every Teebu has their own house but that, sometimes, the Teebus' friends will come over to visit them. During this session, children saw a screen with a set of six identical Teebus, each standing in front of a different house, and were asked to confirm how many Teebus live in one house. If children answered this question incorrectly, the experimenter

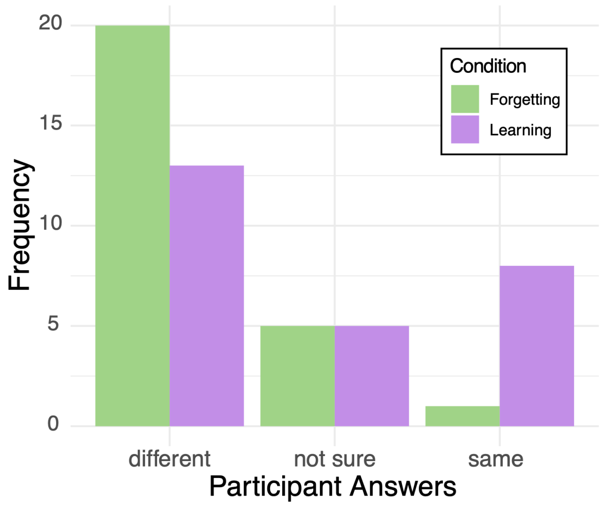

Figure 2. Histogram of participant choices (x axis) as a function of trial type (color coded).

corrected them to make sure children understood that only one Teebu lived in each house. The experimenter then presented a display of a new house with two novel objects placed equidistantly from it (Fig. 1). The experimenter then drew attention to the two objects, told the child that one of them was called a *Sarn,* and asked them to guess which one. Regardless of the child's answer, the experimenter always told them they were incorrect and asked them to guess again. We included this process because research suggests that children can have a strong optimism bias and we wanted to ensure they considered the possibility that ignorant agents can get things wrong (Chen et al., 2015; Garnham & Ruffman, 2001). After their second guess, the experimenter confirmed that they had identified the Sarn. Note that this implies that children's initial choice determined which object would be the Sarn in the task (such that the purple object was the Sarn for 65.4% of participants, and the yellow object was the Sarn for the remaining 34.6%). Participants who chose the same object on both guesses were excluded from the task (see Results).

After this warm-up phase, the experimenter explained that the task was to figure out if there was one Teebu or two different Teebus inside the house (recall that the Teebus are all visually identical). To achieve this, the experimenter explained that they would ask the Teebu to identify the Sarn as soon as they came out of the house.

In the *Learning condition*, the first Teebu walked out of the house, the experimenter asked, "Hi Teebu, one of these toys is a Sarn. Can you stand in front of the Sarn?" and the Teebu then moved towards the incorrect object. When the Teebu stopped in front of the object, the experimenter, talking to the participant, stated, "That's not it. So we saw that the Teebu did not know which toy was the Sarn." The experimenter then provided feedback to the Teebu, saying, "Teebu, that wasn't it! That toy wasn't the Sarn." The Teebu then walked back to the house and disappeared. An identical-looking Teebu then appeared in front of the house and the experimenter once again asked, "Hi Teebu, one of these toys is a Sarn. Can you stand in front of the Sarn?" and the Teebu now moved towards the correct object. The experimenter then confirmed to participants, "That's right. So we saw that the Teebu *did* know which toy was the Sarn." Again, the experimenter provided feedback to the Teebu, "Teebu, that was right! That toy was the Sarn."

The *Forgetting condition* was identical to the *Learning condition,* with the only difference being that the trial order was inverted. Thus, the first Teebu correctly moved to the Sarn, and the second Teebu moved to the incorrect object.

After this, participants were asked to remind the experimenter whether the first Teebu and second Teebu were able to identify the Sarn, asking one question at a time. Participants who answered incorrectly were corrected by the experimenter. Children were asked the test question, "Can you tell me, did we see the same Teebu two times, did we see two different Teebus, or are you not sure?" and children were allowed to answer "same", "different", and "not sure." Children were then asked to explain why they thought so. Finally, as an inclusion criteria, children were asked to confirm which toy was the Sarn. (see Results).

Results

Eight children were excluded from the task due to: failure to identify the Sarn at the end of the task (N = 3), experimenter error (N = 1), prior participation in a similar version of the task (N = 1), and family interference (N = 3).

Figure 2 shows the experiment results. In the forgetting condition, we predicted that children would infer that they saw two different Teebus significantly above chance. Indeed, 20 of the 26 participants (76.92%, 95% CI: [61.54.14% - 100%]) reported there were two different Teebus ($p = .005$ through a pre-registered one-tailed binomial test, binning "same" and "not sure" responses as a single 'incorrect' category, and setting chance to 50%).

In the *learning condition,* the switch from an incorrect to a correct response is consistent with both the idea that this was the same Teebu or two different Teebus. We therefore did not predict any significant effects in this condition. Indeed, participants were distributed across all response types: 13 of the 26 participants (50%, 95% CI: [26.92% - 65.38%]) reported 'different' Teebus, 8 of 26 participants (30.77%, 95% CI: [11.54% - 46.15%]) reported 'same' Teebu, and 5 of 26 participants (19.23%, 95% CI: [3.85% - 34.62%]) reported they were not sure. As predicted, however, this distribution of responses was significantly different from the one obtained in the *forgetting condition* ($X^2(2) = 6.93$; $p = .03$ through a pre-registered chi-squared test). Together, these results suggest that children indeed distinguished between the two patterns of epistemic change, and were more inclined to infer that the apparent loss of knowledge implied the presence of two agents.

Given that children's guess about which object was a Sarn determined which object would ultimately be the right answer, we also tested whether this had any influence on participant judgments. To do this, we tested whether the

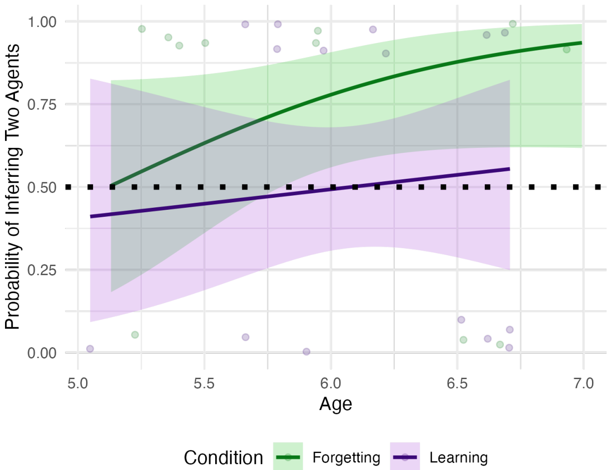

Figure 3. Participants' responses as a function of age. Each dot represents a participant's answer. The x-axis shows their age, and the y-axis shows their answer. The response *"there were two different characters"* was coded as 1, and other answers ("*same"* and *"not sure")* were coded as 0. Data are jittered slightly on the y-axis (but not on the x-axis) for visibility. Each line represents a logistic regression fit to an experimental condition.

distribution of children's answers was significantly different between children who initially thought the purple object was the Sarn (34.6%) and those who initially thought the yellow object was the Sarn (65.4%). Two exploratory chi-squared tests revealed no difference in these participants in either the *learning condition* ($X^2(2) = 3.99$, $p = .14$) or the *forgetting condition* ($X^2(2) = 2.14$, $p = .34$).

Finally, we ran an exploratory analysis testing for potential developmental changes through a logistic regression predicting children's judgment that there were two different Teebus as a function of age, one for each condition. Neither regression revealed a significant effect ($\beta = 1.43$, $p = .10$ and $\beta = .35$, $p = .69$ for the forgetting and learning conditions, respectively, predicting probability of stating 'different' in both cases). Although we did not find age effects, our sample size was selected to have enough power to test children as a whole, not as their age varies. The results from Figure 3 suggest that there might be some developmental change that our study was unable to identify.

General Discussion

People's minds are in constant flux. Consequently, social reasoning requires that we understand the dynamics of mental states. This study explored children's expectations about knowledge change. In the forgetting condition, where an agent initially appeared to know the name of an object, followed by an identical-looking agent who lacked this knowledge, children were significantly more likely to infer they had seen two different agents. Conversely, in the learning condition, where an agent initially appeared to lack knowledge but, after receiving feedback, subsequently had it, children did not have a strong expectation about whether they had seen the same agent or different agents. These results suggest that young children already have intuitions about knowledge change, and that they believe that losing knowledge is less likely to occur than gaining knowledge within a short timeframe.

Our work has several implications. First, our results suggest that young children not only have a capacity to represent mental states but also have intuitions about what types of trajectories mental states are more likely to follow. Such expectations might be particularly important for rapid real-world social cognition. The ability to predict how other people's beliefs, knowledge, and desires will change in an ongoing interaction might help us build quick expectations about others and make the problem of mental-state inference easier.

In addition, our experiment suggests that children in our task were able to integrate an understanding that agents can engage in logical reasoning. This was seen in the *learning condition*. Here, the first agent that selected the incorrect object was only told they were wrong. Given that there were only two options, this logically implies that the other object was the right answer. This inference justifies the possibility that the second agent was the same Teebu as the first one (now having gained knowledge through logical reasoning). Note, however, that 50% of children in the learning condition still stated that they had seen two different Teebus. This opens the possibility that only about half of the children understood that the agent had made a logical inference. Children who did not realize this might have seen the change from no knowledge to knowledge as implying they had seen two different agents. We hope to explore this possibility in future work.

Our work also leaves several open questions. First, during our warm-up procedure, children were asked to guess which object was the *Sarn* and, regardless of their choice, were always told they were wrong and asked to guess again. We included this experience phase due to past research suggesting that children have a general optimism bias (e.g., Wente et al., 2020), and therefore wanted to avoid the possibility that children might not initially associate being correct with being knowledgeable. It is possible that this manipulation was unnecessary, as some research has also argued that children expect ignorant agents to always make errors (Chen et al., 2015; Garnham & Ruffman, 2001), although other research suggests this may not be the case (Friedman & Petrashek, 2009; Aboody et al., 2019). Nonetheless, it remains an open question whether children's intuitions would have changed in the absence of an initial experience of gaining knowledge. In particular, it is possible that this manipulation helped the children in the learning condition who seemed to expect the agent to engage in deductive inference. Future work will explore this question.

A second open question concerns the lack of apparent developmental change. Despite the null result, our visualization of age effects suggests that older children might have been performing better at the task (Fig. 3). This opens the possibility that children's intuitions in this age range are indeed maturing, but that we simply lacked the power to detect developmental change. One potential source of evidence that children's intuitions are still developing comes from the *learning condition*. If all children clearly understood that the change from being wrong to being right was consistent with both one or two agents, why didn't all children respond 'not sure'? The fact that many children sided with either stating there was only one agent, or two agents, opens the possibility that some children might not have realized that the evidence was consistent with both possibilities (and hence, this might develop throughout this age). At the same time, it is also possible that children in our age range were indeed aware that it was not possible to determine whether they had seen the same or two different agents. However, the children who avoided saying 'not sure' may have simply decided to guess. We hope that future work will help distinguish between these possibilities.

Our work focused on children's expectations regarding knowledge gain and loss in short timescales and when these mental state changes were small (the difference between knowing or not knowing which of two objects is the *Sarn*). As adults, our intuitions about mental state change are highly nuanced and influenced by these two factors. It is a common experience to walk into a room and suddenly forget what you are looking for. Or, you might remember the title of a song you like one moment, only to forget it entirely or leave the name at the tip of your tongue the next. Further, it is also easy to imagine scenarios where it may be unlikely to acquire knowledge within a short period of time. Imagine you discovered that your friend just started learning a new language yesterday, and then saw them speaking it fluently the next day. It would seem unlikely that your friend could learn to speak a new language in a day. You might even suspect maybe they'd been lying to you about what they know. This reveals that adult-like expectations about epistemic dynamics are highly complex and context sensitive. Our work is a first step towards characterizing these expectations.

The importance of what information is being shared might also be relevant to children's inferences about mental state change. For instance, you might not be surprised if a new acquaintance forgets your name, but you might be more taken aback if the parents of someone you've been dating do so. In other words, the perceived utility of a piece of information might influence how much you expect another person to remember it and can therefore further shape your intuitions about others' mental state changes.

More broadly, expectations about these dynamics are not limited to knowledge; they also apply to other mental states like goals and desires. For example, people may have different intuitions about mental state change regarding things like food preferences and food craving. Like knowledge, people may assume that food preferences are generally stable over time, whereas view food cravings are more transient. For instance, if an agent switched from liking chocolate to disliking it, we might suspect the presence of two different agents, but this intuition might be weaker if the agent switched from craving chocolate to no longer craving it.

Overall, our work points to a dimension of Theory of Mind that has been historically understudied: expectations about mental dynamics. Our work shows that young children already have some expectations about the dynamic nature of epistemic change, and is a first step towards characterizing the development of people's rich intuitions about how mental states change. Our findings highlight that, from a young age, children are not only capable of reasoning about others' mental state changes but might also expect these changes to happen in predictable ways.

Acknowledgments

This work was supported by NSF award BCS-2045778.

References

Aboody, R., Davis, I., Dunham, Y., & Jara-Ettinger, J. (2021). I can tell you know a lot, although I'm not sure what:Modeling broad epistemic inference from minimal action.

Aboody, R., Huey, H., & Jara-Ettinger, J. (2022). Preschoolers decide who is knowledgeable, who to inform, and who to trust via a causal understanding of how knowledge relates to action. Cognition, 228, 105212.

Aboody, R., Zhou, C., Flowers, M., & Jara-Ettinger, J. (2019). Ignorance = doing what is reasonable: Children expect ignorant agents to act based on prior knowledge. In *CogSci*(pp. 1297-1303).

Asaba, M., & Gweon, H. (2022). Young children infer and manage what others think about them. *Proceedings of the National Academy of Sciences*, *119*(32), e2105642119.

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 0064.

Berke, M., & Jara-Ettinger, J. (2021). Thinking about thinking through inverse reasoning.

Berke, M., Tenenbaum, A., Sterling, B., & Jara-Ettinger, J. (2023). Thinking about Thinking as Rational Computation.

Bohn, M., Zimmermann, L., Call, J., & Tomasello, M. (2018). The social-cognitive basis of infants' reference to absent entities. *Cognition*, *177*, 41-48.

Bonatti, L., Frot, E., Zangl, R., & Mehler, J. (2002). The human first hypothesis: Identification of conspecifics and individuation of objects in the young infant. *Cognitive psychology*, *44*(4), 388-426.

Bridgers, S., Jara-Ettinger, J., & Gweon, H. (2020). Young children consider the expected utility of others' learning to decide what to teach. *Nature human behaviour*, *4*(2), 144-152.

Chen, Y., Su, Y., & Wang, Y. (2015). Young children use the "ignorance= getting it wrong" rule when predicting behavior. *Cognitive Development*, *35*, 79-91.

Croom, S., Zhou, H., & Firestone, C. (2023). Seeing and understanding epistemic actions. *Proceedings of the*

National Academy of Sciences, *120*(49), e2303162120.
Einav, S., & Robinson, E. J. (2011). When being right is not enough: Four-year-olds distinguish knowledgeable informants from merely accurate informants. *Psychological science*, *22*(10), 1250-1253.
Friedman, O., & Petrashek, A. R. (2009). Children do not follow the rule "ignorance means getting it wrong". *Journal of Experimental Child Psychology*, *102*(1), 114-121.
Garnham, W. A., & Ruffman, T. (2001). Doesn't see, doesn't know: is anticipatory looking really related to understanding or belief?. *Developmental Science*, *4*(1), 94-100.
Gates, V., Callaway, F., Ho, M. K., & Griffiths, T. L. (2021). A rational model of people's inferences about others' preferences based on response times. *Cognition*, *217*, 104885.
Jara-Ettinger, J., Floyd, S., Tenenbaum, J. B., & Schulz, L. E. (2017). Children understand that agents maximize expected utilities. *Journal of Experimental Psychology: General*, *146*(11), 1574.
Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child development*, *76*(6), 1261-1277.
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.
Martin, A., & Santos, L. R. (2016). What cognitive representations support primate theory of mind?. *Trends in cognitive sciences*, *20*(5), 375-382.
Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., ... & Knobe, J. (2021). Knowledge before belief. *Behavioral and Brain Sciences*, *44*, e140.
Richardson, E., & Keil, F. (2020). Children use agents' response time to distinguish between memory and novel inference. In *CogSci*.
Richardson, E., & Keil, F. C. (2022). Thinking takes time: Children use agents' response times to infer the source, quality, and complexity of their knowledge. *Cognition*, *224*, 105073.
Rivera, S. M., & Zawaydeh, A. N. (2007). Word comprehension facilitates object individuation in 10-and 11-month-old infants. *Brain research*, *1146*, 146-157.
Rubio-Fernández, P., Mollica, F., Ali, M. O., & Gibson, E. (2019). How do you know that? Automatic belief inferences in passing conversation. *Cognition*, *193*, 104011.
Saylor, M., & Ganea, P. (2018). *Active Learning from Infancy to Childhood*. Springer International Publishing AG.
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.
Wente, A. O., Goddu, M. K., Garcia, T., Posner, E., Fernández Flecha, M., & Gopnik, A. (2020). Young children are wishful thinkers: The development of wishful thinking in 3-to 10-year-old children. *Child development*, *91*(4), 1166-1182.
Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive psychology*, *30*(2), 111-153.
Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*(3), 223-250.
Zhang, C., Kemp, C., & Lipovetzky, N. (2023). Goal recognition with timing information. *Proceedings of the International Conference on Automated Planning and Scheduling, 33* , 443–451.