# Efficient Mindreaders: Flexible belief reasoning and updating in children

## **Paula Fischer**

SUPERVISOR: Ernő Téglás

SECONDARY SUPERVISOR: Ágnes M. Kovács

Thesis submitted for the degree of Doctor of Philosophy (PhD)

Department of Cognitive Science Central European University

Submitted: Budapest, April 2022

#### **Originality Statement**

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at Central European University or any other educational institution, except where due acknowledgment is made in the form of bibliographical reference.

I also declare that the intellectual content of this thesis is the product of my own work, and the claims here reflect my own thinking. However, as I believe scientific research is inherently a social activity, and all work receives support from others in theoretical, methodological, or stylistic matters; I will present this work in first- person plural voice.

**Paula Fischer** 

### Abstract

Human social life highly depends on our ability to comprehend other minds. Theory of Mind (ToM) refers to the ability to make inferences about others' mental states, allowing us to predict and explain others' behaviour. While even infants and young children show some understanding of other people's mental states, it is still unclear how rich their representational capacities are, and furthermore, what processes allow them to revise and update their own belief and those attributed to others in case of an event or information that is inconsistent with their previously acquired knowledge.

The first objective of this thesis is to investigate whether young children's belief attribution ability is flexible enough to operate in various scenarios, similarly to adults. We specifically ask whether they can attribute beliefs contents to others that they can entertain in first person: i.e., beliefs about tool efficiency and causal events. In our first study, we show that children already by their third year of age are able to integrate efficiency information in false belief reasoning: they successfully predict an agent's action based on his false belief about the efficiency of a tool.

The second objective of the thesis is to address children's belief revision strategies. As children rapidly learn new information, and sometimes new information can contradict with their previous beliefs about the world, the question emerges how they maintain consistency. Evidence from studies with adults using conditional premises suggests that adults prefer to keep the observed data and update the conditional rule when encountering new and conflicting information, but little is known how young children deal with such inconsistencies, and whether their strategies are different in first person and third person belief revision. In this line of study, we find that 5-year-old children, similarly to adults, are more likely to revise a rule and preserve the observed data, but only in case the rules are not constrained, and not for rules that are subject

to physical constraints. However, this is different when children are required to update a belief about constrained rules attributed to an agent: in this case, children also become more likely to update the rule itself and retain the data. We argue that as the representation of relational contents is likely to be more complex, and thus more fragile, children are more likely to update it in case of conflicting information.

Together these findings suggest that children possess sophisticated abilities to represent others' beliefs: similarly, to adults, they are able to integrate contents that they can understand in first person when they are reasoning about someone's beliefs. Moreover, when they are learning about the world, and acquire beliefs about contingencies or rules, they reorganize their belief sets in a consistent manner, similarly to adults, when they face counterevidence.

## Acknowledgements

First, and foremost, I would like to thank my supervisors, Ernő Téglás and Ágnes Melinda Kovács. Thank you for guiding me and pushing through my limits. Working with you was highly inspirational, and it changed my attitude towards research and science forever. Thank you for bearing with my messiness.

I'd like to thank the faculty members of the Cognitive Science Department, with whom I had fruitful discussions from the beginning, and who influenced my work and my way of thinking till today: Gergely Csibra, György Gergely, Dan Sperber, József Fiser, Christophe Heintz and Ildikó Király. Special thanks to Csaba Pléh, for providing all the knowledge, books and discussions to deepen and widen my knowledge. Thank you for teaching me to be forever curious, and for giving me a different perspective on science.

I cannot thank enough the members of the Babylab, Mari and Ági, and the assitants, for their unlimited help, especially Zsuzsi and Iuli for helping me with experiments from Study 1, and Dorka, for her patience to test so many kids for Study 3. I'm also thankful to all the children and their parents who participated in the presented studies.

I'm grateful for past and current members of the Cognitive Science Department, who helped me to get started, and provided insightful feedback on my plans and finished works: Dorka, John, Denis, Mikolaj, Katka, Nicolo, Hanna, Tibor, Basia, Petya, Rachel. You were all my mentors at least once.

I'm especially grateful for my PhD cohort and friends from the department: Gina, Francesca, Otavio, Simily, Laura, Barbu, Ákos, Eszter and Liza. I hope we will keep in touch.

I cannot express my gratitude enough towards our coordinators, Edit and Réka. You took so much weight from my shoulders, not just regarding bureaucratic issues. You are a miracle, I wish you were everywhere.

I would like to thank my wonderful family for their endless support. I feel incredibly lucky to be part of your circle.

Thanks to my friends, Àdàm, Bohu, Bori, Koma, Elza, Heni, Lilla, Lili, Sàra, Viki, Zsófi for being there for me, no matter what.

I dedicate this thesis to Máté. The best brother in his very own and interesting way, and second best father in the world. Thank you for teaching me to find my purpose and to follow it. Thank you for teaching me to form my very own opinion.

Abstract
Acknowledgements 4
Chapter 1: Introduction10
Background 10
The development of Theory of Mind: Contradictory findings from traditional tasks and implicit
tasks
Bridging the gap: Alternative explanations and The Two-System account
Evidence and counterevidence for the Two-System Account18
Can infants attribute false beliefs about identities?19
Evidence pointing to the signature limits of the early Theory of Mind abilities
False belief tasks with reduced processing demands25
Uncharted avenues: Investigating the flexibility of children's belief attribution
Uncharted avenues: Investigating the flexibility of children's belief attribution
Uncharted avenues: Investigating the flexibility of children's belief attribution
Uncharted avenues: Investigating the flexibility of children's belief attribution       29         Beliefs about tool efficiency       30         Children's understanding of tool efficiency       32         Attributing beliefs based on causal inferences       33
Uncharted avenues: Investigating the flexibility of children's belief attribution       29         Beliefs about tool efficiency       30         Children's understanding of tool efficiency       32         Attributing beliefs based on causal inferences       33         Updating beliefs in first person and third person       35
Uncharted avenues: Investigating the flexibility of children's belief attribution       29         Beliefs about tool efficiency       30         Children's understanding of tool efficiency       32         Attributing beliefs based on causal inferences       33         Updating beliefs in first person and third person       35         The current studies       38
Uncharted avenues: Investigating the flexibility of children's belief attribution       29         Beliefs about tool efficiency       30         Children's understanding of tool efficiency       32         Attributing beliefs based on causal inferences       33         Updating beliefs in first person and third person       35         The current studies       38         Chapter 2. Can young children integrate information about tool efficiency in false belief
Uncharted avenues: Investigating the flexibility of children's belief attribution       29         Beliefs about tool efficiency       30         Children's understanding of tool efficiency       32         Attributing beliefs based on causal inferences       33         Updating beliefs in first person and third person       35         The current studies       38         Chapter 2. Can young children integrate information about tool efficiency in false belief         reasoning?       41
Uncharted avenues: Investigating the flexibility of children's belief attribution       29         Beliefs about tool efficiency.       30         Children's understanding of tool efficiency.       32         Attributing beliefs based on causal inferences.       33         Updating beliefs in first person and third person.       35         The current studies.       38         Chapter 2. Can young children integrate information about tool efficiency in false belief         reasoning?       41         2.1.       Theoretical background
Uncharted avenues: Investigating the flexibility of children's belief attribution       29         Beliefs about tool efficiency       30         Children's understanding of tool efficiency       32         Attributing beliefs based on causal inferences       33         Updating beliefs in first person and third person       35         The current studies       38         Chapter 2. Can young children integrate information about tool efficiency in false belief         reasoning?       41         2.1.       Theoretical background       41         Theory of Mind abilities in infants and young children       43
Uncharted avenues: Investigating the flexibility of children's belief attribution       29         Beliefs about tool efficiency       30         Children's understanding of tool efficiency       32         Attributing beliefs based on causal inferences       33         Updating beliefs in first person and third person       35         The current studies       38         Chapter 2. Can young children integrate information about tool efficiency in false belief         reasoning?       41         Theoretical background       41         Theory of Mind abilities in infants and young children       43         How infants and children represent beliefs? Some theoretical possibilities       45

## Table of Contents

Theories describing an early developing flexible belief reasoning system	48
Representing a variety of contents: Beliefs about tool efficiency	52
Young children's understanding of tool efficiency	53
2.2. Experiment 1. Beliefs about tool efficiency: the case of 4-year-olds	58
2.2.1. Methods	58
2.2.2. Results	65
2.2.3. Discussion	66
2.3. Experiment 2. Can 4-year-olds represent beliefs about efficiency when realit	y bias is
eliminated?	67
2.3.1 Methods	68
2.3.2. Results	69
2.3.3. Discussion	70
2.4. Experiment 3. 5-year-olds integrate efficiency information in false belief rea	soning (reality
bias version)	
bias version)	
bias version) 2.4.1. Methods 2.4.2. Results	
bias version).         2.4.1. Methods         2.4.2. Results         2.4.3. Discussion.	
bias version).   2.4.1. Methods   2.4.2. Results   2.4.3. Discussion   2.5. Experiment 4. 5-year-olds tracking true beliefs about tool efficiency	
bias version)   2.4.1. Methods   2.4.2. Results   2.4.3. Discussion <b>2.5. Experiment 4. 5-year-olds tracking true beliefs about tool efficiency</b> 2.5.1. Methods	
bias version).   2.4.1. Methods   2.4.2. Results   2.4.3. Discussion <b>2.5. Experiment 4. 5-year-olds tracking true beliefs about tool efficiency</b> 2.5.1. Methods 2.5.2. Results	
bias version)	
bias version)         2.4.1. Methods         2.4.2. Results         2.4.3. Discussion         2.5. Experiment 4. 5-year-olds tracking true beliefs about tool efficiency         2.5.1. Methods         2.5.2. Results         2.5.3. Discussion         2.5.4. Experiment 5. 3-year-olds integrate efficiency information in false belief read	
bias version)   2.4.1. Methods   2.4.2. Results   2.4.3. Discussion <b>2.5. Experiment 4. 5-year-olds tracking true beliefs about tool efficiency 2.5.1.</b> Methods <b>2.5.2.</b> Results <b>2.5.3.</b> Discussion <b>2.6. Experiment 5. 3-year-olds integrate efficiency information in false belief reareality-bias version</b> )	
bias version).   2.4.1. Methods   2.4.2. Results   2.4.3. Discussion <b>2.5. Experiment 4. 5-year-olds tracking true beliefs about tool efficiency 2.5.1.</b> Methods <b>2.5.2.</b> Results <b>2.5.3.</b> Discussion <b>2.6.1.</b> Separ-olds integrate efficiency information in false belief reality-bias version) <b>2.6.1.</b> Methods	
bias version)   2.4.1. Methods   2.4.2. Results   2.4.3. Discussion <b>2.5. Experiment 4. 5-year-olds tracking true beliefs about tool efficiency 2.5.1.</b> Methods 2.5.2. Results 2.5.3. Discussion <b>2.6. Experiment 5. 3-year-olds integrate efficiency information in false belief reality-bias version</b> ) 2.6.1. Methods 2.6.2. Results	

2.7. Experiment 6. Replication of Experiment 5: Excluding alternative explanations	79
2.7.1. Methods	79
2.7.2. Results	80
2.7.3. Discussion	81
2.8. General Discussion	82
Chapter 3. Can young children represent false beliefs about causal events?	88
3.1. Introduction	88
3.2. The aim of the current study	94
Infants and young children's understanding of causal relations	95
Beliefs based on causal inferences	96
3.3. Experiment 1. Can children attribute beliefs based on causal inferences?	
3.3.1. Methods	100
3.3.2. Results	105
3.3.3. Discussion	107
3.4. Experiment 2. Go-No-Go task	108
3.4.1. Methods	
3.4.2. Results	110
3.4.3. Discussion	111
3.5. Experiment 3. Location change task	112
3.5.1. Methods	113
3.5.2. Results	115
3.5.3. Discussion	115
3.6. General Discussion	116
Chapter 4. How do children update their own beliefs and beliefs attributed to oth	ers? 120
4.1. Introduction	120

How do adults update their own beliefs in case of inconsistency?	
How do children revise their beliefs?	
4.2. Aims of the current studies	
4.3. Experiment 1A and 1B. Strategies of 3 and 5-year-old children's for revising	rules that are
causally opaque	138
4.3.1. Methods	
4.3.2. Results	
4.3.3. Discussion	146
4.4. Experiment 2. 5-year-old belief revision of rules supported by physical const	train 147
4.4.1. Methods	
4.4.2. Results	
4.4.3. Discussion	150
4.5. Experiment 3. 5-year-old children belief revision from a third person perspe	ctive 151
4.5.1. Methods	152
4.5.2. Results	
4.5.3. Discussion	155
4.6. Experiment 4. Revising attributed beliefs in 5 and 6-year-old children	156
4.6.1. Methods	
4.6.2. Results	
4.6.3. Discussion	161
4.7. General Discussion	163
Chapter 5. General Discussion	
References	

### **Chapter 1: Introduction**

#### Background

Humans are highly social creatures. From very early on, during our every-day life, we are almost constantly involved in social interactions. Starting from a simple situation like allowing a person passing through a door before us, to more complex scenarios, such as when we try to explain a complicated story to a friend who has difficulties to understand as she was not involved in the preceding events, we face various social scenarios on daily bases. During these events, we highly depend on the understanding of others mental states; we often take into account what others want, know, think and believe. One of the most important aspects of human social abilities is that we recognize and understand what others think and believe, and how the content of the mind influences our and others' behaviour. Theory of Mind (ToM) refers to the ability to make inferences about others' mental states, allowing predicting and explaining their actions (Premack & Woodruff, 1978). Such abilities are also associated with several positive developmental outcomes, from improved cooperation to better academic achievements (e.g., Sally & Hill, 2006; Kloo, Perner, 2008). To gain a better understanding about how this ability enables humans to navigate efficiently in their social environment, some remaining questions need to be answered about the development and the flexibility of such abilities. The current thesis aims to investigate questions regarding how flexible the early developing Theory of Mind system is, and what are the differences between updating first person and third person representations.

## The development of Theory of Mind: Contradictory findings from traditional tasks and implicit tasks

Since Dennett (1978), Bennett (1978) and Harman (1978) pointed out that attributing true beliefs to others cannot be distinguished from reality-based reasoning, the litmus test for ToM abilities became probing false belief (FB) understanding. One typical way to test for the understanding of false beliefs in children is the location-change task (Wimmer and Perner 1983; Baron Cohen, Leslie & Frith, 1985). In such a standard false belief task, participants are exposed to a story involving a set of characters, where the first character has a false belief regarding a location of her marble (as a second character changed its location while she was absent). When the child is asked where the first character will look for her marble, most 3-yearolds fail to take into account her false belief and will answer (or point to) the actual location of the marble (Wimmer & Perner, 1983, Perner, Leekam & Wimmer, 1987) from which it was concluded that children before the age of 4 do not understand that others can have beliefs that are in conflict with reality or their own beliefs. In the last 30 years, a large body of supporting evidence corroborated these findings, consolidating a view according to which a conceptual change takes place in reasoning about mental states at about 4 years of age, when children become capable of understanding that agents can hold and act on false beliefs and thus, grasp the representational nature of the mind. Research from the last 15 years on infants' implicit understanding of false beliefs, however, cast doubts for the proposal that the standard explicit FB task may be the litmus test for false belief reasoning.

To date, over 30 published reports using nonverbal FB tasks have provided evidence for false belief understanding in infants and toddlers (see Scott & Baillargeon, 2017 for a review, but see Kulke & Rakoczy, 2018 for an overview about replications and non-replications). These tasks have involved a variety of paradigms measuring different responses, for instance

spontaneous responses such as infants' looking times (Violation of Expectation paradigms - VoE), anticipatory looking, preferential looking, neural responses or children's spontaneous helping behavior.

A VoE tasks take advantage of infants' natural tendency to look longer at events that violate their expectations. The very first evidence about infants' false belief understanding using this paradigm was obtained by Onishi and Baillargeon (2005). In this experiment, infants first saw an agent hide her toy in a box (e.g., green box) as opposed to the other box (e.g., blue box). During the test trials, the agent either witnessed (true belief trials) or was absent (false belief trials) the location change of her toy, and later she either reached towards the box congruently or incongruently with her (false) belief. Infants expected the agent to reach towards the box where she (falsely) believed her toy was, and they showed evidence for a violation of expectation (measured by their longer looking time) whenever the agent reached towards the location incongruent with her belief. Tasks using such paradigms found evidence of false belief understanding (i.e., involving beliefs about object locations) in infants before their first birthday (e.g., Kovács, Téglás & Endress, 2010). In a more recent study using VoE paradigm Scott (2017) demonstrated that infants seem to understand not only the beliefs of others, but also have accurate expectations regarding different emotional reactions to an outcome depending on the character's beliefs about the given situation. For instance, in a set of experiments infants expected the agent to express surprise when the outcome of a scenario (involving the change of location or identity of an object) was inconsistent with her belief, compared to scenarios when the agent was merely ignorant and therefore held no expectation about the possible outcome, or when the outcome was consistent with her beliefs. Results from this study suggest that infants can flexibly apply their mental state understanding abilities to a range of belief-inducing situations and belief-based responses (such as surprise when someone else discovers that his belief turned to be false), similarly to older children and adults.

Other tasks have asked yet a further question, specifically whether infants are also able to predict how someone would behave based on his beliefs. Whereas studies measuring infants' looking time after an agent performed a certain behaviour (congruent or incongruent with his beliefs) indicate how infants react after an agent executed an action, studies measuring infants anticipatory gaze have suggested that they are also able to predict an agent's action, which is a highly relevant aspect of social interactions, for instance when cooperating with others. In anticipatory looking tasks infants are expected to anticipate that the agent will approach a location where she (falsely) believes her desired object to be, that is, look at that location before the action takes place. For instance, Southgate, Senju and Csibra (2007) found that 2-year-olds are able to predict an agent's behaviour based on their false beliefs, measured by their anticipatory looking. Whereas Kampis, Kármán, Csibra and Southgate (2021) did not replicate this specific study (for replication issues see a review by Kulke & Rakoczy, 2018; Kulke, von Duhn, Schneider & Rakoczy, 2018), other studies using anticipatory paradigms found evidence for false belief understanding in 2-year-olds and even in 17-months-old infants (e.g., He, Bolz, Baillargeon, 2012; Surian & Geraci 2012). Furthermore, studies eliciting 'anticipatory pointing behaviour to correct a mistaken belief (spontaneous use of pointing gesture to indicate where an object is about which an agent has a false belief) have found impressive evidence in this direction in 18-24-month-old infants (Knudsen & Liszkowski; 2012a, b).

However, anticipatory gaze might not be the most suitable measure to measure ToM processes, and this measurement may also have various limitations, as suggested by the replication studies (Kampis et al, 2021). Oftentimes very little anticipation is found, which may be explained by infants' immature visual disengagement (Elsabbagh et al., 2013), in the sense that younger infants may be slower at disengaging from a moving stimulus and move their eyes towards a static one to generate action predictions. However, studies using alternative measurements, such as neural correlates for action predictions, provided further evidence about

infants' ability to predict an agent's action based on his beliefs. Previous studies with infants demonstrated a decrease in alpha amplitude over the sensorimotor cortex (indicating motor activation) when they were presented with an object that implied impending action (Southgate & Begus, 2013). Based on this, Southgate and Vernetti, (2014) conducted an experiment measuring infants' neural responses using electroencephalography (EEG), when they observed an agent who had a belief about an object present at a certain location and could act on that object. Results indicated that 6-months-old infants show an increase in sensorimotor alphaband suppression when they expect an agent to search for an object at a certain location based on her false belief, but this is absent when the agent falsely believed that the container was empty. Another study building on the proposal that when infants ascribe a representation to another person, they would rely on the same representational apparatus used in first person reasoning, found increased temporal gamma-band activation (a neural correlate for sustained object representation) in 8-month-old infants when they observed an agent who falsely believed that there is an object behind an occluder (Kampis, Parise, Csibra and Kovács, 2015). This result suggests that infants attributed a sustained representation of the object to an agent when she lost visual access to the object, and even after when the object disintegrated (and this was only visible to the infants), by successfully computing the agent's visual perspective, and representing the agent's belief about the existence of the object behind the occluder.

Studies measuring neural responses in infants not only showed that infants encode an agent's false belief about the presence of an object, but that they are also able encode another person's belief in communicative contexts. Forgács and his colleagues (2020) found that 14-month-old infants show an N400 effect (a neural signature indicating semantic incongruence) when an object is being labelled congruently from their own perspective but did not match with the communicative partner's false belief. That is, infants are not only able to represent others'

beliefs about objects at certain locations, but they also understand others' misunderstanding in communicative contexts.

Additionally, besides studies as described above measuring infants' implicit responses, some studies tested infants' and young children's active behavioural responses with interactive tasks. In such elicited intervention tasks, children observe a scene about an agent who holds a false belief and then they are prompted to perform an interactive behaviour by taking into account the agent's false belief. These tasks are important indicators about children's understanding of others' beliefs in more natural contexts compared to the previously mentioned studies using implicit measurements. Moreover, in these tasks infants and children are required to interpret and make sense of others' actions based on their false beliefs, and they are required to use this understanding in social interactions. For instance, a study implementing a helping paradigm conducted by Buttelmann, Carpenter and Tomasello (2009) showed that 18-monthold infants took into account an agent's (false) belief, and they helped him based on that, to obtain his desired object from a locked box. In this experiment, infants first witnessed an agent hiding his toy in box A, and then left. While he was away, a second experimenter took the toy and moved it to box B and then locked both boxes. After the agent came back, he tried to open the box where he left the toy. When infants were prompted to help the agent, they went and opened the box that contained the desired toy, suggesting that they interpreted the agent's behaviour (opening the empty box) as aiming to obtain the toy about which he falsely believed to be in that box; however, this was not the case where the agent witnessed the change, and yet aimed to open the box that did not contain the toy – in this case infants interpreted the agent's behaviour as wanting to open this empty box, and helped him opening that one. In an earlier interactive task, Carpenter, Call and Tomasello (2002) conducted an experiment where 3-yearold children watched an agent placing two distinct toys in two different boxes, and later in her absence a second experimenter switched the two objects. When the agent came back and pointed the box where the toy originally was, infants approached the other box to get the toy for her, suggesting that they took into account her false belief about the location of her desired object, and they understood that this is the toy which she wanted.

These results suggest that infants and young children seem to understand others' beliefs and make active use of this understanding during social interactions.

The studies discussed above indicate that the ability to attribute beliefs to others is present from early on. Based on such ability, infants and young children are able to represent others' (false) beliefs, predict others' behaviour, and they are able to operate efficiently during social interactions by taking into account what others may think or believe.

#### Bridging the gap: Alternative explanations and The Two-System account

The studies with infants discussed above have mainly used implicit measures (or implicit tasks requiring active interventions) and have found success at an early age in false belief tasks. In contrast, explicit tasks requiring verbal responses suggest that children under the age of 4 are unable to represent another person's false beliefs (Wellman, Cross & Watson, 2001). Such antagonist findings raise the possibility that an implicit Theory of Mind understanding, and an explicit false belief understanding might involve different abilities.

In order to resolve such discrepancies some accounts have proposed alternative explanations regarding infants' success on these implicit tasks. For instance, Perner and Roessler (2012) describe that in order to correctly predict a character's behaviour during traditional, explicit false belief task, the child is required to perform an intentional switch from her perspective to the character's perspective, but she is unable to do so before the age of 4, due to the lack of conceptual, executive function or linguistic abilities. Perner and his colleagues point out that younger children and infants' success on implicit false belief tasks might be

explained by low-level or association-like processes (e.g., Perner & Ruffman, 2005). Perner and Ruffman argue that, for instance, in the study by Onishi & Baillargeon (2005) when infants watch an agent placing an object to a certain location, they may form a three-way association between the agent, the object and its location, and they are surprised (look longer) if in the test phase they observe a pattern that is incongruent with this association.

Another interesting proposal aiming to explain this discrepancy between implicit and explicit false belief understanding comes from Apperly and Butterfil (2009) who developed a two-system account for Theory of Mind reasoning. They argued that an early developing system, called Minimal Theory of Mind that can track simpler belief-like "registrations", may explain success in implicit false belief tasks, but it does not allow representing false beliefs per se. Based on this account, as infants have limited cognitive resources and little conceptual sophistication, they use this early developing system to track belief-like states. Specifically, the early developing system allows the infant to track where an agent has last seen (registered) an object, and therefore is able to predict where the agent will look for it to fulfil his goal directed action. These registrations only capture the relation between the agent and an object and its location. For instance, if an agent 'encounters' (which is a proxy to the notion 'perceive' an object that she has perceptual access to, based on its physical proximity for example) an object at a certain location, which is then moved to a new location in her absence, the early developing system will be able to predict that the agent will search for the object where she originally registered it. Since these registrations are relations between an agent, an object and a location, representing them does not require metarepresentations (beliefs about beliefs), and neither are aspectual, in the sense that it is not encoded under what aspect an agent represented an object. Thus, it was argued that the early developing system has specific signature limits. This system may represent objects at a certain location, while it is unable to track the aspectuality (e.g., identity) of objects or their function, and neither can deal with quantifiers or logical relations. It cannot distinguish between what is represented and how it is represented (Apperly & Butterfil, 2009, p.963.). However, the early developing system is claimed to be automatic, fast but inflexible, and is unable to handle scenarios with high processing demands, as such scenarios are "placing demands on working memory, attention and executive function, they would be incompatible with automaticity" (Butterfil & Apperly, 2013, p. 629). Moreover, the early developing system, based on its inflexible nature, is largely encapsulated from the rest of the cognition and therefore would be unable to integrate information from other cognitive processes when reasoning about agents' actions (Butterfill & Apperly, 2013).

Importantly, according to this view, only the later developing system, "involves representing propositional attitudes like beliefs, desires and intentions to construct reasongiving, causal explanation of action." (Butterfill & Apperly, 2013, p. 607) The later developing system emerges around the age of 4, is slow, effortful and flexible, can be called the full-blown Theory of Mind, also used by adults, and it is capable of representing a wide range of beliefs and enables children to pass on the traditional explicit tasks. In line with this proposal, Rakoczy (2017) further argues, that children acquire the ability to represent propositional attitudes such as beliefs around the age of 4, when they understand aspectuality.

#### Evidence and counterevidence for the Two-System Account

Some studies cast doubts on the distinction between the two ToM systems. On the one hand, the growing amount of evidence demonstrating infants' understanding of false beliefs about identity questioned the signature limits of the early developing system. Representing and tracking false beliefs about identity require encoding the aspect under which the other person has represented an object, building on prerequisites that fall beyond the boundaries that characterize registrations. On the other hand, other studies have targeted the explicit false belief tasks and by reducing processing demands showed successful false belief understanding with toddlers and young children. Together, studies from these two lines point to a single system for Theory of Mind reasoning, that emerges early and operates more efficiently with other gradually developing cognitive abilities.

Studies provided positive results about infants' ability to represent other's beliefs about object identities as indicated by their looking time (e.g., Song & Baillargeon, 2008; Scott & Baillargeon, 2009), anticipatory looking (Buttelmann & Kovács, 2019), manual search time (Kampis & Kovács, 2021) or by their helping behaviour (Suhrke & Buttelmann, 2015). Furthermore, there is evidence about toddlers and young children's performance on traditional, explicit false belief tasks when processing demands are reduced (Setoh, Scott & Baillergeon, 2016, Scott & Roby, 2015, Surian & Leslie 1999, Rubio-Fernandez & Geurts, 2013, 2016) indicating that children's difficulty in certain false belief tasks are not due to their inability to represent beliefs, but to various performance demands.

#### Can infants attribute false beliefs about identities?

Central to the two-system account is the proposal that the early developing system is unable to deal with false beliefs about identity. In this framework, infants or young children can represent relations between an agent and a physical object (i.e., objects encountered by the agent at a certain location) but lack the ability to represent an object under a certain aspect.

Given the nature of beliefs, it is possible to hold a certain belief when the object of belief is represented under one aspect and a contradictory belief when it is represented under a different aspect. In certain cases, reasoning with beliefs faces inferential challanges, for instance: if a.) "John believes that X", b.) "X=Y" it does not follow that c.) "John believes that Y", unless one is aware about the equivalence relation in b.). The substitution of the coreferents is not truth preserving ("John believes that Superman is strong" can be true even though "John believes that Clark Kent is strong" is false, despite "Superman" and "Clark Kent" referring to the same person). This kind of understanding of the aspectuality of beliefs is crucial for reasoning supported by the belief concepts.

Contrary to the predictions of the two-system account, there are several studies suggesting that infants are capable of such attributions involving aspectuality or identity of objects in false belief reasoning.

For instance, a study by Scott and Baillargeon (2009) presented 18-month-old infants with two identically looking toy penguins. The difference between the penguins was that one of them could be separated into two pieces, whereas the other one not. Participants were familiarised with an agent who always placed her keys into the separable penguin and demonstrated preference by continuously reaching for the separable penguin and hiding the key inside that one, and then joined the two pieces into one. Importantly, the agent always found the separable penguin in a separate form and the other one in a conjoint form. In the beginning of the false belief test phase, the agent was absent, and once she returned, she saw a one-piece penguin under a transparent cover (which was in fact the two piece one put together, but the agent didn't know), while the other penguin was covered by an opaque cloth. Infants looked longer when the agent reached for the penguin under the transparent cover, as they expected her to infer that if the one-piece penguin is under the transparent cover, then two-piece penguin must be the under the opaque cover, which would satisfy her goal (to hide her key). The authors suggested that infants' expectation could be achieved only if they attributed to the agent a false belief about the identity of the two-piece penguin, as during the familiarization trials she always witnessed the penguin in a divisible form. However, Butterfil and Apperly (2013) argue that in this study, infants do not ascribe false beliefs about identities, but rather, false beliefs about object affordances defined by the different object types. Specifically, infants may represent that

the uncovered penguin is the indivisible one that does not afford hiding a key, and as there are always two penguins present, the other one under the cover must be the devisable penguin. However, it is not clear from this criticism, how representing an object as having (or not having) certain affordances and attributing this to another agent is different from representing an object under a certain aspect. Furthermore, this interpretation also assumes that infants have the ability to reason by exclusion (*if* this not the devisable penguin *then* it must be the other one) and attribute this inference to others, which is already a more sophisticated reasoning ability that goes beyond the signature limits the two-system account proposed for the early developing ToM system (Carruthers, 2016).

Another study by Buttelmann and his colleagues (2014) tested 18-month-old infants' ability to attribute false beliefs about unexpected contents. First, infants observed an agent who encountered three boxes that contained a block, and in his absence, participants learnt that the fourth box contained a spoon. Once the agent returned, infants were asked to help the agent and could give either a block or a spoon. – Infants always gave the object that matched the agent's (false) belief: in case the agent did not witness the content of the box, and falsely believed that it contained a block, then infants gave a block to him, however, this pattern reversed when the agent was present and thus was aware that the fourth block contains a spoon. Furthermore, Scott and her colleagues (2015) asked the question whether infants also understand rather complex scenarios in which an agent induces a false belief about an object's identity to deceive another person in a competitive scenario. 17-month-old infants witnessed that the thief is aiming to deceive another agent by stealing his desirable, rattling toy and substituting it with a less desirable, silent toy. Infants realised that this replacement could only be effective if the silent toy is identically looking and expected the deceived agent to mistake her original toy for the silent one after she returned, indicating that infants are able to understand not only others'

beliefs about identities, but also the actions of an agent who is aiming to induce false beliefs to someone else based on the features of objects.

Furthermore, a recent study by Kampis and Kovács (2021) investigated whether 14month-old infants' searching behaviour would be modulated by an agent's partial knowledge about an object (that had two different aspects) which led her to be mistaken about the objects' identity. Crucially, in the false belief condition the available information for the agent was compatible with two different objects. Infants witnessed two agents, one of them put an object under a certain aspect (hedgehog) into a box, and then the second agent left. In her absence, the first agent took the object out and demonstrated its different identity by transforming the hedgehog into a different plush figure and put it back to the box. The agent who was previously absent came back and retrieved the object (in its new form) which led her to falsely believe that original object (the hedgehog) must still be inside the box (as she was not aware of the double identity of the object). Next, infants were allowed to search in the box. Infants searched longer when the agent was absent during the transformation demonstration suggesting that their search was modulated by the false belief of the agent that there must be yet another object in the box. Their search times were shorter in the scenario where the agent witnessed the object's transformation and therefore expected only one object in the box which was then retrieved.

Overall, these results suggest, that the early developing system may not be limited to tracking beliefs about object locations, infants also understand another agent's false belief about the identities of objects. This supports the possibility that from early on, infants may be able to entertain various types of belief contents, contrary to the signature limits suggested by the minimal theory of mind account. Infants also understand others' actions and intentions of implementing false beliefs about object identities to others, and moreover, their behaviour can be modulated by someone else's false belief about the identity of an object.

#### Evidence pointing to the signature limits of the early Theory of Mind abilities

Although there are several positive findings concerning infants' abilities to represent false beliefs about object identities, other studies revealed signature limits of young children's mindreading abilities. Low and Watts (2013) investigated 3 and 4-year-old children's and adults' anticipatory gaze to determine whether they automatically represent other's false beliefs regarding identities of objects. Participants were tested with an anticipatory looking paradigm with two different scenarios, the first one involving a location change of an object, whereas the second an identity based false belief task. Participants were also required to give verbal predictions in the test phase about the agent's behaviour (where she will search the object). The results showed that in the location change task, participants across all age groups correctly anticipated towards the location which was congruent with an agent's (false) belief, however, majority of the participants anticipated to the incorrect location (incongruent with the agent's false belief) when the agent had a false belief not about the location, but the identity of the object. Regarding participant's verbal predictions, it was found that accurate responses increased with age; 3-year-olds were below chance in both identity and location change tasks, whereas 4-year-old children where above chance on the location, but at chance on the identity task, and adults gave correct verbal responses in both tasks.

Fizke and her colleagues (2017) further investigated whether 2- and 3-year-old children attribute beliefs about identities similarly to beliefs about the location of objects with an active helping paradigm. In one experiment, children were required to take into account an agent's (false) belief about the location of an object, whereas in a second experiment they were required to attribute (false) beliefs about the identity of an object. Whereas children distinguished between false and true belief conditions about the location of the object and helped the agent accordingly, they were unable to do so when the agent had a false belief about the identity of the object. However, as Fizke and her colleagues (2017) also note, children's performance did not differ significantly between the two experiments regarding the false belief conditions (location change vs, aspectuality), therefore these results may not support a strong difference in children's ability to represent false beliefs about location vs. identities of objects. However, Oktay-Gür, Schulz and Rakoczy (2018) used an analogous design and found a difference in children's performance when they were required to take into account an agent's false belief about the location, versus the identity of objects, that is success in the first and failure in the latter case.

These results are in line with the two-system account, supporting an early developing system that is fast and automatic and able to operate only with simple scenarios, and a second system that develops later and may be slow and effortful, but can operate with more complex scenarios, for instance when reasoning about others' beliefs about the identities of objects. However, Carruthers (2016) pointed out, that the identity studies conducted by Low and Watts (2013, 2014) likely involve greater inferential demands compared to change of location tasks, as in such experiments, participants are required to encode an object from an agent's perspective, and later, when they discover that it is double sided (blue on one side and red on the other), they need to update their belief about the object which was previously encoded under one aspect, and understand that it's an object with double identity. Furthermore, they need to infer that the agent encoded the object's identity (being red or blue) based on its other side (based on only one aspect, as the agent did not have visual access to the other side of the object). These sorts of inferences may require high working memory demands and involve visual rotation, resulting in a greater cognitive demand compared to the change of location tasks, and infants' failure in certain experiments involving aspectuality could be explained by the executive demands of the tasks, rather than their lack of competence to attribute beliefs about identity.

In the following parts, studies involving standard explicit false belief tasks with reduced processing demands will be discussed to further explain young children's difficulty on traditional tasks.

#### False belief tasks with reduced processing demands

Different attempts at reducing processing difficulties focused on lowering the inhibitory control demands in false belief tasks. Leslie, Friedman and German (2004) advanced an interesting proposal to explain children's difficulties during explicit false belief tasks. Based on this, when children are required to answer the test question "Where Sally look for her marble?" they are required to overcome their own correct answer based on the marble's actual location in order to correctly predict Sally's behaviour based on her false belief. Studies have found that when such inhibitory control demands are reduced, for instance the target object is removed from the scene, children under the age of 4 succeed in explicit false belief tasks (e.g., Rubio-Ferndez & Geurts, 2013; 2016).

Besides lowering the inhibitory control demands triggered by the target object's actual location, there are other ways to decrease processing demands in false belief tasks. For instance, younger children might have issues with interpreting the "where" question in the test phase due to pragmatic difficulties. A study from Setoh, Scott and Baillargeon (2016) aimed to investigate this issue. The authors tested two-and-a-half-year-old children's ability to succeed on the explicit false belief location change task. One of the assumptions behind this study was that when children are asked the standard question, they must interpret the question, hold it in memory and generate a response, while inhibiting the prepotent response (e.g., the object's actual location). Overall, these demands might impact children's performance, which may result in random responding. In the first experiment children listened to a story accompanied

by a large picture book about a character who placed her apple in a container, and while she was absent, her brother took the apple away. To test whether children's performance could be improved by lowering response generation demands, before the actual test question, two practice trials were used, where children were required to answer some simple questions regarding the location of the apple's or that of another toy of Emma by using the "where" question (e.g., Where is Emma's apple? Where is Emma's ball?). These practice trials were intended to support the response generation process, as during the test trial, after the apple has been taken away in the character's absence, children were asked where she will look for her apple after she came back. Results indicated that 33-month-old children succeeded on the task with the help of these practice trials. To further assess which factors contributed to children's success, the authors conducted another experiment where the practice trials remained, but the story was modified in a way that in the false belief test the other character did not simply take the apple away, but put it in another container, posing a greater inhibitory demand on children's performance as they were required to inhibit the object's actual location to answer correctly. Results from this experiment showed that children's performance fell below chance, suggesting that for young children to succeed, both response generation demands, and inhibitory demands need to be reduced. However, Rubio-Fernandez, Jara-Ettinger and Gibson (2017) argues that children might be prompted by the factual "where" question during the training to always point to the last location where the apple was.

Besides such difficulties, there might be other factors that influence children's performance, for instance the need to keep track of the protagonist's perspective throughout a false belief task might also pose difficulties for younger children. Rubio-Fernandez and Geurts (2013) investigated whether children under the age 4 are able to pass the standard false belief task if they are being helped to keep track of the perspective of the agent who holds the false belief, and when an open-ended question that requires an act out response is being used instead

of the standard test question. This open-ended question was implemented to create a more interactive scenario without any break between the narrative and the response elicitation, and moreover, whereas the standard question focuses on the target object, the open-ended question does not mention it at all, reducing processing demands by not requiring children to inhibit the object's actual location. In this experiment, children observed a Duplo character who placed her banana in a fridge (e.g., green fridge) and then left the scene; at this point, to help children to keep track of the perspective of the agent, they were asked if the character can see them from where she is now. This was followed by the replacement of the banana from the green fridge to another fridge (e.g., blue fridge) and the question again was repeated whether the character (who was away) could see what happened from where she is now. Once the character came back, the experimenter gave the Duplo figure to the participant and instead of asking the standard test question, which involves mentioning the target object, she asked the children to act out what is going to happen ("what will the character do now?"). Results indicated that young children succeed on such tasks and take into account the character's false belief, but only if both modifications (the perspective tracking questions and the open-ended test question) are implemented. If only one of the factors is implemented, then children's performance fell below chance. It was concluded from this result, that even children under the age 4 are able to pass the standard false belief tasks, if the both the disruption of the perspective taking processes and inhibitory demands are minimized. Another study by Rubio-Fernandez conducted with adults (2017) further supports the hypothesis that the participants' focus of attention (the perspective of the protagonist) is the key to their performance in false belief tasks.

In another set of studies, Rubio-Fernandez and Geurts (2016) further examined how children's attentional processing difficulties may affect their performance during the standard false belief task. The focus of this study was to address how the use of the standard question impacts children's performance during the task, given that by mentioning the object, it directs children's attention to the object's actual location that has to be disregarded. In two experiments, the standard question (e.g., "Where will Lola look for her toy?") was modified in the following ways: in the first experiment, children were still presented a binary choice between two locations where the character could search, but the question did not involve mentioning the target object (e.g., they simply asked where Lola will go now?). In the second experiment, the target object was mentioned before the open-ended false belief test question was asked. The results indicated that children succeed on the standard task even if they are required to make a binary choice, when the target object is not mentioned, however, they performance is below chance if the target object is mentioned during the test phase, even if not directly in the test question. Thus, children before the age of 4 may fail the standard tasks because these tasks pose greater demands on their attentional processes, independently of their Theory of Mind abilities. However, Kammermeier and Paulus (2018) only partially replicated the pattern obtained by Rubio-Fernandez and Geurts (2013). Based on the results, only children above the age 4 passed the Duplo task, although 3-year-old children's performance was significantly better in the Duplo task than in the standard false belief task.

Relatedly, others have also argued for the role of pragmatic factors and high processing demands of the standard false belief tasks. Salter and Breheny (2019) proposed that the *where* question mentioning the target object in a standard task causes an increased activation in memory for the actual location of the target, making it harder to inhibit. However, these difficulties may not be merely caused by reality bias, but there could be also a bias coming from responding based on the shared information between the experimenter and the child. To test this proposal, Salter and Breheny (2019) conducted a study that aimed to lower such pragmatic difficulties for 3-year-old children. In this experiment, the experimenter asking the test question either witnessed or not the location change of the target object. Children performed better in the condition where the experimenter asking the test question did not witness the location

change, suggesting that removing the shared information in the test phase between the experimenter and the child may be a significant predictor for children's success in false belief tasks.

The studies described above suggest, that contrary to some of the signature limits indicated by the two-system account, infants and young children are able to represent beliefs about identities in various scenarios, and even before the age of 4, they are able to pass a standard explicit task, if certain processing demands are reduced. These findings provide evidence for an early developing, single ToM system, that may become more efficient through infancy and childhood given the development of other abilities (e.g., executive control, pragmatic abilities), and that young children's failures in certain tasks may be explained by factors related to high performance demands, but not necessarily a lack of ToM competence. As Carruthers (2016) states, data collected supporting the two-system account may be better explained by the distinction between tasks, that do or do not place high demands on executive control.

#### Uncharted avenues: Investigating the flexibility of children's belief attribution

Up to date, to our knowledge, the studies investigating early Theory of Mind abilities mostly tested simple scenarios involving beliefs about objects at a certain location or with a certain identity. However, in principle, adult Theory of Mind reasoning should be able to deal with any possible scenarios involving various belief contents. Virtually any kind of content, that one can entertain in first person, can be also attributed in third person belief reasoning. Since only specific contents were investigated regarding infants and young children's belief reasoning abilities, little is known whether young children, similarly to adults, are able to flexibly apply their mindreading abilities, which can operate with various scenarios, involving different contents, or as proposed by the two-system account, specific limitations would apply.

In the current thesis, we aim to investigate this issue. We will focus on knowledge structures that are well documented in early infancy, such as causal inferences required for understanding tool efficiency or reasoning about conditional relations and investigate whether they can be used in belief attribution.

#### **Beliefs about tool efficiency**

We have targeted examining children's belief reasoning abilities regarding tool efficiency for several reasons. First, the representation of tool efficiency may be more complex compared to the representation of objects at a certain location, and their examination may answer some of the theoretical questions outlined above about the flexibility or limits of the ToM system. While the representation of beliefs about objects at a certain location requires the ability to represent objects with spatial indexes, the representation of tool efficiency requires integrating an external object (tool) and embed its function in the motor coordination of actions. Moreover, it requires the representation of possible outcomes based on the actions performed with the tool. This representation is supposed to be hierarchical, suggested by neuroimaging results indicating that tool use and processing sentences with complex syntax may rely on shared neural substrates, and training in one positively impacts performance and understanding on the other one (Thibault, Gervasi, Salemme, Koun, Lövden, Boulenger, Roy & Brozzoli, 2021). Moreover, understanding the causal relations required when reasoning about the efficiency of a tool involves richer forms representations than probabilistic dependency between multiple elements. Sloman and Lanagdo (2015) argued that thinking about causal knowledge as a representation of probabilistic dependencies would leave out important aspects of causal cognition, in particular, how aspects of a certain mechanisms by which causes lead to effects, whereas it would overestimate people's access to distributional knowledge and the ability to use them in computations. In this sense, causal reasoning involves inferences about causes and effects (and is not merely an association, as it is not a representation of a correlation), in a way that they support intervention. Furthermore, "causal relations are not just stipulated but rather represent mechanisms in the world that take input (causes, enablers, disablers, preventers) and generate outputs (effects). These causal relations unfold over time such that effects cannot precede their causes" (Sloman & Langado, 2015, p. 227).

Are young children able to integrate causal knowledge in a false belief task? There are set of differences between attributing beliefs about an object's location or identity, and causal relations involving the efficiency of a tool. To attribute a false belief about an object location, the child first needs to represent the physical location of the object, and the agent's representation about the location of the object. After the false belief has been implemented (e.g., the object changed location while the agent was absent), children form an updated representation about this object's location from their own perspective, without updating the representation from the agent's perspective. In contrast, the representation of efficiency builds on the relation between several elements e.g., which intervention has to be applied on which manipulandum in order to elicit the desired outcome. Thus, false beliefs about efficiency may require more complex representations. An important question is whether the representational structure available for children is compatible with such more complex content? If children are able to integrate information about efficiency in a false belief task, it would indicate the existence of an adult-like belief tracking system that can represent beliefs with various kinds of contents and not only limited representations that can be explained by encoding registrations between agents, objects and location as suggested by the two-system account.

Previous studies described above about infants' understanding of beliefs about identities suggest that infants and toddlers are able to attribute beliefs with different contents and that they are able to take into account not only what, but how an object is being represented. However, some studies found discrepancy between infants' ability to represent beliefs about objects and beliefs about identities (e.g., Oktay-Gür, Schulz and Rakoczy, 2018). Thus, the evidence about infants and toddlers' ability to represent beliefs beyond those predicted by the two-system account seems still inconclusive. Therefore, more studies are needed that test false belief understanding in different scenarios. The first study of the current thesis aims to provide such a test case. With the experiments presented in Chapter 2 we test young children's ability to represent others' beliefs about tool efficiency.

#### Children's understanding of tool efficiency

An early understanding of the causal relations that define tool efficiency may be important for several reasons: Infants grow up in human societies and they must extract the causal structures from their environment, as well as the causal relations that are important for the understanding of human-made artefacts and their functions. A central concept in tool use is understanding efficiency. Issues concerning efficiency understanding can apply to several levels of tool use. First, one has to evaluate whether the action applied over the tool is appropriate to elicit the expected effect. Second, and more importantly, one has to evaluate whether the tool is an efficient means to achieve the desired goal. If the conditional rule capturing the relation between the manipulation and the desired effect turns to be false, then we can conclude that the artifact cannot fulfil the expected function, or at least has a limited potential of doing so. Understanding how human made artifacts, like tools work is of central interest in developmental and comparative studies. Tools are important cultural items that come into existence as the result of a chain of cumulated cultural knowledge. In order to grasp the function of a tool, the learner has to discover a set of causal structures: which actions to be used on which manipulandum in order to elicit the desired effect. Finally, for an appropriate functioning, tools should be efficient means to a goal. Children consider the physical properties of tools in in their evaluations of tool efficiency already in their second year of life (Bauer, Schwade, Wewerka & Delaney, 1999; Fagard, Rat-Fischer, and O'Regan, 2014). From around 3-years of age they can distinguish between broken and intact tools, and they can judge how these physical properties impact efficiency, which allows them to select the correct tool in order to obtain rewards successfully (Albiach-Serrano, Sebastian-Enesco, Seed, Colmenares & Call, 2015).

Based on these studies, one can assume that young children have already a causal understanding about how efficient tools impact others and their own goal directed actions. The question emerges whether, assuming that they can entertain such knowledge in first person, they would be able to attribute beliefs to others about tool efficiency.

Examining the attribution of beliefs about such structures would allow us to gain a better knowledge about the representational abilities of young children. If children are able to attribute beliefs about tool efficiency to others, such findings would provide the first evidence that they may be able to deal with more complex, relational contents as well, similarly to adults.

#### Attributing beliefs based on causal inferences

The ToM studies reviewed at the beginning of this introduction involved scenarios where children reasoned about an agent's belief based on what kind of information was perceptually available for them (e.g., seeing that an agent does or does not have visual access to a location change of an object). Situations where the available perceptual information is insufficient to predict someone's actions are frequent in everyday life. Previous research provided ample evidence that children may be highly competent in attributing beliefs based on the observable, but can they do so with beliefs that can be only inferred? Can children attribute beliefs when perceptual justification is insufficient?

To our knowledge, there is no previous research targeting children's ability to attribute beliefs based on inferences, however there is evidence from recent studies indirectly supporting this possibility. One such evidence comes from a study by Cesana-Arlotti, Kovács and Téglás (2020). Through a set of experiments, the authors investigated whether 14-month-old infants are able to apply disjunctive inferences when they are reasoning about an agent's goals and actions. Based on the results, infants were able to identify an agent's goal when direct perceptual evidence was not available, by applying disjunctive inferences. Another study, crucial to our research question, was conducted by Ting, He and Baillargeon (2021), who tested whether 5-month-old infants' ability to track what information an agent garners through inference, relying on their general physical knowledge about objects' size. It was found that infants expected an agent to search for his preferred wide object in a wide container and not a narrow one, thus, attributing the inference to the agent that wide toys can only fit into wide containers. It can be assumed, that if already infants can attribute inferences to an agent and use the result of these inferences to predict his action, older children should also be able to do so. One of our goals, thus, is to explore whether children are able to take into account an agent's beliefs and the causal inferences she may perform in order to arrive to these beliefs. A study that examines the representational repertoire that allows children to integrate causal inferences in belief reasoning would enhance our knowledge about the flexibility of the early developing belief reasoning system.

#### Updating beliefs in first person and third person

Another important aspect of children's belief reasoning abilities concerns how they reorganize their beliefs if those are in conflict with previously acquired beliefs. Children rapidly learn new information in their everyday life, and it is often the case that life provides unexpected outcomes. Two interesting questions emerge regarding how children track and resolve such inconsistencies. First, one might ask which type of beliefs they prioritize over other in case they update their belief sets, and second, how do they expect someone else to resolve such inconsistencies and how do they expect others to update their beliefs during social interactions.

Take this example: Mary knows that whenever grandma is visiting, they are going to bake some cookies. However, one day Mary learns that grandma is coming over, but mom did not get any ingredients to bake cookies. Would children revise the rule, that whenever grandma is coming, they bake together, or would they simply discard the evidence thinking that grandma is likely coming over another day? And how would they expect another person, e.g., their brother to think about such scenario? As children acquire new knowledge every day, which often involve revising their prior beliefs these questions have high relevance in infancy and early childhood. While there is ample evidence about how adults update their beliefs in case of a conflicting information, little is known whether children apply similar strategies.

In a pioneering project, Elio & Pelletier (1997, but see also Elio, 1997; Politzer & Carles, 2001) tested adults' revision strategies in a deductive framework (e.g., Elio, 1997; Elio & Pelletier 1997; Politzer & Carles, 2001). In these experiments, they asked participants to draw inferences about conditionals (by relying on premises of the form "if P then Q" expressing a conditional rule and a non-conditional information "P", that served as evidence, allowing to draw a conclusion). Then they administered a piece of information contradicting this conclusion. These studies investigated the idea of epistemic entrenchment: specifically, that
some elements of a belief set are more deserving to be retained over others in case of a conflicting information. There are two theoretical possibilities regarding which belief set is more privileged in situations as described above. One possibility is that some knowledge structures that express regularities and have a high predictive power (e.g., regularities about the world, or laws of physics) should enjoy greater entrenchment in face of contradiction, as they are more useful. The other perspective claims exactly the opposite: the data that one observes is the one that one can be more certain about, therefore it has priority to be retained, whereas regularities expressed in a conditional form are simply hypothesis about the world and can be abandoned in case of conflicting evidence. In these studies with adults, Elio and colleagues (1997) documented a general tendency to update the conditional rule and retain the observed evidence, however this strategy depended on the type of knowledge involved in the update processes (e.g., it applied to situations where the causal rule was weak, because factors preventing the effect were easily identifiable).

To our knowledge, there are only a few studies investigating children's revision strategies. While different studies targeted children's ability to update their previously acquired beliefs in case of additional, new information (e.g., Bonawitz, Fischer & Schulz, 2012; Ganea et al. 2007; Ganea & Harris, 2010), to our knowledge only one study asked specifically which update strategy children would apply in case of inconsistency and when different kinds of updates are possible. In this study, Van Hoeck and colleagues (2012) tested 7-year-old children with a counterfactual reasoning task implemented in a pretend play context. Children learned a story about heroes and thieves, and that heroes wear white hats, whereas thieves wear black hats. Later children were presented with a conflicting statement while they saw a picture of hero, and they were asked to pretend that he wears a black hat. However, in this case, most of the children selected the option to revise the data indicating that this person is a hero when they were required to resolve the inconsistency. One may note that this task, due to its pretense

context might differ from an every-day life situation and may be updated differently (as one might treat observed evidence differently in a pretense context).

Another important question regarding children's belief updating strategies concerns whether these update processes would be different in a third person reasoning. How would children resolve a conflict from a third person perspective, and when such events coincide or not with their own perspective?

We hypothesize, that as one has direct access to one's own beliefs, but others' beliefs can only be inferred, the update process in third person might be different from the ones in first person. There are three frameworks that would support such a possibility. First, the representation of others' beliefs about relational contents may be more complex than beliefs about an object location, which could lead to a more fragile representation. Thus, such a framework would predict that children will be more likely to update the rule itself (and not the data as in Van Hoeck et al., 2012) when they need to resolve inconsistency regarding attributed beliefs. A second framework with similar predictions can be derived from studies investigating the impact of co-witnessing (e.g., seeing an object together with another agent) on adults' and children's encoding of events. For instance, Gregory and Jackson (2021) found that gaze cueing results in better memory encoding for a cued target in adults. Moreover, Howard and Woodward (2019) showed that social context enhances encoding objects in memory even in infants. Based on such findings, it could be assumed that witnessing events with another person would strengthen the encoding of the observed data and would lead to a prioritization of the observed data over the rule, in case of conflicting information, as to our knowledge there is no evidence suggesting that co-witnessing would also enhance rule learning. In contrast to the first two, according to a third framework, one could predict that children would expect an agent to abandon the observed data when the agent faces contradicting information; as children might not be certain about what an agent has seen or observed, or at least he might not have encoded it in the same way as they did. As people do not have direct access to others' perception, one might encode the data from another's perspective as less certain. In this case, children would be less likely to abandon the rule, but they would revise the observed data from someone else's perspective.

#### The current studies

The current thesis aims to investigate children's belief reasoning abilities in the following ways. First, we test the flexibility of young children's Theory of Mind abilities by tasks that compared to the previously used change of location or identity tasks implement different contents in belief attribution and ask whether can children attribute beliefs about tool efficiency involving causal inferences. Second, we test children's belief revision strategies regarding simple conditional rules that are either tied to physical constraints or not. We hypothesize that children's belief revision strategies may be similar to that of adults, and that they would reorganize their beliefs in a specific manner to preserve consistency. We further tested whether children expect others to use these revision strategies as well by investigating their belief revision regarding attributed beliefs. Below we briefly summarize the studies presented in each chapter.

*Chapter 2* explores whether 3- to 5-year-old children are able to integrate efficiency information in their belief reasoning. To gain a better understanding about the flexibility of the Theory of Mind abilities available for children, different contents need to be implemented in false belief tasks. We hypothesize that similarly to adults, any kind of content that children can entertain in first person, they should be also able to attribute in third person. As children's understanding about the causal information required for tool efficiency is well documented, we predict that they are also able to attribute beliefs about tool efficiency to others. We test these

questions across 6 experiments with 3, 4 and 5-year-old children with a modified standard false belief task, in which however, children are required to predict an agent's action based on his belief about the efficiency of tools. We find that by the age three children are already able to predict the character's behavior by taking into account his false belief about the efficiency of a tool, when a reality bias is eliminated (thus processing demands are reduced). Furthermore, 5year-old children can predict the character's actions by considering the agent false or true belief about tool efficiency even in a task with a reality bias. We suggest that children's ability to integrate different contents in belief reasoning emerges earlier than indicated by some of the signature limits proposed by the two-system account.

In *Chapter 3*, we investigate whether children are able to represent beliefs about causal inferences. The majority of the false belief tasks tested children's ability to attribute beliefs based on perceptual information. The study aim is to test whether children can reason about others' beliefs, when additional inferences are required. In three experiments implemented a touch screen task, we test 3- to 5-year-old children's ability to attribute beliefs about the causal structure of a device. As previous studies with infants suggest that already 5-months-old infants may be able to attribute inferences to an agent, we predict that young children would be able to take into account an agent's false belief even when these beliefs are based on inferences, and not merely perceptual information.

In *Chapter 4* we aim to investigate how children revise their previously acquired beliefs if they experience conflicting evidence. Children acquire new knowledge with an incredible speed. Thus, they may experience conflict between their beliefs and the collected evidence relatively often. In this study, we explore whether they apply specific strategies to preserve consistency. In an online study, we asked children to predict the location of an object based on two information: the size of the object, and a rule conditional that was dependent on the size. When their predictions were contradicted by reality 5, but not 3-year-old children applied

revision strategies similar to the one used by adults: they revised the rule, but only if those rules were not physically constrained. Moreover, 5-year-old children used also tended to update attributed beliefs about constrained rules, when they detected conflicting evidence from an agent's perspective.

# Chapter 2. Can young children integrate information about tool efficiency in false belief reasoning?

### 2.1. Theoretical background

Humans possess highly sophisticated abilities to think about others' minds. In our everyday life, during social interactions, we highly depend on our understanding of what others think, know or believe, for instance, to collaborate or communicate efficiently. Most of these inferences unfold in highly dynamic situations, where one must take into account the continuous changes involving social interactions as well as changes to the non-social environment. Efficient mental state tracking requires a system that allows to integrate the information about the environment to which the interlocutor has access to in belief reasoning. Such integration comes naturally to adults: we can track others' beliefs, knowledge and preferences even when we perform rather complex tasks together, such as renovating a house with a limited set of available tools. For instance, after leaving the house the thought may strike us that our partner will likely use an inefficient tool while we are away, as we forgot to tell her that the new and more efficient tool has arrived. To avoid the unwanted consequences of her ignorance, we will urgently call her to update her knowledge.

Adults can understand, explain or predict other people's actions, even when these actions depend on beliefs and desires that they do not share or have never experienced. In theory, any thought we may think of can also be attributed to others, even if it involves highly complex scenarios and complicated inferential processes. However, it is still unclear, how such capacity develops and whether early developing Theory of Mind abilities can incorporate more complex contents, or they are restricted to more simple ones. Consider the below two scenarios:

First, imagine a young boy who has received a fishing toy for his birthday. At the end of the day, he and his mother decide to put away the toy in the cupboard. The next day, the child witnesses his father who brings some new equipment, so he needs to reorganize the cupboard, moving the fishing toy into the drawer, while the mother is not present. At bath time the child asks the mother if he could play with the toy. Would the child be able to infer that his mother has a false belief about the location of the toy, and expect her to search for it in the old location? This may seem as a relatively simple scenario, where the child needs to represent the belief of his mother about the location of the object. Once the father replaces the object to a new location, the child needs to form an updated representation about its spatial location but keep the old location representation from the perspective of his mother.

Now let's consider a more complex scenario. Imagine that the fishing toy works in such way that in order to obtain the rewards (toy fish), a magnetic fishing rod has to be used to catch the fish. The boy discovers with his mother how the fishing rod works. One day, the small magnet falls out from the rod while the child is playing, and the mother is absent. The father makes a new magnetic stick, yet the mother is unaware of these events. Next day, the mother invites the child to play and looks at the two rods. Would the child expect his mother to try to use the old but inefficient rod, or would he expect her to approach the new and efficient tool? Would the child realize that the mother does not know about some crucial events and inform her about which rod is working and which is not before they start playing? In this scenario, compared to the previous one in which only a simple location change happened, the child first has to attribute specific knowledge to the mother regarding the functioning of the toy, specifically a relation between an object, an action and the outcome: *if* the fishing rod works, *then* the fish will be obtained. Next, he updates his own beliefs in a way that this rod is no longer functional, therefore a new rod is necessary, but should keep a representation about his mother's false belief about the functionality of the toy.

In short, the above scenarios involve different sets of computations: in the first scenario, one can attribute a false belief about a toy's location, while in the second scenario, one has to attribute a false belief about the efficiency of a tool. Would children be able to draw the correct inferences about other's mental states in both cases?

#### Theory of Mind abilities in infants and young children

Earlier studies have provided ample evidence regarding when children can explicitly predict other's behavior based on their (false) beliefs about the location of an object. Most of the evidence comes from a location change false belief tasks that have been traditionally used to measure ToM in children (Wimmer & Perner, 1983; Baron-Cohen, Leslie & Firth, 1985, Wellman, 2001). The task involves two characters, e.g., Sally and Anne and a situation where one of the characters has a false belief. In this task, the child is first introduced to a character (e.g., Sally) who puts her marble in a specific location (e.g., basket) and later leaves the scene. While she is away, a second character (Anne) replaces the marble from the first location to a second one (e.g., box). Upon Sally's return, the child is asked to predict where Sally will look for her marble. Given its widespread use, this task became the so-called standard false belief task.

In the last 35 years, studies have converged towards that children tend to succeed on such tasks after the age of four, whereas younger children fail, and upon Sally's return they indicate the marble's actual location (for a review see Wellman, Cross, Watson, 2001). Based on such findings it was assumed that children undergo a shift in their representational abilities around the age of 4 (e.g., Gopnik, 1993; Wimmer & Weichbold, 1994), in the sense that only at this age they become able to represent others' mental states.

This conclusion has been challenged by experiments using modified tasks to demonstrate false belief understanding earlier in development. Some of these tasks do not require answering a direct test question about the behavior (or the belief) of the agent, but they assess young children and infants' understanding about an agent's false belief differently, for instance via interactive helping paradigms (Buttelmann, Carpenter & Tomasello, 2009) or via measuring their looking behavior (e.g. Onishi, Baillargeon, 2005; Kovács, Téglás, Endress, 2010) or brain responses (e.g. Southgate & Vernetti, 2014; Kampis, Parise, Csibra & Kovács, 2015). Other studies have aimed at reducing the processing demands of the task (e.g., Scott & Roby, 2015, Surian & Leslie, 1999; Rubio-Fernandez & Geurts, 2013; 2016). Studies using implicit measures revealed sensitivity to false beliefs in infants already in the first year of life (Onishi & Baillargeon, 2005 as the very first evidence for 15-month-old infants but for a review including younger age see Scott & Baillargeon, 2017, and see Poulin-Dubois, 2018 as a reply to a commentary).

The great majority of studies focusing on how Theory of Mind abilities develop have tested scenarios where an object is moved to a certain location unbeknownst to the agent. Thus, these studies have targeted belief contents of a certain type: attributing beliefs about the location of an object. Contents other than objects at certain location were seldom tested in belief attribution tasks with children. There are some remarkable exceptions, as also discussed in details in the general introduction: for instance, some studies tested young children's false belief understanding regarding object identities or unexpected contents (Scott & Baillargeon, 2009; Buttelmann, Shurke & Buttelmann, 2015, Low & Watts, 2013; Fizke, Butterfil, van de Loo, Reindl & Rakoczy, 2017; Moll, Khalulyan & Mofett, 2017), or false belief about the traits or emotions of agents (Choi & Luo, 2015; Smith-Flores & Feigenson, 2021). That is, while it has been widely investigated whether Theory of Mind abilities are present at certain stages of the

development, little is known about how flexible these representational abilities are, and whether young children can integrate different belief contents when reasoning about other minds.

Our project aims to chart whether early ToM abilities are flexible enough to go beyond tracking beliefs about objects' presence at a certain location and integrate more complex content in belief attribution. While it is natural to suppose that adult belief reasoning can deal with all possible belief contents, it is still an open question whether young children are capable of integrating various contents (that they themselves can understand) in social situations, when predicting and explaining others' actions. For instance, as described in our example in the beginning of this chapter, if a young child understands how a fishing tool works, can he attribute such inferences to his mother regarding the efficiency of the tool, and sustain them although they will conflict with what he has experienced (that the efficiency has changed)?

#### How infants and children represent beliefs? Some theoretical possibilities

One can argue that the cognitive prerequisites of such a reasoning system imply an underlying representational structure that supports the integration of a broad range of contents. Whether initial belief reasoning is restricted to certain domains and whether it follows a developmental path from being able to ingrate a narrow set of contents to a broader set is reflected in antagonist theoretical positions that were advanced in the literature.

As outlined in the introductory chapter, to date, there are two main lines of theoretical proposals concerning early Theory of Mind abilities, one arguing that children undergo a representational change and can reason about other's beliefs only around the age of 4 (e,g, Perner & Ruffman, 2005; Apperly & Butterfil, 2009, Butterfil & Apperly, 2013; Low, Apperly, Butterfil & Rakoczy, 2016), and the other arguing that these representational abilities are

present early on though they might get more sophisticated across development (together with other cognitive abilities) (e.g. Carruthers, 2013, 2016; Scott & Baillargeon, 2017).

Based on the representational change accounts, infants early on do not possess ToM abilities, and are able to succeed on certain false belief tasks due to different reasons other than representing an agent's false beliefs. For instance, some non-mentalistic accounts claim that infants succeed on implicit tasks because they form associations between the agent-object and its location and are surprised if they consequently see a pattern that does not fit to this association (Perner & Ruffman, 2005). According to the two-system accounts, infants possess a so-called minimal Theory of Mind in the sense that they are able to track belief-like states, but not beliefs *per se* (Apperly & Butterfil, 2009; Butterfil &Apperly, 2013) These theories state that a shift occurs in the preschool years, when children are able to pass the so called traditional false belief task and become able to represent others' beliefs in various scenarios.

In contrary, according to one-system or continuity accounts, false-belief reasoning emerges early in life and becomes more efficient in the course of development. According to this view, infants and young children's difficulty with certain false belief tasks are more related to processing demands than the lack of belief reasoning abilities.

In the following sections of this chapter, we will first analyze some of these theoretical proposals regarding infants' and young children's belief reasoning abilities. We argue that different belief contents should be tested in false belief tasks to better understand the flexibility of the early developing belief reasoning abilities. If various contents can be integrated in in belief reasoning from early on, it would support a single Theory of Mind system efficiently operating in various scenarios.

#### The two-system account of Theory of Mind

As discussed earlier, research from the past decades investigating early Theory of Mind abilities yielded positive findings with infants and young children (for a review see Scott & Baillargeon, 2017; Scott, Roby, Baillargeon, 2022;) but see a review about non-replications see e.g., Kulke & Rakoczy, 2018). Most of these studies suggest, that if the commonly used false belief task (e.g., the Sally Ann task) are modified in certain ways, infants and young children show success in representing an agent's false belief. In contrast, explicit false belief tests suggest that children under the age of 4 are incapable to represent others' beliefs (Wellman, Cross & Watson, 2001). One of the attempts to explain this discrepancy was the two-system account developed by Apperly & Butterfil (2009). Based on the two-system account, the early mindreading system is unable to represent false belief per se and is only able to track beliefs regarding certain contents, e.g., objects at a certain location. They call this first system the minimal Theory of Mind. Only the later developing system, which is the full-blown Theory of Mind also used by adults "involves representing propositional attitudes like beliefs, desires, and intentions to construct reason-giving, causal explanation of action." (Butterfill & Apperly, 2013, p. 607). They state that infants' success on the modified tasks can be explained by the Minimal ToM, and later, with the development of the second system, children become able to pass the false verbal false belief task after 4 years of age.

The early developing system is described to be fast and automatic, but inflexible, whereas the later developing system is slower and effortful, but also more flexible. The early developing system can track goals, perceptions and belief-like states, but importantly, without representing them as such.

Besides infants being able to attribute goals to agents (e.g., Király et al. 2003; Luo & Baillargeon, 2005; Buresh & Woodward, 2007; Csibra, 2008) they are argued to track

*registrations*, based on which they are able to interpret and predict an agent's action in certain (non-verbal) false belief tasks. Based on this theory, registrations only capture the relation between an agent, the object and its location. However, the representational repertoire of this system is specialized to a certain input domain (e.g., agent-object-location relations). Any content that is incongruent will fall beyond the postulated representational capacity. Importantly, according to this proposal only the later developing, full-blown Theory of Mind will be able to grasp propositional structures. Therefore, children's early developing Theory of Mind abilities are limited to tracking only certain belief contents. Based on this account, infants and young children should not be able to track an agent's belief about how (under which aspect) an object is represented, or for instance, what function one believes an object to have.

On the contrary, if infants and young children are able to integrate various kinds of complex contents in their belief attributions, the question emerges how should a representational apparatus, which makes this possible, look like. There is no doubt that adults possess sophisticated abilities to infer others' thoughts and beliefs independently of their contents. However, it is a remaining question whether children from early on can integrate different contents when reasoning about others' beliefs. Furthermore, what representational structure would allow them to infer and update beliefs on various contents in an efficient and fast manner, as it is required in everyday life? And what are the core characteristics of such an apparatus that may be available already for young children?

## Theories describing an early developing flexible belief reasoning system

Proponents of the other main theory that was advanced in the field suggest that infants from early on may possess full-blown Theory of Mind abilities. We describe here three proposals that also aim to characterize the representational structure of beliefs that are flexible enough to deal with various contents. One of the earliest theories dates to Leslie, 1987. By taking as a case study the analysis of the processes explaining children's pretend play, Leslie put forward a proposal to explain how young children may deal with multiple conflicting representations attributed to the self and others. According to this proposal, in pretense, to avoid representational confusion, pretended representations need to be detached from primary representations. For instance, in pretense, the expression of a banana as a telephone is detached from banana as a banana as primary representation ("direct semantic relation with the world", Leslie 1987, p. 414). These detached representations are linked to primary representations via an informational relation that has a computational function, relating together agents, decoupled expressions, and primary representations. Leslie's main suggestion was that the representational structures underlying pretense are similar to the ones necessary for computing other's belief representations. For instance, if we go back to the fishing example described at the beginning of the chapter about the false belief on the efficiency of a tool, the child will hold a primary representation about the tool which has become inefficient, and in order to avoid representational confusion, he will also hold a detached expression representing her mother's false belief (which is about the tool being efficient). In sum, Leslie proposes a three-term relation between the agent, decoupled expression, and a primary expression: in this case, the agent (mother) her decoupled expression (efficient rod) and primary expression (inefficient rod). This three-term relation can be also applied in general to false belief reasoning: to the relation between the agent as the belief holder, the belief content, and the actual situation.

Another theoretical description of an early developing, flexible belief reasoning system comes from Baillargeon, Scott and He (2010). The authors assume that infants from early on are equipped with a psychological reasoning system to interpret others' actions by attributing them mental states, and that this system involves the operation of two subsystems. Subsystem 1 enables infants to attribute motivational and reality congruent states to agents and is present already from the first months of life. Subsystem 2 enables infants to attribute reality incongruent states such as false beliefs with the help of a decoupling mechanism that allows infants to maintain two different representations about the world – one from the agent's perspective that might be different from the reality, and one from their own perspective which is congruent with reality. System 2 allows infants to predict and interpret an agent's action based his beliefs about various contents, and it is operational already from the second year of life. Based on this theory, these two systems are present early on, and operate in a parallel manner.

The third theory aiming to tackle on the representational structure supporting belief attribution comes from Kovács (2016). Kovács suggests that the basic representational structure, the belief file would provide the foundation for efficiently encoding and updating information about others' beliefs during social interactions. Central to this proposal is a multicomponent system, that - similarly to Leslie - emphasizes the role of the variables for the agent as the belief holder and a variable for the belief content that are sustained in parallel with one's own representation of the real world. Furthermore, for achieving the required flexibility the structure of the belief file allows its subcomponents to be separately updated in order to help fast and efficient belief encoding and updating. They allow belief reasoning even if their content is underdetermined (Kovács, Téglás & Csibra, 2021) Thus, in principle, such representational structures would allow encoding and updating beliefs with complex relational contents. For instance, updating beliefs about efficiency in a manner that certain components from the relational content that need no update (e.g., the relation between the tool and the elicited effect) would be preserved whereas others would be updated (e.g., the change regarding the efficiency).

To summarize, crucially to our question, the theories proposed by Leslie (1987), Baillargeon and her colleagues (2010) and Kovács (2016) suggest that various contents can be integrated in false belief reasoning, independently from the complexity of the belief content. Thus, if young children are able to integrate different contents when reasoning about others' beliefs, it would support a one system Theory of Mind, and could be described by the representational structures proposed by the theories above.

To better understand young children's belief tracking capacitates, it becomes a crucial question whether they can integrate information on various contents in false belief reasoning. Since predominantly specific belief contents were investigated in studies involving young children, it is unclear whether early in development, children, similarly to adults, can flexibly integrate diverse contents in their belief reasoning or as proposed by the two-system account, specific limitations would apply. In the studies presented in this chapter, we plan to investigate this issue.

Let's consider our first example, a case when a child's toy becomes inefficient in the absence of her mother. Initially, the child knows that *if* he is using the magnetic rod, *then* he will be able to get the rewards from the toy fishbowl. Therefore, he holds a representation that whenever the manipulation (stick with a magnetic rod) is present, the effect (successful catching) can be elicited. The child also knows that his mother holds the very same representation regarding the toy. However, in the absence of his mother, the rod would brake (e.g., the magnet would fall out from the stick) which would result in an updated representation from the child's perspective: the manipulation of the original rod can no longer elicit the desired effect. However, he still needs to hold a non-updated representation about her mother's false belief: that she believes that the intervention would still elicit the effect. Both representations should be kept in a way that they should not lead to a confusion regarding the child's representation of the world: even though he holds a representation regarding her mother's false belief (that the rod is efficient), and a first-person representation (that the rod is not efficient) he should not be confused about the efficiency of the rod.

Based on the theoretical proposals suggest by Leslie (1987), Baillargeon and colleagues (2010) and Kovács (2016) even the early belief reasoning system should be flexible enough to track and update beliefs with any kind of contents. In contrast, if the early belief reasoning capacities are not flexible enough, young children should not be able to track beliefs about different contents, and this would support accounts suggesting two systems for theory of mind reasoning, where the early developing system cannot deal with beliefs with complex contents.

#### Representing a variety of contents: Beliefs about tool efficiency

We aim to test young children's ability to integrate their understanding of tool efficiency in false belief reasoning. We targeted tool efficiency understanding as it is relatively well documented in the literature that children by their 3 years of age possess an understanding of what properties make a tool efficient (e.g., Bauer, Schwade, Wewerka & Delaney, 1999). Our first target is integrating causal information necessary for tool use in belief reasoning. Understanding causal information in tool use from early on may be an important skill for several reasons. Infants growing up in human societies have to extract the causal structures from their environment, as well as the causal relations that underlie tool use.

In order to encode the functioning of different tools, one has to have a good grasp of the relation between a manipulation and an effect: for every well designed and well-functioning tool each time the appropriate manipulation is present, the expected effect will be elicited. Efficiency of the tools, however, may be perceived as a categorical attribute (i.e., efficient/inefficient, when a short stick cannot reach a distant object, while a long stick can), or it can be graded (i.e., flexible stick, that is long enough to reach a specific distant object, thus its expected utility is a function of its elasticity), and it may call for different representations (i.e., probabilistic dependency between multiple elements). Either way, representation of tool

efficiency requires a causal understanding of the relation between an object's function, the possible intervention and outcomes, the considerations of the possible preventing factors when completing an action to achieve a goal with the tool.

Attributing a representation about tool efficiency may be more complex than attributing representations about object location. Beliefs about object location may be described as representations of spatial locations. In contrast, forming a belief about tool efficiency requires the integration of an object, its use and the possible outcomes and embed its functional structure in the motor coordination of actions, possibly resulting in a hierarchical representation, which will have to be further integrated in belief attribution.

Examining the attribution of beliefs about such structures would allow us to gain a better insight about the representational abilities of young children. Evidence that children are able to attribute beliefs about tool efficiency to others would indicate that they can go beyond the signature limits suggested by the minimal ToM approach, and it would suggest that children from early on may be able to deal with more complex, relational contents as well, similarly to adults. Before turning to the experimental part where we aim to directly target these questions, we will review research that has targeted tool efficiency understanding in young children.

#### Young children's understanding of tool efficiency

The efficiency of a tool in achieving the desired effect is often defined by a set of physical properties: for instance, its shape, length, continuity, or rigidity. A stick that is too short or too flexible will be inefficient in moving a target object closer. A rope that is broken is inefficient to pull a food item closer. A short tool may be inefficient for extracting objects from a container. When do children begin to understand such physical properties and become able to select efficient tools for certain actions?

Some studies (Bauer, Schwade, Wewerka & Delaney, 1999) showed that children are able to analyze cause-effect relations between an action on an efficient or inefficient tool and its outcome from around their second birthday: they are able to anticipate the outcomes of these actions and plan their behavior accordingly. A longitudinal study conducted by Fagard, Rat-Fischer, and O'Regan (2014) has investigated the development of the learning mechanisms that underlie tool use. They found that infants from early on have a basic understanding of how to use a tool properly in order to obtain a reward; for instance, 12-month-olds already know that they can move one part of a rigid object by moving another part; 18-month-olds can use a rakelike tool to obtain an out of reach object.

Various studies conducted with non-human animals and children investigated the understanding of continuity of objects in tool use (e.g., broken vs. intact tools, Albiach-Serrano, Sebastian-Enesco, Seed, Colmenares & Call, 2015). These studies mostly focused on whether non-human animals and children understand more abstract physical properties that contribute to the efficiency of the tool (for instance that a broken tool will be inefficient to obtain a reward). Results from these experiments suggest that by 3 years of age children understand the aspects of physical differences between broken and intact tools and how they impact efficiency, and this allows them to select the efficient tool in order to obtain objects. Not only children, but non-human animals such as capuchins and great apes are also sensitive to these properties of tools.

To summarize, these results indicate that children from around their third birthday have a relatively good understanding about what physical properties contribute to the efficiency of a tool.

In our experiment, we developed a task involving tool efficiency that requires the understanding of what is the sufficient length of a tool to obtain a reward that is out of reach. We aim to investigate whether children are able to integrate information about efficiency when they reason about others' beliefs.

# Study 1: Can children integrate efficiency information into their belief reasoning abilities?

Our main research question was addressed in a set of experiments investigating 3-, 4and 5-year-old children's ability to integrate efficiency information in false belief reasoning, using a behavioral paradigm. In these experiments, children are required to predict a character's behavior who has a false belief regarding the efficiency of the tool available to obtain rewards. Specifically, the tasks implement a change in the object's functionality (an efficient tool becomes inefficient, and the character is not aware of this), but not in its location.

The current experiments would allow us to gain a better understanding whether young children's mindreading abilities are flexible enough to integrate various contents when reasoning about others' beliefs.

#### Aim of the present study

The aim of the present experiments is to investigate young children's ability to represent beliefs about efficiency information required for tool use.

There are a set of differences between attributing beliefs about an object's location or identity, and causal relations involving the efficiency of a tool. To attribute a false belief about an object's location, the child holds a representation from the agent's perspective about the location of the object. After the false belief manipulation has been implemented (e.g., the object changed location while the agent was absent), children can form an updated representation about this object's location from their own perspective, without updating the representation from the agent's perspective. An important question is whether the representational structure available to children is compatible with more complex contents? As discussed earlier, it can be assumed that representations of false beliefs about efficiency are more complex simply because they may require representing the relation between several elements e.g., which intervention has to be applied on which manipulandum in order to elicit the desired outcome. If an initially learned cause-effect relation becomes false (while we still attribute it as someone's belief content), one needs to update these relations according to the change of the causal structure yet keep a not-updated representation for the agent.

If children are able to integrate information about efficiency in a false belief task, it would indicate the presence of an early developing Theory of Mind system that can represent beliefs about various kinds of contents. This system is not only limited to representations that can be explained by the minimal Theory of Mind account, as in this case children are required to track an agent's representation about the function of the object, and not simply his registration about the location of the object.

In one of our experiments, children observe a plush bear who has to achieve a goal: he has to fish out his favorite balls from a long tube using sticks ending in a hook. The first stick is long enough only to catch the balls on the top of the tube, but not the next ones, therefore, to obtain all his balls he uses a second, longer and efficient stick. In the false belief condition, the bear places the two sticks in two different boxes in a way that only its ending is visible and leaves the scene. Then, the experimenter breaks the long stick and reattaches its ending to the short stick, thus making this previously inefficient stick efficient, however the bear cannot see this. In the true belief condition, these events take place in the presence of the bear. In test phase, children are required to predict where the bear will search for his stick.

Since in our studies we aimed to test also 3-year-olds, we have considered tasks using a false belief procedure with reduced processing demands, which showed false belief understanding in children even under 4 years of age (e.g., Scott, Roby, Setoh, 2020; Scott &

Roby, 2015, Surian & Leslie 1999, Rubio-Fernandez & Geurts, 2013, 2016). Therefore, we tested children with an alternative procedure of the standard false belief task, based on Rubio-Ferndandez and Geurts (2013) who modified the procedure in two ways that led to successful performance in three-year-olds. First, the authors implemented specific prompts before the test trial regarding the character's perspective: when the character left the scene, children were asked whether he can see them from where he is now, allowing children to keep track of the character's perspective. The same question was asked again after the location change (in the scenario) was implemented. Second, instead of using the standard test question "Where will Sally look for her banana?", which would generate high inhibitory demands (as the object is explicitly mentioned and children have to inhibit the current location of the object, e.g. Carlson, Moses, Hix,1998), the authors used an act-out, open question ("Now it is your turn in the story: what will happen next?"), considered to be easier for children at this age compared to the standard test question. Different conditions in this experiment showed that using both the perspective tracking prompts and the open question were necessary for children's successful performance. In case one of them was not included, children's performance fall below chance.

In our study, we incorporated elements of this experimental protocol, that may allow children below 4 years of age to pass a modified version of the standard false belief task. However, we aimed to address the following theoretical question: can children predict a character's action based on his (false) belief about tool efficiency? Our task involved a change of the efficiency of an object (the tool became inefficient), while the central character had false belief about the efficiency of the tool that he would use to obtain a desired goal.

We tested children across three age groups: 3, 4 and 5 years of age. Given the novelty and complexity of the current task, we began by testing 4-years-olds who typically pass the standard location change false belief task.

#### 2.2. Experiment 1. Beliefs about tool efficiency: the case of 4-year-olds

#### 2.2.1. Methods

#### **Participants**

Unless stated otherwise, each experimental group of children involves 24 participants. This sample size is initially chosen arbitrarily, but typical for the field. Considering binomial test, it allows detecting large effects (d=.6 for  $\alpha$ =.05;  $\beta$ =.80)

Twenty-four children (12 females) participated in the study. Their age ranged from 4 years 1 months to 4 years 11 months (M=4 years 5 months). Additional 11 children participated but were excluded from analyses either because they failed to give a correct answer to one of the post-test questions (described in the procedure) (N=10), or because of parental intervention (N=1). Participants were recruited and tested at the Central European University's Child Lab at the Budapest Zoo. Parents gave their informed consent for participating in the study. Participants received a small toy as a gift, independently of their performance. This research received ethics clearance by the United Ethical Review Committee for Research in Psychology (EPKEB) in Hungary.

#### Apparatus

The apparatus consisted of a transparent plastic tube (35 cm long and 8.5 cm diameter wide) and 8 colored plastic balls placed inside the tube (4 cm diameter wide) that had 0.5 cm wide holes on their surface so that a hook could be inserted in the holes. We used two wooden sticks: one of them was short (27 cm long), and the other was longer (49 cm long). Each stick had a colorful hook ending (approx. 5 cm long and 1.5 cm large) allowing the user to extract the balls from the tube. The longer stick served as the *efficient tool*, because it allowed collecting

every ball, including those that laid at the bottom of the tube. In contrast, the short stick was an *inefficient tool*, as its use did not allow for extracting the balls that lay at the bottom of the tube. The sticks were introduced sequentially in the appropriate phases of the experiment. Two coloured boxes, a green and a blue (47 x 2 x 1 cm each) served as stick holder boxes. Once the balls were collected the experimenter stored them in a small bag.

The events also involved a polar bear character (a plush hand puppet). The experimenter used this character to introduce the true or false belief scenarios. Central to each scenario was the bear's intention to collect the balls from the tube.

#### Procedure

Children were tested in a separate room in the Zoo. The whole procedure lasted around 10 minutes. The child was seated in front of the experimenter on a small chair, and the parent(s) were seated in the corner of the room and were asked to not to intervene. The session started with a warm-up phase where the experimenter played a game unrelated to the experiment that involved a shopping scenario, wooden vegetables and the bear. This phase had the role of warming up the child in a turn-taking game where children were asked to buy vegetables with the puppet from the experimenter.

The experimental task had two phases: familiarization and test. First, in the familiarization phase we introduced information about tool efficiency and then we tested whether our participants could integrate this information into false belief attribution. The experiment was followed by control questions checking children's memory of the events and their efficiency understanding. We describe these phases of the experiment in turn.

#### **Familiarization phase**

First, the participants were familiarized with the experimental apparatus. The transparent tube with the 8 balls inside was placed at the furthermost edge of the table from the child, in the middle. The stick holder box and the two sticks were placed under the table out of view from the participant.

The experimenter began to demonstrate how the apparatus works involving the puppet; "Now let's play a game! Let the bear show you first how does this game work. These are the bear's favorite balls, but unfortunately, now they fell into this tube, and he can't get them out because his paw is too short" At this moment, the experimenter demonstrated that the puppet tried to get his balls with his paw, but it's too short to reach for them. Then, she introduced the sticks:

"Fortunately, the bear has some sticks! For example, there is a stick here. Let's see if bear manages to get some balls out with this one!" At this point, the experimenter with the puppet in her hand, picked the first, short stick, approached the tube, and extracted two balls, one after another. Then he (the bear) tried to reach for the rest of the balls, however, the length of the stick was insufficient to reach for the remaining balls. The experimenter continued the story: "the bear would like to get all the balls out, but this stick is too short for that. Luckily, there is another stick around here..." At this point, the experimenter placed the short stick in one of the two coloured boxes in way that only it's ending was visible (approximately 5 cm) and the demonstration followed with the second, long stick: "This one is long enough! I'm sure the bear will be successful with this one too." And he pulls out two more balls with the long stick. Figure 1 demonstrates the set up and the phases in the familiarization phase.



*Figure 1*. Familiarization phase. Demonstrations with the short and the long stick: after collecting certain objects with the short stick (b) the increased distance of the remaining objects from the opening of the tube turns the short stick an inefficient means to achieve the goal (c) while the long stick guarantees the successful completion of the task (d).

#### **Test phase**

After the familiarization phase, the experimenter said that now the bear has something to do, so he is leaving for a short walk. The bear placed the long stick (while the experimenter commented "The bear is going to put his stick now here") into the other coloured stick holder box in the same way as the short stick was placed and put the box on the table opposite from the box with the short stick. After the locations of the tools were established, she walked away with the puppet, and placed him in the corner of the room facing the experimental scene with his back, approximately 1.5 meter away from the scene. To help the child keep track of the puppet's perspective (as in Rubio-Fernandez & Geurts, 2013), the experimenter went back to the child and asked the child: "Can the bear see us from there?" If the child did not reply or gave a wrong answer (e.g., yes, he can see us) the experimenter corrected the child and provided

the right answer ("He is not able to see us from there"). As in the original study, this was not a check question, but rather just a prompt to emphasize the knowledge state of the protagonist.

To implement false belief, the experimenter told the child that they will play a trick on the bear, and at this point, she took the long stick out of the box and broke it into two pieces (at its half-so that now it became as long as the short stick) and placed it back in the stick holder box in the same way as the original long stick was placed, so that its ending was only visible. She commented on her action using a secretive voice: "Let's play a trick on the bear! I'll take the long stick now, break it (shhhh), so now it is short! And now, let's put this short one in the other box, in the same way as he has left the long one." In addition, the experimenter took the other ending of the broken stick, opened the box with the stick that was short from the beginning, took it out, and attached the ending of the long stick to the short stick with a help of scotch tape. Then she commented: "Look! Now this stick became long!" and she placed back the new long stick in the box.

Then the experimenter asked the child again: "Did the bear see what we did?" Again, if the child did not reply or did not give the right answer, the experimenter answered, "He must not have seen us from there". The set up of test phase is illustrated in Figure 2.



*Figure 2.* Test phase illustrating the efficiency manipulation: the puppet placed both sticks in boxes then left the scene (e) in his absence, the long stick was broken by the experimenter (f-i) and the resulting part was taped to the short stick (f-ii). Thus, the originally long stick became short, while the originally short stick became long (g).

The experimenter then brought back the puppet to the scene and gave it to the child, asking her to continue the story: "It's your turn now, what will happen next? What will the bear do?" If the child did not reply to this open question for 10 secs, the experimenter asked a second question while giving some cues: "What will the bear do if he wants to get the balls out? Where will he search for the stick?"

The colour of the boxes (blue/green) and the sides were counterbalanced across children.

#### **Post-tests**

In order to guarantee the validity of our dependent variable, we have to be certain that the participants reliably encoded the necessary information. We checked their memory for the location of the tools and the correctness of their efficiency judgments. Participants who did not remember correctly where the originally long stick was placed (N=9) or did not answer correctly to the efficiency question (see below) (N=1) were excluded from the analyses.

#### Memory

To make sure children's answers are not affected by memory issues, (for instance, they might not remember where the original stick was placed) after the test phase, children were asked two memory-related questions: first, the experimenter asked where the puppet left the originally long stick before breaking it, and then asked where the puppet left the originally short stick. The memory post-test questions helped us to make sure that children encoded the original location of the sticks, which was necessary to represent the puppet's false belief correctly. Participants whose response was incorrect for the second memory question regarding the originally short tool's location were not excluded (N=3), as the first memory question may have biased their answer and in any case remembering where the originally shot stick was placed is not a precondition for correctly computing a false belief.

#### **Efficiency judgments**

To ensure that children had sufficient understanding of the efficiency of tools involved during the experiment, after the memory questions children were asked which stick they would pick to pull out the remaining two balls, and a new long stick that was as long as the originally long one and the originally short stick were placed in front of the child. The post-test question targeting efficiency aimed to make sure that children understood correctly the level of efficiency provided by the initially available sticks, and to ensure that children integrated their correct understanding of efficiency in their false belief reasoning.

#### 2.2.2. Results

We coded children's answers from the point where the puppet was brought back to the scene and the experimenter handed it to the participant and asked the open question. Children could indicate one of the boxes placed in front of them, either by i) taking the puppet there (e.g., pretending that the puppet will approach that box), ii) pointing at, or iii) verbally indicating one of the two boxes. Only these three kinds of behaviors were considered as valid answers. However, only 9 of 24 participants gave a valid answer to the first, open question, by indicating one of the boxes. From the rest of the sample a response was elicited by the cue giving questions (What will the puppet do if he wants to obtain the remaining balls, where will he search for the stick?). This question was similar to the one used in the standard ToM task ('Where will the puppet search for the object?'), however in this case it did not involve a possible bias towards the actual location of the object that has to be overcome so an empty location can be indicated, as it was found in earlier studies (e.g. Rubio-Fernandez & Geurts, 2013, 2016) because both locations contained a stick. Therefore, we assumed that it will not pose any additional demands on children's performance. Children's answer was coded valid if they reacted to any of the questions in a way described above and given that we had only 9 children giving valid responses for the open question their answers were aggregated across the two types of questions (open, or standard question). Note that a child could have a valid response either from the open or from the standard question, but not from both, as the standard question was not asked if the child provided a valid response to the open question already.

The below analysis is performed on these aggregated scores. Fourteen (5 in response to the open, 9 to the standard question) out of 24 children (58%, p = .541, binomial) answered correctly, by indicating the box that originally contained the long stick. Thus, in this first experiment, children did not predict the puppet's behavior based on his false belief above chance level. In addition, we calculated binomial Bayes Factors (BF) contrasting the null hypothesis (equal probability of indicating the originally long and the new long stick) to the alternative hypothesis of higher probability for the correct answer (originally long stick) (using the default hyperparameters in JASP), and the estimated bayes factor did not favor the null hypothesis (BF=2.922).

#### 2.2.3. Discussion

Children's chance performance observed in this experiment may be related either to their difficulty to integrate efficiency in false belief resoning, or alternatively, to the fact that they should have inhibited the correct answer (the newly built, efficient stick) from their own perspective, and they were not very successful in that. This latter possibility would be in line with previous experiments showing that children may have difficulties to inhibit such reality bias, and this may interfere with their ability to answer correctly (Birch & Bloom, 2003; Leslie, German, & Polizzi, 2005; Mitchell & Lacohee, 1991). Mentioning the target object here probably did not impact children's performance, because both boxes contained a stick (the broken which was originally long and the newly built long one). Note that compared to the study conducted by Rubio-Fernandez and Geurts, in the current study, the act-out open question elicited responses only in 9 4-year-old children, however, also note that the current task was likely more complex. Most of the children seemed puzzled or were uncertain what to do, and only a few answered immediately. The rest of the group's answer was elicited by prompts similar to the standard question. This might be explained by the fact that here the open question

was slightly more ambiguous compared to the location change scenario used by Rubio-Fernandez and Geurts (2013). Once the puppet came back and was handed to the participants, children may have assumed that he simply wants to continue to play the game and fish out the balls, and therefore were confused how to react. Once some prompts were given (e.g., the puppet wants to collect the balls, where he will look for the stick?) most of the children gave relevant answers.

To investigate whether a possible reality bias may have influenced children's performance, we tested a second group of 4-year-old children in a modified version of the task where the reality bias was eliminated in a way that children were simply required to keep track of the perspective of the agent, but there was no correct answer from their own perspective, therefore no inhibition was required (no-reality bias version)

# 2.3. Experiment 2. Can 4-year-olds represent beliefs about efficiency when reality bias is eliminated?

To test whether 4-year-old children's poor performance was related to their inability to infer others' beliefs regarding tool efficiency, or to the processing demands posed by a reality bias, we conducted a second experiment. Given the structure of the task, it is possible that children were simply unable to inhibit their own perspective regarding the efficient tool, and thus this interfered their ability to correctly predict the character's behavior based on his false belief. We modified our first experiment in a way that there was no correct answer from the child's perspective in the test phase. This modification allowed us to test the prediction that 4year-old children might be able to integrate information on efficiency in false belief reasoning when inhibitory control demands (that are unrelated to false belief understanding) are reduced.

#### 2.3.1 Methods

#### **Participants**

24 children (10 females) participated in the study. Their age ranged from 4 years 1 month to 4 years 11 months (Mean age=4 years 5 months). Additional 14 children participated but were excluded from analyses either because they failed to give a correct answer to one of the post-test questions (N=7) or failed to give a relevant answer to the false belief test question (N=4) or because of parental intervention (N=1) and experimental error (N=2).

### Procedure

The same procedure was repeated as in Experiment 1, with one modification. In order to reduce reality bias, we changed the test phase in the following way: when the experimenter broke the long stick in half instead of attaching the ending to the originally short stick, she removed the ending from the scene. As a result, there were only two short sticks in the two boxes. We assumed that in this way the reality bias was eliminated, as there was no correct answer from the perspective of the child, about how to get the remaining balls. The test phase without reality bias is demonstrated in Figure 3.



*Figure 3*. Test phase without reality bias: The puppet leaves the long and the short sticks in two boxes, and in his absence, the long stick becomes short

# 2.3.2. Results

Children's answers were coded as valid as in Experiment 1, if they indicated one of the boxes by either pointing or verbally to any of the questions (open ended N=10 or standard question N=14).

In Experiment 2, children successfully predicted the puppet's behavior based on his false belief: Upon the puppet's return, 20 (open N=10, standard question N=10) out of 24 children (83%, p=.001, binomial) predicted that the puppet would search for the stick in the box where the originally long stick was left (and now became short) and indicated this box verbally or by pointing. Estimated Bayesian Factor strongly supported the alternative

hypothesis (BF= 126). The graphs representing children's performance in Experiment 1 and 2 is shown in Figure 4.



*Figure 4.* 4-year-old children's performance (correct answer to the false belief test question) in Experiment 1 (reality bias) and 2 (no reality bias)

For Experiment 1 and 2, we analyzed children's choice (box congruent/incongruent with the puppet false belief) using a binomial GLM. We included as predictor variables experiment (Experiment 1 and 2, FB test with and without reality bias) and their choice as outcome (correct or incorrect box). Children in the no-reality bias experiment tended to choose correctly more often the box which was congruent with the puppet's false belief (est.=1.27, se.=.69, p = .06) compared to the other experiment with reality bias.

#### 2.3.3. Discussion

Results from Experiment 2 indicated that once the reality bias is eliminated, 4-year-old children can successfully predict an agent's behavior based on his false belief about the

efficiency of the tool. This is in line with the theories and accounts claiming that children's poor performance on certain explicit tasks are related to processing demands and not a lack of competence (e.g., Leslie, German, & Polizzi, 2005). We also compared children's performance in the two-different experimental groups and found that children tend to correctly predict the puppet's behavior based on his false belief in the group with no reality bias (Experiment 2) more often, compared to the group of children who participated in the experiment with high processing demands (Experiment 1). Results from these experiments suggest, that around the age of 4, when children typically pass the standard false belief tasks that involves representing beliefs about the location of objects, they are also able to integrate more complex contents in their belief reasoning system. However, the current task might pose greater processing demands on children's performance. Compared to a location change scenario, where they are simply required to maintain that the agent's goal is to get her object and attribute a representation of his (false) belief about the location of the desired object, here they are required to maintain that the agent's goal is to obtain rewards with the help of a tool. They are required to integrate the agent's causal knowledge and represent his belief about which tool is efficient for getting the rewards. These inferential demands in the current task might impact children's performance in a way that it becomes even harder to inhibit the reality bias compared to a standard location change task. In the following experiment, we aimed to test older children with more mature cognitive abilities, to investigate whether later in development children would be able to overcome these reality biases even in scenarios involving higher inferential demands.
# 2.4. Experiment 3. 5-year-olds integrate efficiency information in false belief reasoning (reality bias version)

We hypothesized that if 4-year-old children have difficulty inhibiting a reality bias, 5year-old children could succeed in the task from Experiment 1 given their more mature inhibitory control abilities (e.g., Zelazo & Müller, 2002). Therefore, we conducted a third experiment with 5-year-old children.

### 2.4.1. Methods

#### **Participants**

24 children (9 female) participated in the study. Their age ranged from 5 years 1 month to 5 years 11 months (Mean age: 5 years 3 months). Additional 12 children participated but were excluded from analyses either because they failed to give a correct answer to one of the post-test questions (N=9) or failed to give a valid answer to the false belief test question (N=1) or because of parental intervention (N=2).

#### Procedure

The same procedure was repeated as in Experiment 1.

#### 2.4.2. Results

In Experiment 3, 5 children gave a valid answer immediately to the open question, and the remaining 19 children provided a valid answer to the standard question. Results indicated that 5-year-old children tended to predict correctly the puppet's behavior based on his false belief: 17 (open N=3, standard N=14) out of 24 children (70%, p=.063, binomial) indicated

that the puppet would search his stick in the box where the originally long stick was left (which in the meantime became short). Estimated Bayesian factor moderatly supported the alternative hypothesis (BF=3.794).

However, when comparing results from Experiment 1 and 3, the GLM analysis did not support a difference between the groups, in the sense that children in the older age group would be more likely to choose the box which was congruent with the puppet's false belief (, est.=.55, se=.61, p = .37).

#### 2.4.3. Discussion

Result from Experiment 3 suggest that older children tended to be successful when they were required to predict the puppet's behavior based on his false belief about the efficiency of the tool. However, when we compared the two groups, we found no significant difference between younger and older children's performance. Surprisingly, the current task seems to be not very easy even for older children who should be able to deal well with the extra demands posed by the standard tasks, as only 70 % of them show successful performance. These results further suggest that the current task might pose specific demands on children's inhibitory control abilities, and with the maturation of cognitive control they are more likely to perform efficiently.

#### **2.5. Experiment 4. 5-year-olds tracking true beliefs about tool efficiency**

To further confirm that in Experiment 3 children's performance was related to their understanding of the puppet's (false) belief, we conducted a true belief control version of the task with another group of 5-year-old children.

We hypothesized that children should be able to predict the puppet's action upon his return, congruently with his true belief about the efficiency of the tool. Therefore, we predicted, that upon the puppet's return children should take into account the puppet's true belief and indicate that he would approach the box with the new long stick, as it is more efficient compared to the short stick that was previously long.

#### 2.5.1. Methods

### **Participants**

24 children (10 females) participated in the study. Their age ranged from 5 years 1 month to 5 years 11 months (M= 5 years 4 months). Additional 10 children participated but were excluded from analyses either because they failed to give a correct answer to one of the post-test questions (N=9) or failed to give a valid answer (children's answers were coded as relevant as in the previous experiments) to the true belief test question (N=1).

#### Procedure

The same procedure was repeated in this true belief experiment as in Experiment 3, with the only difference, that the bear did not leave the scene, but he broke the stick and replaced the ending to the originally short one, and no secretive voice was used, the experimenter simply said that the bear would like to play a trick now with us. Furthermore, in order to avoid that the child may simply select the stick that was last manipulated by the puppet and the experimenter, for half of the participants, the experimenter (with the puppet in her hand) first broke the long stick, but did not put it back to its box, instead, she reattached the ending to the short stick and placed the broken stick in the box only after the new long stick was already placed.

#### 2.5.2. Results

Out of 24, 5 children gave a valid answer immediately to the open question and 19 provided a valid answer to the standard question. Children correctly predicted the puppet's behavior based on his true belief: 23 (open question N=5, standard N=18) out of 24 children, (95%, p<.0001, binomial) predicted that the puppet would search for the stick in the box where the new, efficient stick was placed. Estimated Bayes factor favored the alternative hypothesis (BF=55924)

We compared 5-year-old children's performance in the true and false belief experiments (Experiment 3 and Experiment 4), by contrasting how likely they were to choose a specific box. Fisher exact test showed that children were more likely to choose the short (broken) stick's box in the false belief condition, when they were asked to predict the character's behavior, compared to the children in the true belief condition, who were more likely to choose the new long stick's box (p<.001). Children's performance across the false and true belief trials in represented in Figure 5.

Results based on GLM analysis comparing children's answers in the true and false belief conditions suggested that children were less likely to choose the old now inefficient stick in the true belief condition compared to the false belief condition (Est=-4.02, se=1.12, p<.001).



*Figure 5*. 5-year-old children's correct answers to the false and true belief test question of Experiments 3 and 4

#### 2.5.3. Discussion

Results from Experiment 4 suggest that children can successfully take into account the puppet's true belief about the efficiency of the tools available. That is, once they are aware that the puppet knows which stick is efficient, they would predict that the puppet would approach the box with the newly fabricated efficient stick. Compared to Experiment 3 where children were more likely to choose the previously efficient stick, which after the false belief implementation became inefficient, in the true belief task most children indicated that the puppet would approach the box with the new efficient stick. When we compared children's choice regarding the sticks, we found that children are significantly more likely to choose the new efficient stick in the true belief condition compared to the children in the false belief condition. That is, children's indication of the location of the stick depended on whether the puppet had a false or true belief about the efficiency of the sticks. We also aimed to eliminate any possible bias towards the last manipulated object, as half of the sample witnessed that the puppet is putting back the old efficient stick into the box only after he placed after he placed

the new efficient stick into the box. If children in Experiment 3 would have been biased by the last manipulated object, they should have indicated the old efficient stick's box more often, but only one participant did so by answering incorrectly.

# 2.6. Experiment 5. 3-year-olds integrate efficiency information in false belief reasoning (no-reality-bias version)

To test our main research question, we further investigated whether younger children would be able to attribute information about the efficiency of tools in belief reasoning. Given that 4-year-old children successfully predicted the character's behaviour based on his false belief when the reality bias was excluded, we investigated whether younger, 3-year-old children, would be able to succeed in the same task.

Alternative accounts (e.g., Apperly and Butterfil, 2009) would suggest that the early developing system that is operating in children younger than four is inflexible and has signature limits, therefore would be unable to deal with an agent's false belief regarding contents involving efficiency. However, based on the one system accounts, here we predict, that given that 3-year-old children are able to pass an explicit false belief task with reduced processing demands, and they have a good understanding of tool efficiency, they might be able to integrate efficiency information in their belief reasoning.

2.6.1. Methods

#### **Participants**

24 children (17 females) participated in the study. Their age ranged from 3 years to 3 years 11 months (Mean age= 3 years 5 months). Additional 14 children participated but were

excluded from analyses either because they failed to give a correct answer to one of the posttest questions (N=6) or failed to give a relevant answer to the false belief test question (N=1) or because of parental intervention (N=2) /lack of cooperation (N=5).

# Procedure

The same procedure was repeated as in Experiment 2.

### 2.6.2. Results

11 children gave a valid answer to the open question and 13 children gave a valid answer to the standard question, by indicating one of the boxes. Children successfully predicted the puppet's behavior based on his false belief: 20 (open question N=9, standard N=11) out of the 24 children (83%, p=.001, binomial) predicted that the puppet would search his stick in the box where the originally long stick was left (and now became short).

These results indicate, that when reality bias is not inferring with children's ability to respond accurately, even 3-year-old children are able to integrate information on tool efficiency when reasoning about others' belief, and they are able to predict the puppet's behaviour. Estimated Bayes factor favored the alternative hypothesis (BF=126).

#### 2.6.3. Discussion

3-year-old children successfully predicted the puppet's behavior based on his false belief about the efficiency of the tool. However, one might argue that there can be an alternative explanation to this result, specifically that young children may be biased to search at the last manipulated location. In this experiment the experimenter always put first the originally short stick into one of the boxes, then in the test phase, broke the originally long stick, and placed it back to its box. To ensure that children did not indicate the box, which was last manipulated by the experimenter, we conducted a further experiment to replicate the results with the 3-yearolds while controlling for such bias.

# 2.7. Experiment 6. Replication of Experiment 5: Excluding alternative explanations

To exclude that 3-year-old children simply indicated the box which was last manipulated by the experimenter in Experiment 5, we modified the task in the following way: after the experimenter broke the long stick and placed it back to the box, pulled out the originally short one and commented: "look, now the long one became short as this one", and then placed it back to its box. Thus, now the last manipulated box was the non-belief relevant.

We predicted that if children can integrate information about efficiency with the puppet's false belief, and not simply indicate the box which was last manipulated, they should behave as in Experiment 5, indicating the box where the originally long stick was placed.

### 2.7.1. Methods

#### **Participants**

24 children (15 female) participated in the study. Their age ranged from 3 years 1 month to 3 years 11 months (mean age 3 years 5 months). Additional 12 children participated but were excluded from analyses either because they failed to give a correct answer to one of the posttest questions (N=9) or failed to give a relevant answer to the false belief test question (N=2) or because of parental intervention (N=1) and experimental error (N=1). Children were tested in the Central European University's Babylab (Cognitive Developmental Center).

# Procedure

The same procedure was repeated as in Experiment 5, with the modification discussed above.

### 2.7.2. Results

Six children provided a valid answer to the open question, whereas 18 children indicated one of the boxes after the standard question. Children successfully predicted the puppet's behavior based on his false belief: 20 (open question N=6, standard question N=14) out of 24 children, (83%, p=.001, binomial) predicted that the puppet would search for his stick in the box where the originally long stick was left (which became short in the meantime). Estimated Bayes Factor favored the alternative hypothesis (BF=126). Children's performance in Experiment 5 and 6 is shown in Figure 6.



Figure 6. 3-year-old children's correct answers to the false belief test question in experiments 5 & 6

# 2.7.3. Discussion

Experiment 6 successfully replicated the results from Experiment 5, excluding a possible alternative hypothesis. Children predicted the puppet's behavior based on his false belief, even when the last manipulated object was the originally short stick. If in Experiment 5, children simply indicated the box with the stick that was last manipulated, then in Experiment 6 children should have selected the last manipulated box more often. To the contrary, children indicated the correct box as often as in the previous experiment. These results from Experiment 5 and 6 are in line with our hypothesis that children from relatively early on, once they have a good understanding about tool efficiency, are able to predict a character's behavior who has a false belief about the efficiency of the tools. Moreover, these results are not in line with the signature limits suggested by the two-system account, based on which children would be unable to attribute beliefs to agents about *how* an object is represented.

# 2.8. General Discussion

In the present research, we investigated 3- to 5-year-old children's ability to reason about others' belief about tool efficiency. Previous experiments targeting early Theory of Mind abilities, tested children's ability to attribute false beliefs about location or identity of objects. In contrast, the current study involved a scenario when children had to integrate into their belief reasoning an agent's causal understanding about the efficiency of tools.

# **Overview of the findings**

Throughout the experiments children first observed a plush puppet whose goal was to collect balls from a long tube using a hook ended stick that varied in length. During the familiarization trials, children learned that with his short stick the puppet can reach and only two balls that were on the top of the pile, but he is able to collect the remaining ones with the help of his long stick. Therefore, the length of the stick directly determined the expected success, turning the long stick an efficient the short stick an inefficient means to achieve the goal. Before the puppet finished the collection of the balls, in the false belief condition, he left the scene and the experimenter changed the efficiency of the sticks in a way that she either broke the long stick into two and reattached its ending to the originally short stick (Experiment 1, 3) or took the ending away (Experiment 2, 5 and 6) or the puppet itself changed the ending of the sticks (Experiment 4), so it became inefficient.

In Experiment 1, 4-year-olds failed to predict the puppet's behavior based on his false belief: almost half of the children indicated that the puppet would approach the box with the new, efficient stick. As previous research on Theory of Mind development demonstrated that children around the age of 3 are more likely to infer correctly a character's false belief when they are less biased by their own knowledge (e.g. Gherear, Baimel, Haddock & Birch, 2021; Birch & Bloom, 2007; Southgate, Senju, Csibra, 2007), we decided to modify the task in order to reduce reality bias, given that the current tasks may pose difficulties for older children as well.

In Experiment 2, the only modification compared to the first experiment was that in the test phase the experimenter after breaking the long stick, did not reattach its' ending to the short stick, but simply put it away. Therefore, upon the character's return, there were two short sticks placed in the two different boxes, so that the child was not required to inhibit his own knowledge about a new, efficient stick, he simply needed to track the puppet's false belief about the originally long stick. Results from the second experiment showed that 4-year-old children successfully predicted the puppet's behavior based on his false belief; upon his return, most of the children indicated that he would approach the box where he falsely believes that the efficient stick is, even though children were aware of that the stick is already inefficient. Therefore, 4-year-old children's performance increased once the reality bias was removed.

In Experiment 3, we hypothesized, that 5-year-old children, given their possibly more mature cognitive control abilities, would succeed in the original task without removing the reality bias. Therefore, we performed the first experiment with 5-year-old children and the results showed that participants tended to predict the puppet's behavior based on his false belief, by indicating the box with the originally efficient stick that later became inefficient in the absence of the puppet. However, when the two age groups were compared, we found no significant difference in children's performance, which to our surprise may suggest that the current task involving reality bias may pose difficulties even for 5-year-old children. In Experiment 4, we tested a true belief condition with 5-year-old children, where the only modification was that the character did not leave the scene but reattached the ending of the originally efficient stick to the short, inefficient stick. In this case, most of the 5-year-olds predicted that the character would approach the box with the new efficient stick. We compared

children's performance between the two conditions and found that children's indication of the boxes (either the box containing the originally efficient, but now inefficient, or the box containing the originally inefficient but now efficient stick) depended on whether they participated in the false or true belief conditions. Children were more likely to predict that the puppet will approach the box with the inefficient stick in the false belief condition, whereas they predicted that he would approach the box with the efficient stick in the true belief condition.

Next, given that 4-year-old children were successful in the task once the reality bias were reduced (Experiment 2), in Experiment 5-6, we tested whether 3-year-old children, who have a good efficiency understanding and they also tend pass tasks with reduced processing demands, would be able to attribute false beliefs to an agent about tool efficiency. Three-year-old children were tested with the same procedure as in Experiment 2, Results showed that 3-year-old children were successful at taking into account the puppet's false beliefs about the efficiency of the tool, as most of the children correctly predicted that the puppet would approach the box where he falsely beliefs his efficient stick is placed. To exclude alternative explanation in Experiment 6 we replicated these results with a modified version.

Overall, these results suggest that already by the age of 3, children can integrate information about tool efficiency and another agent's causal knowledge in their false belief reasoning. The results are in line with the theories proposing a single Theory of Mind system that is present from early on and it is able to operate efficiently in various scenarios, similarly to the adult Theory of Mind system (Leslie, 1987; Kovács, 2016), which however may become more sophisticated due to the development of other cognitive abilities with age (Carruthers, 2013).

#### **Discussion of the findings**

Compared to previous studies testing Theory of Mind abilities in young children, the current experiments involved a more complex scenario: children were not merely required to represent an object's location from someone else's perspective, but they were required to take into account the object's function, and more importantly, the causal relation between the efficiency of the object and the outcome it can produce. As discussed in the introduction, there are set of differences between attributing beliefs about objects and their location (or their identity,) and objects and their efficiency. For instance, if children were simply required to attribute beliefs about the sticks location which was replaced in the absence of the puppet, they would have been simply required to update the spatial coordinates of the stick when the false belief has been implemented and keep a non-updated representation about the location of the stick from the puppet's perspective to predict his behavior. However, in the current experiments, children were first required to understand which tool is efficient for the puppet to obtain the desired rewards. During the training trials, the puppet did not show a preference for any of the sticks, he simply demonstrated, that the short stick allowed him to reach only the balls on the top, only the long one was efficient enough to obtain the balls from the bottom of the tube. Once he left the scene, and the false belief was implemented, children did not have to update the location of the stick, but its efficiency as it has changed (it became inefficient, and in the reality bias version, the short stick became efficient). Therefore, children were required to update the relation between the stick's efficiency and its effect and keep a non-updated representation from the puppet's perspective. Whether the puppet was aware of the change of the efficiency or not was the main factor determining their prediction about the puppet's behavior. For instance, when the puppet broke the of the long stick and reconnected the resulting part with the short one, he must have true belief about that the originally short stick becoming efficient. As expected, 5-year-olds successfully predicted that the puppet would approach the box with the new efficient stick. When the puppet didn't see this transformation, they predicted the opposite search pattern. Therefore, older children were able to take into account whether the puppet has a false or true belief about the efficiency of the sticks and predicted his behavior accordingly.

The results obtained with the 3-year-old group are not in line with the signature limits proposed by the two-system account. Based on the two-system account, children below the age of four are able to track only that an agent registers an object at a specific location, e.g., location A. If the object is moved to location B in the absence of the agent, children would expect the agent to search it in location A, as this is where he registered it. However, in the current experiment, the agent represents the object not (or not only) at a certain location, but also having a certain efficiency. In the absence of the agent, the location of the object does not change, but its efficiency does. Based on the two-system account, the early developing system that available for children before the age 4 would be unable to deal with an agent's beliefs about *how* this object is being represented. However, our results suggest that even younger children are able to predict an agent's false belief about the efficiency of the tools.

As every experimental study, this study may also have specific limitations. One could argue that children simply tracked that the agent registered one of the sticks at a certain location, and they simply expected him to approach the box that previously contained it, or they simply represented that the puppet wants to long stick, without representing his beliefs about the efficiency. There are two reasons that could possibly exclude these interpretations. First, if children tracked that the puppet registered the long stick at a certain location, they should also track his registration about the short, inefficient stick. The only reason why they expect the puppet to approach the long stick's box, is because they attributed a belief to the puppet about *how* this stick is being represented. Second, we did not train children that a puppet has a specific preference for any of the sticks. Children simply learnt that the puppet has the goal to collect

the balls, and in order to achieve his goal while few balls were only in the tube, he needs the long stick. Therefore, our familiarization ensured that children did not attribute a preference to the puppet for a specific stick, instead, they attributed an understanding about the tool's efficiency. Once the false belief has been implemented, they integrated his understanding into their false belief reasoning to correctly predict the puppet's actions.

To summarise, these findings suggest that children already by their age 3 are able to represent beliefs that go beyond beliefs about object locations. They seem to track an agent's goal and causal understanding about the efficiency of tools available and integrate this information while reasoning about his beliefs. Future research may target even younger children (e.g., 2-2.5-year-olds) who already have a good understanding about tool efficiency, and who were reported to be able to pass specific ToM tasks with reduced processing demands (e.g., Setoh, Scott & Baillargeon, 2016).

# Chapter 3. Can young children represent false beliefs about causal events?

# **3.1. Introduction**

To make sense of others' behavior, we generally need to consider what sort of information is available to them that could guide their actions. Just like our own perception usually provides adequate reasons to believe the evidence acquired in this way, detecting whether another person does or does not have perceptual access to the state of the art may have an important role in justifying the beliefs we attribute to others. However, some beliefs we form go beyond perceptual evidence. Consider the following examples:

- 1a.) The piano is in the living room
- 1.b) Peter believes that the piano is in the living room
- 2a.) I have a migraine
- 2b.) Peter believes that I have a migraine
- 3a.) The elevator is broken, therefore I must take the stairs
- 3b.) Peter believes that the elevator is broken so I will take the stairs instead

Statements like 1a.) may be the result of visual evidence collected during the exploration of the room we are standing in, statements like 2a.) about introspection resulting from an evaluation of our internal physiological states, 3a.) information accessed through inference.

Some of these, like the one in 1a) may be directly observable and attributable to others as belief contents like 1b).

Evidence about others' internal states (like 2a) can be acquired if someone shared this information, communicating it to an involved recipient, whereas 3a) is information that we obtained through inference (for instance by noticing that the elevator fails to operate despite our action on the control panel). Thus, as we could see from the above examples, belief attribution may be supported by information from various sources: it may be supported by perceptual evidence, by verbal reports or based on inferences. Neurotypical adults face no difficulty to explain or predict others' actions based on their knowledge, desires or beliefs, even in situations that require integrating information from various sources with others' mental states, allowed by their Theory of Mind abilities. A crucial question is how does this ability develop? Children from early on are often involved in social interactions where tracking the inferences what others may perform is a relevant source of information. In the current chapter, we aim to investigate how children can deal with social situations where the perceptual justification of beliefs is insufficient.

Developmental research from the past decades indicates that the ability to make sense of agent's intentional states, such as goals, preferences and desires emerges early in infancy, within the first two years of life (for a review see Baillargeon, Scott & Bian, 2016). A line of developmental research dedicated a special focus on how infants can track others' epistemic states such as ignorance and knowledge (e.g., Luo & Baillargeon, 2007; Moll, Carpenter & Tomasello, 2007), revealing that infants from early on recognize what an agent might or might not have seen, and they are able to predict or explain agents' actions based on such information. Further studies investigated these abilities beyond knowledge and ignorance attribution by testing infants' and young children's ability to represent beliefs. Some of the early research investigating children's Theory of Mind abilities with explicit tasks, showed that children were able to pass false belief test only over the age of 4 (Wellman, Cross & Watson, 2001) while implicit studies revealed success already in infants (see Onishi & Baillargeon, 2005 as the very first piece of evidence). Additionally, studies using tasks with reduced cognitive demands provided positive evidence of explicit false belief understanding even in toddlers (e.g., He, Bolz & Baillargeon, 2011; Scott, He, Baillargeon & Cummis). That is, based on these previous studies, infants and young children are able to reason about others' beliefs, beyond knowledge or ignorance, when making sense of others' actions.

However, most of these studies with infants (e.g., Onishi & Baillargeon, 2005, Southgate, Senju, & Csibra, 2007; Song, Onishi, Baillargeon, & Fisher, 2008 Buttelmann, Carpenter, & Tomasello, 2009; Kovács, Téglás & Endress, 2010) and young children (He, Bolz & Baillargeon, 2011; Scott, He, Baillargeon & Cummis, 2012; Rubio-Fernandez & Geurts, 2013, 2016, Scott & Roby, 2015) tested scenarios when an agent had a (false) belief regarding the presence of an object at a certain location, indicating that participants expect an agent to act according to their beliefs. In a typical false belief scenario if an agent was absent during the replacement of a desired object, participants expected them to search for a toy where they saw it was hidden, and not where the toy really was. In these studies, infants and young children reasoned about an agent's behavior by taking into account what kind of information was available to them based on perceptual evidence. A question emerges, whether young children, like adults, are also able to track what information an agent can arrive to through inferences.

For instance, let's consider a young child and his mother. The child goes to a nursery located on the 3rd floor of a building. He knows that every day in the afternoon his mom waits for him right by the elevator on the ground floor where the teacher accompanies them. One day he learns that the elevator is broken, and he needs to take the stairs. Will he become upset upon his arrival to the ground floor by the staircase when seeing that his mom is not there? Or would he think that his mom is likely waiting for him by the elevator, around the corner because she doesn't know that it is broken, and she believes he will arrive with the elevator?

While earlier research demonstrated that children are relatively good at making sense of others' actions when they need to consider information that is available to them through perceptual evidence (e.g., what they can see and know, e.g., Pratt & Bryant, 1990), and they can attribute false beliefs in "location change" situations, it is still unclear whether they can go beyond these simple scenarios. Perceptual justification is assumed to be a prerequisite for this ability, because oftentimes it plays a crucial role in initializing belief attribution. From this perspective, someone seeing something at a certain location leads to having a belief about it. Is the early developing belief attribution system restricted to perceptual evidence? Can infants and young children's track the inferences that others may perform during social interactions? The interactions that children encounter often unfold in a rich web of causal relations, but it is not yet known whether they can integrate their knowledge about various causal events in belief reasoning.

Our previous example with the elevator meant to capture such a scenario. If the child understands that the elevator doesn't work, and that the broken elevator restricts possibilities, forcing someone to take an alternative route, will he attribute this inference to his mother when reasoning about her belief? While in some cases, the available perceptual information is sufficient to inform our causal inferences (for example, if we put an object into a rolling ball's way, we can infer that it will block the ball's path), this is not always the case. The elevator's malfunction may not be detected before an intervention over the control panel is implemented (for instance, the elevator fails to react, despite the operation of the appropriate buttons). Belief tracking requires the integration of one's knowledge about the causal structure (if there is an obstacle, the ball won't move) with the representation of someone's belief (if Peter sees the obstacle, then he will think that the ball won't move) or even in a false belief or ignorance scenario (if Peter did not see that there is an obstacle, then he believes that the ball will move).

Such situations, in which available perceptual information is insufficient to predict an agent's actions, are frequent in everyday life. The agent might not see something, objects might hide other objects, or some relevant details might be missed by the agent. We know that infants are able to form hypotheses about the beliefs of others based on the observable. Can they also do it from the inferable?

There are no previous studies about how infants and children attribute false belief to agents based on inferences. Two findings, however, indirectly point to this possibility. First, there is evidence that infants can infer the goal of an agent even if the goal object is not visible at the moment of the agent's actions (Cesana-Arlotti, Kovács & Téglás, 2020). Second, they can attribute inferences to others based on their general physical knowledge about objects (Ting, He and Baillargeon, 2021). In the following section, we will unpack these findings, in order to suggest that children's ability to attribute beliefs about inferences to others might be present early in development.

A few studies investigated infants' logical abilities before their acquisition of logical vocabulary, focusing on disjunctive reasoning. In a disjunctive inference, two (or more) alternative hypothesis are framed in a logical relation that at least one of them is true. If one of the hypotheses is eliminated than one may infer that the other hypothesis is true, based on the elimination of the different alternative(s). Whereas earlier studies showed that children under the age of 3 fail to retrieve a goal object based on the elimination of the other alternatives (Mody & Carey, 2016), a recent study (Cesana-Arlotti, Martín, Téglás, Vorobyova, Cetnarski & Bonatti, 2018) indicated that even younger infants may rely on early logical capacities to solve tasks that involve disjunctive inference by using implicit measurements such as infants' looking times and pupillometry. In another set of studies Cesana-Arlotti, Kovács and Téglás (2020)

investigated whether 14-month-old infants are able to apply this type of inference to process others' actions, and whether this can aid social learning when direct perceptual information is not available. Specifically, they have asked whether infants can identify the goals of others solely based on disjunctive inference. Crucially, infants were able to identify the preference of an agent even if the goal-object was not visible, but its identity could be deductively inferred. This result shows that the outcome of a deductive inference can serve as input for goal attribution, guiding this way infants' learning about the social world.

Most importantly for our current questions, a study by Ting, He and Baillargeon (2021) provided evidence about infants' abilities to attribute inferences to others while reasoning about their goals. The authors tested 5-month-old infants' ability to track what information an agent garners through inference, with the use of their general physical knowledge about objects' size. While a small object can be completely hidden both in a small and large box, a large object will fit only in the large box. In this experimental situation, however, the size of the objects and the size of the boxes allow only one spatial arrangement where all objects can be completely hidden at the same time. The experiments were built on a Violation of Expectation paradigm. First, infants were familiarized with an agent expressing a preference for a wide over a narrow toy, then, in the agent's absence, the wide toy was hidden in a wide, and the narrow one in a narrow container. The available physical information was sufficient for infants to infer the location of the objects and based on this they should have assumed that the agent will identify location of the preferred object by working through the steps of the same inferences. Results were in line with this assumption: infants looked longer (expressed surprise) at scenes where the agent returned and reached for the narrow box. This revealed that infants' expectations were congruent with the assumption that infants attributed inference to the agent. This study suggests that young infants spontaneously track the inferences others may perform and integrate the outcomes of these operations with the goals and preferences that guide their actions.

It can be assumed, that if preverbal infants can attribute inferences to others, young children would face no difficulty in such situations. However, it is still an open question whether children can apply such inferences when reasoning about others' beliefs, and whether they are able to represent someone's false belief when the information about the belief content is only available through inferences? Children comprehend causal relations from early on (for a review see Muentener & Bonawitz, 2018). However, it is not yet known whether the inferential apparatus used in first person with a great ease can help them to reason about others' beliefs. As so far research on Theory of Mind development mainly tested situations where an agent has a false or true belief about the location, or the identity of an object, further research is needed in order to investigate whether the early developing belief reasoning abilities are flexible enough to operate in different scenarios, for instance when the information about someone's (false) belief requires additionally the integration of a causal inference.

### **3.2.** The aim of the current study

The present study aims to explore whether children are able to take into account an agent's belief and the causal inferences she may have performed in order to arrive to these beliefs. We developed a new task that required children to predict an agent's action that has a false or true belief regarding the causal structure of a ball dispenser machine. The device that stays in the centre of the events has a central container filled with balls and two lateral arms, operational through the activation of a simple release mechanism: by pushing the button on the top of the central containers, the balls will roll out to each side through the arm. However, if an obstacle is placed on one side of the machine, the balls will only roll out to the other side. After children are familiarized with an agent who operates this machine, a belief manipulation will be implemented in the test phase in a way that the agent will either have a false or true belief

about the location of the obstacle. These scenarios are generated in a way that the location of the balls can be inferred from the position of the obstacle.

We developed a task which requires that children understand both the consequences of interventions and the role of obstacles for changes in motion trajectories. In the following section, we describe some evidence justifying infants' and young children's understanding about causal relations.

# Infants and young children's understanding of causal relations

Infants are sensitive to causal events and their violations (Kotovsky & Baillargeon, 1994; Leslie, 1984; Leslie & Keeble, 1987; and see Scholl & Tremoulet, 2000 for a review), possibly guided by a set of causal primitives available very early on, e.g., about contact mechanics. To achieve this, infants exploit the same spatial and temporal features that influence adult causal perception (e.g., Mascalzoni Regolin, Vallortigara & Simion, 2013). However, detection of causal relations is not limited to contact mechanics characterizing the collision of moving objects: 4-year-olds can infer more complex causal structure even when there is no physical contact between the cause and the effect (Kushnir & Gopnik, 2007) simply by exploiting conditional interventions across different domains (Schulz & Gopnik, 2004; Schulz, Gopnik & Glymour, 2007, Schulz, Bonawitz, Griffiths, 2007) suggesting a domain general causal cognition throughout the development. Learning about causal structures is not restricted to first person experience. Children can efficiently extract causal structures by observing others' goal directed interventions already at 24-month (Meltzoff, Waismeyer & Gopnik, 2012, see Luchkina, Sommerville and Sobel, 2018 about accidental interventions). Uncertain causal structures drive exploration: children are more likely to explore a function object if they can generate disambiguating evidence (Schulz and Bonawitz, 2007). Toddlers are found to be sensitive to the presence of an agent behind causal events; Bonawitz and colleagues (2010) found that 2-year-olds represent predictive events of a causal chain but only in the case when the events were initiated by an agent.

These studies suggest that young children have a rich understanding regarding causal properties, they understand how interventions reveal a causal structure, and use this knowledge to predict the outcome of interventions. Here we aim to test whether they are able to integrate this understanding with their mindreading abilities. If they understand the relation between cause and effect – e.g., an obstacle will prevent an agent from obtaining a reward -, would they be able to attribute beliefs about similar inferences to others?

As mentioned previously, no research so far has focused on whether infants and young children are able to represent beliefs about various contents, for example, causal relations. Representing causal inferences in belief reasoning may be fundamentally different from representing objects' locations or identity. One may think that representing causal events are simply more complex, as they involve the representation of various elements, and the representations of the relation between these elements, such that they can be mapped to a cause/effect description. Cases involving more than one possible effect, require holding multiple representations grasping the future possible outcomes that can be caused by the interventions.

# Beliefs based on causal inferences

Belief attributions involving objects and their locations require that one needs to uphold a representation of the object itself and its position in space from the perspective of an agent (see example 1.b.: "Peter believes that the piano's spatial location is in the living room"). When the location changes and the belief holder will have a false belief about the current location, the person attributing the beliefs simply needs to update and encode a new location for this object from a first-person perspective and keep a representation about the old location from the other person's perspective. However, belief updating may be different when attributing beliefs about causal events (see example 3b. presented a case where beliefs about the correct or incorrect functioning of an elevator change the route one takes). In this case, new evidence has different consequences: the person holding the attributed belief needs to update the relation between two elements (e.g., the elevator and the destination in example 3b). While grasping causal relations ("A causes B") may take form of causal conditionals (like "if Cause A then effect B") allowing in principle for a set of deductively valid inferences given the available information about the cause or the effect (Goldvarg & Johnson-Laird, 2001), there is ample evidence from adults that in many real-world scenarios the richness of causally relevant information impacts reasoning (e.g., alternative causes, enablers, preventers; see Cummins et al 1991, Cummins 1995; Markovits & Potvin, 2001 for a set of early findings). Thus, instead of causal conditionals, adults may rely more faithfully on the mechanisms that they identified as responsible for connecting causes to effects in their causal reasoning. For their causal inferences, adults exploit the interventions and their consequences that unfold in time, the conditional dependencies existent in a causal chain (see Sloman & Lagnado, 2015; Waldmann & Hagmayer, 2013 for reviews). From a logical point, changes in the causal structure may involve changes in the truthvalue of the conditional that was initially established as a description of these relations. Instead, changes in the causal structure may force updating the causal chain itself, where the causal expectations will be guided by the inferential power of interventions.

In our experiments, we plan to use a device and manipulate an actor's knowledge about the mechanisms defining the functioning of this device. Although people may have very little knowledge about the mechanisms that make our modern-day tools possible, causal concepts are necessary to understand how to manipulate them in order to elicit the desired effect. In our experiment, however, the mechanism that makes the device work is fully "transparent", from the children's perspective. The well-coordinated causal chains evoked by intentional actions in order to achieve the desired effect presupposes a rich understanding of causal relations involving conditional interventions. But how efficient are children in integrating this sort of understanding of causal relations in their reasoning about other's beliefs? This is what we aim to examine in the present study. We aimed to explore whether young children's belief tracking abilities can be generalized to different scenarios, specifically, when the belief content to be attributed is available only through inferences, and not based on perceptual information.

# 3.3. Experiment 1. Can children attribute beliefs based on causal inferences?

The aim of the first study was to test 3-4 and 5-year-old children's ability to attribute false beliefs regarding casual events. In order to test this, we designed an experiment involving a touch screen task presenting video animations. Crucially, the videos presented a device and a character's interventions over this device. The device is a kind of ball dispenser: it has a symmetrical structure defined by a central container storing a set of balls and two arms where the balls could roll out (Figure 1, panel a). These arms were elongated horizontal structures with their upper side open. Depending on the height of its front wall the balls moving in the arms were visible or hidden from the observer. Whenever the person in the videos performed an intervention on the device (e.g., every time the agent pushed a button on the top of the machine) the balls rolled out to the left or right side allowing for the agent to collect them.

However, on some trials an obstacle inserted to either the right or the left side of the device could change the motion trajectories of the exiting balls (Figure 1, panel b). This obstacle reversed the motion paths: after a ball bounced into the obstacle it will move towards the opposite side (Figure 1, panel c, d). Therefore, the presence (or absence) of the obstacle changed

the causal structure of the device. For instance, *if* the obstacle was inserted on the right arm, *then* the balls would roll out on the left side. Crucially, in a certain number of training trials, and more importantly, in the test trials, the side of the horizontal arms of the device was high enough to block visual access to the moving balls, but the position of the obstacle remained visible. In the absence of visual access to the motion trajectories, the participants could predict the final location of the balls only as a result of their causal reasoning ability: they needed to infer where the balls will arrive based on the location of the obstacle. Children were asked to guess where the observed character will search for the balls.



*Figure 1*: The ball dispenser and the trajectory of the balls after the introduction of the obstacle. The balls placed in the central container (a) move toward the lateral arms if the button is pressed (b). The possible trajectories of the balls are constrained by an obstacle placed in one of the arms (c) determining the final location of the balls (d). The arrows mark the expected motion trajectories, given the intervention over the central container and taking into account the location of the obstacle.

During test trials, a second character arrived in the scene and changed the location of the obstacle. The within-participant conditions determined whether the agent witnessed this change. In order to correctly predict where the agent will search for the balls in the test trials, children needed to take into account the agent's true or false beliefs regarding the causal structure of the device. In the *True Belief condition*, the agent witnessed the repositioning of the obstacle. If children can attribute false beliefs about causal events to others, then they will expect the agent searching for the balls in the arm opposite to the obstacle. In the *False Belief condition* the visual access of the agent was blocked by an occluder. Lacking information about the change of the obstacle, children should expect the agent to search on the side of the tube where the obstacle is placed.

#### 3.3.1. Methods

#### **Participants**

Thirty-two children participated in the study. Children in the younger age group ranged from 35 months 15 days till 46 months 12 days (N=18, Mean: 40 months 1 day). Participants in the older age group were between 48 months 1 day and 72 months 1 day (N=14, Mean 62 months 23 days). 6 additional children participated but were not included in the final analyses because they either failed to complete the test trials (4) or were not willing to cooperate (2).

Participants were recruited from the larger Budapest area through Central European University's database. Their parents gave informed consent for participating in the study. Participants received a small toy as a gift. The study was approved by the United Ethical Review Committee for Research in Psychology (EPKEB) in Hungary.

# **Stimuli and Apparatus**

The movies were designed as 3D animations in Autodesk Maya 2016. They were then exported as movies at 25 fps and further edited together with real-life footage using QuickTime Player 7 software and Final Cut Express. The stimuli were presented on an ELO touchscreen, placed in front of the child.

#### Procedure

The experiment was conducted by a female experimenter in the Central European University's Babylab, in a quiet testing room. Children were seated in their parent's lap in front of a touch screen. Stimuli were presented on the touch screen. The experimenter sat on the other side of the table with a computer monitoring the experiment, in a way that the participant didn't see the experimenter's screen. In some cases, a warm-up play session preceded the experiment.

The experiment consisted of a familiarization, a training, and a test phase. The training phase had 10 trials. Test had 2 false belief and 2 true belief trials (ABBA order) counterbalanced across participants.

# **Familiarization Phase**

The aim of the familiarization movies was to introduce the structure of the device to the child, who was not required to give any answer at this point.

First (0-4000 ms) the agent addressed the child saying "Szia baba! Nézd csak!" (Hello baby, look!) and then pressed the button on the top of the device (4000-8100 ms). After the agent pushed the button, the balls rolled out to each side of the device (8000-12000 ms), as there was no obstacle present (see Figure 2). During this introductory trial, the experimenter explained the device to the child and said: "Look, the girl behind the device is trying to collect her favourite balls. She presses the button on the top in order to release the balls, so that they can roll out on each side, and then she can collect them."



*Figure 2*. Trial structure in the Familiarization and Training Phase. Trials in the Familiarisation phase start with the agent greeting the child and introducing how the device works. The participants can see that in absence of an obstacle, by pressing down the arm on the top of the central container the balls roll out to both arms of the device. Trials in the Training Phase follow the same structure, but now an obstacle is present. The child is required to predict the agent's action by tapping the relevant side on the screen signalled by the question marks. At the end of each Training trial participants received feedback: the agent reaches for the balls in the correct location. The Objects Tracking Trials allow visual access to balls; thus, participants can track their motion trajectories and for establishing their final location a causal inference was not necessary. In the Inference Trials the final location of the balls can only be established through causal inference.

#### **Training Phase**

Children received two sets of training trials, 6 trials where the route of the balls was visible (object tracking trials) and 4 trials where the route of the balls was not visible (inference trials) resulting in 10 training trials in total.

# **Object tracking trials**

In the object tracking training trials, just like in the familiarization trials, both the agent and the participant could see the balls' trajectory. In the object tracking trials, however, an obstacle was placed to either the left-hand or the right-hand side of the device's "arm". This blocked the balls from exiting on that side so that they exited on the other side (see Figure 2 panel on top). The participant had to press the side of the screen where they expected the balls to turn up.

Similarly to the familiarization trial, the agent first waved at the child (3000 ms), pushed the button (3000-9000 ms), which caused the balls to roll out on one side of the device, opposite the obstacle. This was followed by a pause, when two question marks appeared on each side of the screen. At this point, the child was asked to touch the side where they thought the agent would search for the balls. The child's response was followed by visual feedback. The feedback sequence started with the agent's ostensive communication ("Look!") and was followed by her reaching towards the correct side (16000-18000 ms). Finally, the experimenter explained to the child whether they acted correctly or not, reiterating the causal rule: "yes, correct, you see, the obstacle is on the right side and therefore the balls rolled to the left side, therefore, the agent will search there." If the answer was incorrect, or the child failed to respond, the decision was corrected and accompanied by the explanation of the appropriate behavior (e.g., that the agent will search on the other side contralateral with the obstacle).

#### **Inference trials**

The event sequences and their timing remained the same as in the object tracking trials, with the modification that the agent and the participant could see the upper part of the obstacle but not the trajectory of the balls (see Figure 2 panel on the bottom). Thus, their final location could only be inferred.

The instructions were the same as in the previous training trials, however participants received no verbal feedback (i.e., the experimenter did not correct the child's behaviour), however the visual feedback remained. At this point, the child was simply allowed to continuously play the game.

# **Test Phase**

The test phase presented two types of animations, corresponding to a false belief or a true belief scenario, respectively. Each test trial started with the agent waving at the child (3000 ms). In the **false belief test trial**, after the agent greeted the child, three occluders lowered from the top of the screen (5000-6000 ms). In this way, two occluders on each side of the device covered the obstacle's location, whereas the middle occluder covered the agent's site. The belief induction phase started when a second character, a frog (8000 ms) came in the scene, from the side, waved at the child, grabbed the obstacle (15000 ms) and brought it and placed the obstacle to the opposite side of the device (21000 ms) and left the scene (24000 ms). This was followed by the middle occluder raising (26000 ms) so that the agent was again visible. As the agent did not witness replacement of the obstacle and thus had a false belief regarding where the obstacle was. Therefore, this led her to have a false belief about the causal structure of the device. The agent addressed the child, saying "Look!" (30000-36000 ms) and pressed down the button on the top of the device. Similarly to the training trials, the child was required to touch where he thinks the agent will search for her balls, but no visual and verbal feedback was given, the following test trial immediately started after the child's response.

The **true belief trial** started the same way as the false belief trial, with the difference that the second character arrived (at 4000 ms), and the occluders only lowered down after the character has implemented the change (20000-27000 ms). Therefore, the agent had direct visual access and witnessed the change of the obstacle therefore had a true belief about the causal structure of the device.

The test phase consisted of 4 trials, 2 false belief trials and 2 true belief trials (counterbalanced in ABBA order). Giving a correct answer in a test trial required that the participant take the agent's (true or false) belief into account, regarding the location of the balls,

and, indirectly, the causal structure of the machine. A schematic representation of the test trials is presented in Figure 3.



*Figure 3*. Test trials: True belief manipulation (on the top line) Before the agent pushes the arm on top, a frog arrives (1) replaces the obstacle to the opposite side and the agent observes the change (2) frog leaves. False belief manipulation (on the bottom line): Before the agent pushes the arm on top, 1 occlude rolls down covering the agent's site, then a frog arrives and replaces the obstacle to the opposite side (2) frog leaves (3). Test phase (right side): middle occlude rolls up: child is required to predict where the agent will reach based on her false or true belief

# 3.3.2. Results

Children's responses were coded as correct if they touched the correct side of the screen upon predicting the agent's action. Each child responded to 4 questions in the training and 2 questions per condition in the test phase. We calculated an average for each child in each test condition.

#### 3.5-year-old children's performance

Children's average accuracy was 69% on the transparent training trials and 70% on the opaque training trials. Their accuracy on the true belief test trials was 75% (see Figure 4). We tested whether the mean of children's averages per test condition were different from chance using a two tailed t-test with a chance level .5. Children were significantly better than chance in the true belief condition (t=2.47, df=17, p=.02, 95% CI [.54,.96]).

However, only 3 children answered correctly to both test trials and 2 to one test trial in the false belief trials translating to a 22% accuracy. This was significantly worse than chance in the false belief condition (t=-3, df=17, p=.007, 95% CI [.02,.4]).

This result suggests that children understood the causal structure of the device but did not take into account the agent's false belief about this structure.

# 5-year-old children's performance

Children's average accuracy was 97% on the transparent training trials was and 94 % on the opaque training trials. They answered 85% of the true belief test trials accurately (see Figure 4). This was significantly better than chance in the true belief condition (t=4.37, df=13, p<.001, 95% CI [.68,1.03]). Only 2 children answered correctly to one of the questions in the false belief trials, resulting in a 7% overall accuracy. This was significantly worse than chance in the false belief condition (t=-8, df=13, p<.001, 95% CI [.033,.176]). Children's performance is shown in Figure 4.

This result suggests that much like younger children, older children understood the causal structure of the device, but did not take into account the agent's false belief about this structure.



*Figure* **4**. Results of Experiment 1 Children's mean accuracy (%) on transparent and opaque training trials, true and false belief test trials across the two age groups.

#### 3.3.3. Discussion

Based on their performance produced during the training and true belief trials, children in both age groups were able to understand the causal structure of the device correctly predicting where the agent would search for the balls where this was consistent with the agent's true belief. However, most of the children failed to predict where the agent would search when she had a false belief regarding the causal structure of the device. One possible reason for their failure is that, despite the instructions, children simply learned that the rules of the game was to point where the balls were and not where the character will search. Thus, they might have ignored or were not motivated to track the agent's perspective in the false belief trials. As they received 10 training trials, they were encouraged to detect the ball, but they were not directly instructed to pay attention to the perspective of the agent. It could be that they were simply habituated to detect the balls' location and they were continuing with this strategy in test trials as well. In
addition, their failure might have been caused by task design: in the test trials the use of the occluders might have confused them instead of highlighting the agent's perspective.

In order to motivate children to track the agent's perspective during the game, we modified our task into a go no-go task, where children were required to answer only when the agent could see the position of the obstacle in the beginning of the training trials.

# 3.4. Experiment 2. Go-No-Go task

Children's poor performance in Experiment 1 might be explained by the fact that they were simply trained to track the location of the balls, but not the agent's perspective. In order to motivate children to track the perspective of the agent, we modified our task using a go no-go paradigm. This involved the following changes: the beginning of the video, before the agent started to act on the device, the obstacle lit up. However, in some trials (go trials) the obstacle was visible from the perspective of the agent, in other trials (no-go trials) the obstacle was covered with two side occluders (same as the one used in the test phase in Experiment 1) and therefore was not visible to the agent. During the training trials, children were required to only press the correct side of the screen, where the agent saw where the obstacle was from the beginning of the trials (as there were no occluders blocking her site).

#### 3.4.1. Methods

#### Stimuli

The same stimuli were used in Experiment 2 as in Experiment 1 with the following exceptions: In the beginning of each trial the obstacle lit up (1000 ms) before the agent started acting. This was either visible from the perspective of the agent or covered by two side

occluders (Figure 5). This way we aimed to motivate children to track the agent's perspective from the beginning, and we hypothesised that if the occluders are introduced in the beginning, their presence will not confuse children during the test trials.



*Figure 5:* Example of the start of a No-Go trial – the obstacle lit up before the agent pushed the button, but this was not visible from her perspective

#### **Participants**

16 children participated in the study. Their age ranged from 44 months 1 day to 73 months 3 days (M= 60 months 1 day). Additional 8 children participated but were not included in the analyses either because they failed to cooperate (N= 6) or did not complete any of the test trials (N=2). Children were recruited from the Budapest Zoo visitors' center and were accompanied by their parents who gave their informed consent for participating in the study. Participants received a small toy as a gift.

#### Procedure

Children were tested in the Central European University's child lab at the Budapest Zoo, in a quiet testing room. Procedure was similar to Study 1, with the following exceptions regarding the instructions: within the training trials, children received Go and No-Go trials. The experimenter was saying the following in the beginning of the training trials: "Now your task is to guess where the girl will search for her balls and press that side of the screen. But listen, you only need to answer, when the girl can see where the obstacle is. If the girl cannot see the obstacle, you don't need to do anything.

### **Training trials**

Children received 16 training trials in total, 4 object tracking training trials where the agent witnessed the obstacle (Go trials) 4 object tracking training trials where the place of the obstacle was occluded from the agent's perspective (No-Go trials) and 4 Go and 4 No-Go inference trials, where the trajectory of the balls was not visible anymore. One of the main differences from Experiment 1 was that children were only required to provide an answer after the Go trials, therefore, visual feedback (the second part of the video where the agent reaches towards the balls) was only presented after Go trials. Similarly to Experiment 1, verbal feedbacks were provided till the second opaque trial.

#### **Test phase**

The test phase was the same as in Experiment 1, with the only difference that the obstacle lit up in the beginning (1000 ms), which the agent always observed.

# 3.4.2. Results

We analysed children's performance on the Go/No-Go trials separately to ensure that they understood the paradigm. Children were required to provide an answer by pressing either side of the screen to the Go trials and inhibit their answers by not pressing anything during the No-Go trials. Regarding the Go trials, children were 79% during the object tracking trials, whereas 86% during the inference trials. During the No-Go trials their accuracy was accuracy 73 % in the object tracking trials, and 78% in the inference trials. We further analysed whether children pressed the correct side of the screen during the Go trials to ensure that they managed to predict the agent's behavior. On average, children's accuracy was 76% on the object tracking trials and 75% on the inference trials. Children's success was 87% on the true belief test trials which was significantly better than chance (t=6.70, df=15, p<.001, 95% CI [.7,.9]).

However, only 3 children answered correctly to one false belief question resulting in a 9% accuracy on the false belief test trials, which significantly worse than chance (t=-8.06, df=16, p=.000, 95% CI [.01,.2]). Figure 6 represents children's performance in Experiment 2.



*Figure 6.* Results of Experiment 2. Children's mean accuracy (%) on Go and No-go training trials, and on the True and False belief test trials.

#### 3.4.3. Discussion

Despite the modification of the paradigm, children failed to track the agent's false belief, even in the older (4-5-year-old) age range. In this study children were encouraged to track the agent's perspective, yet they failed to take into account her false belief during the test trials. It could be that the modifications of the go no-go paradigm was simply not enough to encourage children to track the agent's perspective and her beliefs about the balls. Furthermore, children received 8 training trials out of 16 when they were encouraged to detect the location of the balls. It is possible that children were simply habituated to detect the location of the balls in both studies, despite the changes in study 2 to motivate them to track the agent's perspective. Alternatively, the second study involved even more training trials than the first one, and with the manipulation of the Go No-Go task might have simply resulted in a task that was difficult for children to pass.

Furthermore, it is also possible that children's false belief understanding is limited to simple events like a location change of an object. To test this, we redesigned our previous experiment as a location change task. As reported in the introduction, children at this age should be able to pass the false belief test involving a simple location change event. If children are able to pass our experiment when they are required to attribute beliefs about the location change of a ball, it could mean that their representational abilities are not flexible enough to attribute various content when reasoning about other's beliefs at this age, and this would explain their failure in Experiment 1 and 2. However, if they fail to attribute false beliefs about the location of a ball in our task, this would mean that it is the design of our task that causes their failure in the current experiments.

#### **3.5. Experiment 3. Location change task**

Our third experiment aimed to test whether children's poor performance on the first two experiments was because they were unable to take into account others' beliefs involving causal inferences, or other task-related issues. This experiment differed from the first two in such a way that no causal events were involved; only one ball rolled to either side of the device, which was replaced by the second character during the test phase.

# 3.5.1. Methods

# Stimuli

The same stimuli were used as in Experiment 1, with the only difference, that instead of an obstacle, a tunnel was placed on either side of the device. Furthermore, instead of several balls, only one ball was present. After the agent pushed the button on the top, the ball rolled out to the left or right side of the device, independently from where the tunnel was placed (order counterbalanced). The use of the tunnel served only for creating a stimulus which is visually similar to the ones used in the first two experiments.

Example of the stimuli used in Experiment 3 can be seen in Figure 7.



*Figure7.* Example of the training stimuli used in Experiment 3. The agent pushed the button that caused the ball to roll out to either on the side of the tunnel or the opposite side, independently of the position of the tunnel.

# **Participants**

22 children participated in the study. Their age ranged from 48 months 2 days to 71 months 5 days (M=59 months 1 day). Additional 2 children were tested but not included in the analyses because they were not willing to cooperate.

#### Procedure

### **Training trials**

The procedure was similar to the one used in Experiment 1, however only one set of training trials were used, where the route of the ball was always visible from the perspective of the agent. This is because no inferences were required here. Children received 8 training trials in total, counterbalancing the order of which side the ball rolled out, independently from the tunnel (either the same side or the opposite side of the tunnel).

#### **Test trials**

Test trials were the same as in Experiment 1, with the only difference that after the agent intervened on the device, the second character changed the location of the ball from one side to the other. This change again was either visible from the agent's perspective, or her sight was covered by the occluders used through the previous experiments. Therefore, to answer correctly in the test trials, there was no need for the children to take into account the agent's true or false belief regarding the causal structure of the device, but simply her false belief about the location of the ball. Children received the same instructions as in the previous experiments (Guess where the girl will search!)

#### **3.5.2. Results**

Children's mean accuracy was 86% on training trials. This suggested they understood the task. Most of the children successfully predicted the agent's action in the true belief test trials, resulting in a 90% accuracy (only 1 child failed, and 2 children provided answers by touching the relevant side of the screen only for one true belief trial). This was significantly better than chance (t = 7.65, df = 21, p<.001, 95% CI [.7,1.02]). However, children again failed to answer correctly in the false belief test trials. Only 2 of them answered correctly to one false belief test trial each, resulting in a 4% mean accuracy. This was significantly worse than chance t = -14.491, df = 21, p<.001, 95% CI [.01,.1]). Results are shown in Figure 8.



*Figure 8.* Results from Experiment 3: Children's Mean accuracy (%) in the Training, True belief and False belief trials

#### 3.5.3. Discussion

Despite children's successful performance on the training and true belief test trials, their performance on the false belief trials remained poor. As several studies report that children at this age are able to pass even the standard false belief tasks involving simple location change

scenarios, these results suggest, that children's failure in Experiment 1 and 2 may not be due to their inability to represent false beliefs about causal relations, but task related issues.

### 3.6. General Discussion

The three experiments described above were aiming to test young children's false belief understanding when these beliefs are based on causal inferences. In Experiment 1 we tested children in 2 age groups (3-4 and 5-year-olds), asking them to predict an agent's behaviour after she had a false belief about the causal structure of a ball dispenser device. Our results indicated that although children performed well on the training trials, and they successfully predicted the agent's actions on the true belief trials, they failed to do so on the false belief trials. One of the main concerns about their poor performance was the possibility that during the training trials, they were habituated to detect the location of the balls and they were not motivated to track the agent's perspective on the false belief test trials. In order to motivate children to track the agent's perspective, we transformed Experiment 2 into a Go No-Go task. In this experiment, during the training trials, children were required to predict the agent's action only if she saw the obstacle lit up in the beginning of each trial (Go trials) whereas they were required not to act when the obstacle was covered from the agent's perspective (No-Go trials). Again, children's accuracy on the training trials and on the true belief test trials was relatively high, yet their performance remained poor in the false belief trials. It is possible, that this kind of Go No-Go task did not motivate them to track the agent's perspective as expected, and they only learnt the rule that they shouldn't act whenever they saw an occluder in the beginning of the trials, and they kept ignoring the agent's perspective.

To test whether their poor performance can be explained by the limitations of the belief tracking system of the 5-year-olds to represent beliefs based on causal inferences, in

116

Experiment 3 we modified our experiment into a location change task. As Children after the age of 4 are able to pass explicit false belief tasks involving simple, location change scenarios (Wellman et al., 2001), we hypothesised, that if they are able pass our experiment involving a location change of the ball, their poor performance might be explained by their limitations to track beliefs about causal inferences. To our surprise, while our participants' performance was high in the training and true belief trials, yet, their performance was below chance in the false belief trials, where they acted as the character had a true belief. Again, probably they simply tracked the location of the ball, and predicted goal directed actions while ignoring perspective.

These results suggest that it is very likely that the failure of the experiments and the poor performance of the participants originates in certain properties of our design, rendering these experiments inefficient test to study our initial questions. Children' failure in the false belief trials across the three experiments has various explanations.

One possibility is that the training sessions were two long (10 training trials in Experiment 1 and 3, 16 training trials in Experiment 2). Given the length of the training children may have learned to indicate the location of the ball, a response that they could not supress in the false belief trials. It is clear, that they performed the appropriate causal inference routinely, based on the location of the obstacle they could predict the motion trajectory of the objects with a great easy. Despite their competence with the causal task, they seldom integrated the result of these inferences with the agent's perspective when it was required in the false belief trials.

Another possibility is that pragmatic factors might have played a role in their failure. Children were instructed to "guess where the agent will search for the ball". "Guess" is a mental verb; children are able to comprehend it in certain contexts but have an adult-like understanding only around the age 7 (Miscione et al. 1978). Thus, we cannot exclude, that our participants had difficulties in understanding instructions with 'guessing" in relation to this task. Instructions that are not clear enough, might hinder the expected achievement, preventing them to take into account the agent's true or false beliefs during test trials.

A third possibility is that the participants simply did not pay attention at the end of the task (e.g., due to fatigue). and they were simply tried to correctly detect the ball's location.

Moreover, it is also likely that certain visual features of our study caused children's poor performance. For instance, it could be that the occluders were not successful in blocking the agent's line of sight. If the participants didn't have the impression that the agent did not see the change in the scene, may have turned our experimental manipulation inefficient. The use of the occluders during the training trials in Experiment 2 did not improve children's performance either. Overall, it is possible, that the visual scene with the moving occluders was simply too complex and confusing for our participants.

Considering the reasons mentioned above, the material and the method of our experiment can be improved. For instance, fewer training trials may prevent habitual responses. Introduction of more implicit measures would allow us to test whether children would take into account the agent's perspective in the false belief trials even without a lengthy familiarisation. Another possibility is to modify certain visual features of the task. For instance, instead of the use of the occluders, the agent can simply turn away from the scene while the change takes place by the second character. Such design needs to ensure that the disappearing of the obstacle won't pose great demands on children's memory: for instance, if the obstacle disappears from both the participant and the agent's perspective, children are required to remember its' previous location and predict the location of the ball by keeping this in mind.

Finding a suitable modification of the experiments described above would help us gain a better knowledge about young children's belief reasoning abilities, and whether they can attribute beliefs about inferences similarly as they attribute beliefs about locations. Ting and her colleagues (2021, reviewed at page 6) tested infants' ability to attribute knowledge and

118

ignorance about a location of a toy based on an agent's general physical knowledge about objects. If infants and young children can attribute beliefs about inferences, they should look longer in a scenario where the agent expects the location of the preferred toy (in this case for example, the narrow toy, due to physical constrains) in the narrow box, however, in their absence the narrow toy would be replaced into the wide box. Another possibility is an experiment where participants would witness a big and a small ball rolling down in an inverted Y pipe system, in a way that the small ball can roll through the narrow, but the big ball can only roll through the wide pipe. In the beginning of the test phase, the agent would witness the big ball being inserted in the pipe system and would leave the scene. Later, the big ball would be placed into a box under the narrow pipe, and therefore causing the agent to have a false beliefs about inferences, they should expect the agent to reach for the box under the wide pipe and would be surprised if the agent reaches for the box under the narrow pipe. Evidence from these suggested experiments would further explore the strengths and limitations of the mindreading system at different stages of development.

To conclude, the current study failed to bring new evidence about older children attribute belief attribution, children's low performance in the simple location-change task casting doubt about the reliability of our paradigm as an efficient false belief test. Thus, whether children are able to attribute beliefs about causal inferences, will remain an interesting question for future scientific inquiries.

# Chapter 4. How do children update their own beliefs and beliefs attributed to others?

# 4.1. Introduction

Human behavior is driven by a series of factors, such as the perceived states of affairs, one's current goals, as well as by the beliefs, knowledge and norms one is endorsing. However, oftentimes life provides us with different kinds of unexpected events and evidence that contradicts our prior beliefs, and we must resolve the perceived inconsistencies between the different factors, to pursue coherent action plans. Thus, two interesting questions emerge regarding how we track and resolve such inconsistencies, for instance, which beliefs or norms will prevail over others, and whether we also track inconsistency resolutions from other people's viewpoint, for correctly interpreting or predicting their behavior in social interactions.

Everyday examples are abundant. Imagine a case in which one holds the assumption that if Peter's car is in front of the house, he is at home. One day the car is right where Peter always parks it, yet we don't find Peter at home. How will we resolve such a situation, will we change our original assumption? Or let's take another case, in which for instance we have learned that whenever our neighbor's granddaughter is visiting, they always bake chocolate chip cookies together, and bring some over. We have seen that the granddaughter has arrived some hours ago, yet there are no signs of chocolate chip cookies. These events are not extraordinary ones: we learn simple contingencies or rules, and later we receive conflicting information. However, for maintaining a consistent belief set about the world, we should aim to eliminate inconsistency, for instance by discarding our original assumptions or the evidence we encounter. Importantly, this is not a first-person problem only. During social interactions, we may encounter situations in which others experience conflicting information. We may safely assume, that aiming at consistency must be a general feature of the cognitive system, and other people would also aim at holding a consistent set of beliefs. What are the principles that guide the revision of the previously acquired belief sets to overcome inconsistency? Would we expect our peers to think in a similar way as we do? Are they guided by the same principles in their belief revision as we are? In other words, would we expect them to arrive to the exact same conclusions as we do? How do such abilities develop?

Consider now a further example that involves two characters which is an event occurred in a personal experience. One day while taking the elevator, the following conversation took place between a 5-year-old boy and his mother. The mother was claiming that daddy must be home. When the child asked why she is saying that his curiosity was satisfied with appealing to the following rule: "if his car is in the front of the building, daddy is at home". When they entered their apartment, the mother expressed surprise as she didn't find her husband home. The child argued that he might have left for a walk, persuading his mother that this is a viable possibility, expressing surprise that his mom did not arrive to the same conclusion, whereas the mother thought that maybe she simply mistook daddy's car for another one.

In this case, experiencing evidence incongruent with their prediction, the child and his mother seem to have committed to different revision strategies:

i.) the mother abandoned what she has observed earlier, updating her belief about the car's presence, while preserving a belief about the predictive power of this cue and accepting the possibility that it might not be daddy's car that she has just seen in front of the house

ii.) the child updated the rule: the presence of the car may not be a perfect predictor, since there are a set of alternative scenarios justifying the presence of the car and the absence of Peter (e.g., that daddy took a taxi to get to work, or he went for a walk after he turned back home with his car from work and so on) iii.) the child expected his mom to also update the rule (by expressing surprise that she didn't), that whenever daddy's car is in front of the building, he must be home.

While belief inconsistent experience may be pervasive in early development, we expect inconsistencies to trigger revision from early childhood to adulthood. To better understand others and their behavior, for instance, for efficient collaborative action planning, it would be useful to understand how others deal with inconsistencies from early on. However, not having direct access to their thoughts complicates this challenge. Whether we expect others to converge to our strategies or not, we might assume that belief revision is generally guided by one principle: that similarly to us, others are also motivated to preserve consistency in face of new conflicting information, and such assumptions may guide reasoning early on.

In the current chapter, we aim to explore how children update their beliefs when they receive new, conflicting information with their previous belief set, and importantly, we also investigate these update processes when children have to deal with such inconsistencies from another person's divergent perspective. Do children expect others to use the same strategy when they decide how to deal with conflicting information?

Belief sets may not only contain beliefs about the state of the world (e.g., that the cup is in the cupboard), but also regularities (e.g., *if then* relations), like in the everyday examples illustrated earlier. Updating episodic information (e.g., the cup which was in the green cupboard is now moved to the blue cupboard) may sound intuitively simple. One has to assign new spatial coordinates to the object represented in the object file. However, updating rule-like relations may be more complex. One has to decide what to keep and what to revise from the previously acquired multicomponent belief sets in order to conserve consistency. However, people's revision strategies may depend on the content to be updated: people may revise beliefs differently about relations that contain stable properties or have a strong explanatory power about the world (e.g., laws of physics), and about relations that are extracted based on statistical regularities.

Children acquire new knowledge every day, and new information may conflict with their previously acquired beliefs. Thus, these questions may have a high prevalence in early childhood. While evidence indicates that there are specific patterns regarding adults' belief updating strategies (e.g., Elio & Pelletier, 1997, Elio, 1997, Polititzer & Carles, 2001), little is known about how children revise their beliefs when they are faced with new information. To our knowledge, most of the studies investigated how children revise their previously acquired belief sets when they encounter additional information, but only a few studies tested how they update their beliefs when they face inconsistency, and whether they use similar strategies to adults when they update their beliefs. Crucially, whereas infants and young children show some understanding of other's mental states, we know very little about how they revise a belief from a third person perspective in case of inconsistent information.

### How do adults update their own beliefs in case of inconsistency?

Identifying how people - children and adults - adjust their beliefs when contradictory information is detected may shed light on our understanding of how people construct their model of the world. The first empirical studies that aimed at revealing the specificities of belief revision investigated logical reasoning with conditional premises in adults (e.g., Elio,1997; Elio & Pelletier, 1997; Politzer & Carles, 2001). The authors investigated the concept of epistemic entrenchment – the intuitive notion that some elements of a belief set are more deserving to be retained than others in the face of contradiction, and that this may depend on the type of knowledge involved in belief revision. In this approach, beliefs about relational knowledge (often capturing causal information) plays a central role. Relational beliefs, induced here as

conditional premises, postulate a relational rule, possibly congruent with past observations. Establishing whether a contradiction is present is determined by the formal properties of the framework. Thus, whenever the antecedent of a previously established conditional premise is confirmed, the consequent should be concluded. Any classical contradictions of this conclusion, leads to inconsistencies to be resolved. This is exactly what these studies tested by examining adults' belief revision strategies.

There are different theories and perspectives on epistemic entrenchment. These investigations (e.g., for a review, see Elio & Pelletier, 1997, p. 426-428) are centered around a utility argument, suggesting that some beliefs are epistemically more privileged, because they are more useful. However, it is not entirely clear how one may decide what counts as a more privileged content. When facing information contradicting a previously acquired belief set, either the conditional rule, or the observed data can be revised. There are two possibilities concerning why one should be entrenched over the other.

One possibility is that some knowledge structures (together with the propositions that capture them) that express and transmit important regularities with high predictive power, (e.g., laws of physics) that are the outcome of extensive experience and reliable inductive processes, eventually on top of possibly innately available naïve physics intuitions, and therefore are subject to a greater entrenchment in face of contradiction than other beliefs that reflect the current state of the art.

The other perspective suggests exactly the opposite: what has priority to be retained in belief revisions is the observed data because these are the ones about which one can be most certain, whereas regularities expressed in a conditional form are merely hypothesis about the world that can be abandoned in case of a conflicting evidence. In the example mentioned earlier, one would revise the "rule" that whenever Peter's car is in front of the building, he is home, and keep the belief about the data, i.e., the observation that the car is downstairs.

124

As noted in the beginning, these strategies of entrenchments may depend on the kind of belief content involved in belief reasoning. Elio (1997) investigated whether the types of regularities - in these experiments expressed as conditional rules taking the form of if then relations - have an impact on belief revision strategies. Elio's assumptions derive from previous works on causal reasoning (e.g., Cummins, 1995) reporting that reasoning about causality may depend on the number of alternative causes explaining the effect (expressed in the consequent of a conditional) and the number of disabling factors, as these factors impact and prevent effects from occurring even in the presence of possible causes and deviate this way from the deductively validity of these inferences. This study tested different types of conditionals: familiar and unfamiliar definitions, promises and causal rules expressed as conditional statements (Elio, 1997). Furthermore, causal relationships were further divided into subtypes defined by the number of disabling factors. One of the main questions was how the amount of evidence supporting (or undermining) the believability of the conditional relation itself influences the kind of belief revision decisions that adults make. Participants were presented with a conditional statement, and a "data" statement that meant to establish an initial belief set that permitted an inference. Then a new, contradicting information was introduced, and people were asked to indicate which of the initial information - transmitted by the conditional or the non-conditional (data) statement - they no longer believed in order to overcome the contradiction.

Take the following examples based on a previous work from Cummins (1995):

- *I.* Premise 1. *If Jenny turned on the air conditioner, then she felt cool.* (p->q)
  Premise 2. *Jenny turned on the air conditioner.* (p)
  Contradicting information: *Jenny didn't feel cool.* (-q)
- 2. Premise 1. If Mary jumped into the swimming pool, then she got wet. (p->q)

Premise 2. *Mary jumped into the swimming pool.* (p)

Contradicting information: Mary did not get wet. (-q)

This study examined whether the type of knowledge people acquired has an influence on people's revision strategies. It was found that in case of causal relationships, subjects were more likely to revise their belief set by disbelieving the conditional statement, but only in the case of many disabling factors. For instance, it is easier to imagine several disabling factors that can suspend the conditional relation, in the first example (e.g., a case where Jenny was wearing a jumper, therefore she didn't feel cool, or she turned on the AC on a low level) than in the second example (where getting wet in the pool is straightforward, and it is hard to consider any disabling factor that would falsify the conditional relation).

As shown previously, people prefer to abandon the conditional over the observed data, but in case of causal relationships, the amount of disabling factors associated with the causal relation itself has an impact on people's revision strategies in a way that in case of few disabling factors, people are more unlikely to revise the rule. In all the other cases, abandoning the conditionals in case of a conflicting information can be explained by the idea that conditionals expressing loose regularities or associations are less deserving of entrenchment than data statements or observations in face of a conflicting information.

In another set of experiments, Elio and Pelletier (1997) used the very same paradigm to test people entrenchment strategies, but instead of comparing different types of knowledge, the authors simply investigated whether participants would use different strategies regarding conditional statements involving modus ponens (MP) or modus tollens inference (MT). In these experiments, participants were presented with a problem which consisted of an initial belief set, contradicting information and three alternative revision choices. For the MP and MT problems, the original sentence set included a conditional of the form if p *then* q and either the antecedent p or the negated consequent  $\sim$  q, respectively.

See the following example for an MT inference:

Premise1: *If Anna is home, the light is on* (p->q) Premise2: *The light is not on*. (- q) Conclusion: *Anna is not home*. (-p) Contradicting information: Anna is home. (p)

Whereas for MP, the example looked like the following:

Premise1: If it is Monday, Jonny wears a green shirt (p->q)

Premise2: It is Monday (p)

Conclusion: Jonny wears a green shirt (q)

Contradicting information: Jonny wears a red shirt (-q)

The new information in both cases contradicted the derived inference. Participants could choose between three revision types to reconcile the contradiction in logically different ways: 1. deny the conditional or 2. retain the conditional but reverse the truth status of the ground sentence (p) that permitted the inference, or 3. label the ground sentence (p) as uncertain. Their results indicated a difference between the modus ponens and modus tollens; participants were more likely to disbelieve the conditional when they were presented with MP, however regarding MT statements, they were more likely to label the truth status of the initial ground sentence (p) as uncertain. This might be explained by the possibility that MT problems are harder to generate, and participants may not accept them as consistent in general.

Politzer and Carles (2001) further investigated these issues. They used the same paradigms as Elio and Pelletier (1997), however in these experiments participants always had the opportunity to label the set of beliefs as uncertain, and not only to categorically deny it. The results were consistent with the previous studies in a way that participants were more likely to doubt (and label as uncertain) the conditional statement itself rather than to disbelieve the ground sentence. The tendency to reject the conditional belief was a function of the strength of their belief in the conditional relation itself: high subjective certainty leads to reduced update. They argue that it is reasonable to alter the degree of belief than to fully give up.

To summarize, the studies described above found stronger evidence for people's preference to revise, disbelieve or label uncertain the conditional statement rather than the observed data or the direct evidence itself. This was slightly different in the case with causal relations, where people's revision strategy depended on whether the causal relation contained sufficient disabling factors to disbelieve it in face of a new contradicting evidence. As discussed in the beginning of the introduction, these results are consistent with the perspective that what enjoys priority in belief revisions is the observed data. This might be explained that observed data are which one can be most certain, whereas regularities expressed in a conditional form are merely hypothesis about the world that can be abandoned in case of a conflicting evidence. However, the likeliness of the revision of conditional rule may depend on the type of conditional too.

In the following part of the chapter, we review the existing studies targeting children's belief revision abilities, and consequently we propose a set of experiments to investigate whether children's revision strategies are similar to adults from first person and third person perspectives, when it comes to revise beliefs regarding causal relations with different amount of disabling factors.

#### How do children revise their beliefs?

To our knowledge, only a few studies investigated how children revise their beliefs, and most of these studies tested whether they are able to revise their previously acquired belief sets when they are presented with some additional information which was unknown to them before. However, little is known about how children resolve inconsistency, when they face not simply additional, or previously unknown information, but information that is contradicting their previous beliefs. First, we will discuss research investigating children's abilities to revise their beliefs in various contexts (e.g., Bonawitz, Fischer & Scuhlz, 2012), and later address studies by Van Hoeck and his colleagues (2008, 2012) who tested children's ability to revise their beliefs in face of a contradicting information in a counterfactual reasoning.

A study by Bonawitz, Fischer and Schulz (2012) have asked whether young children favor plausible causal mechanism over statistical evidence, as it was suggested by earlier research (e.g., Schultz, 1982). Contrary to previous findings, where children simply sticked to their own initial domain-restricted arguments when they were asked to give an explanation about a physical state of a character (e.g., puppy's tummy hurts because he fell on the stairs), after participating in different training sessions (i.e., training inferences with multiple causal relations), they were more likely to revise their beliefs and rely on new evidence from a different domain, and they have not unduly entrenched their previous causal beliefs. For instance, after introduced with a character that had a bodily effect caused by a psychological state (for instance he was nervous because it was his first day at school), children readily updated their explanation: e.g., puppy's tummy can also hurt when he feels anxious, not only when he falls on his tummy.

Macris and Sobel (2017) investigated how 4-5-year-old children revise uncertain beliefs in light of disconfirming evidence, and whether the variability of counterevidence, and children's explanation about their initial beliefs impacts their belief revision abilities. In these experiments, children were presented with a machine (an ambiguous causal system) that was activated by different kinds of objects, and they were asked to make a guess about which object activates the machine. The system was ambiguous in a way that the object activating the machine had two properties: it was either cube shaped, or an object that had an interesting small internal part (e.g., it had small hole with a foamy texture inside). Later, the experimenter provided two new objects that varied in their features from the previous ones in the following manner: a cube (without an internal part) and a triangle with an internal part. At this point, children were asked to generate their hypothesis about which of these new objects will make the machine go. Later they were presented with evidence, which was not in line with their initial hypothesis, while in a second experiment they observed a more diverse set of stimuli that provided counterevidence (for instance, instead of two differing objects, six differing objects activated the machine). Furthermore, half of the children were asked to explain why a certain object would make the machine work. Results indicated that across both experiments, and in regardless of whether they were asked for explanation, children were more likely to revise their initial hypothesis than expected by chance, and the amount of counterevidence did not have a significant impact on their abilities to revise their beliefs about the causal system.

These studies mentioned above investigated whether children would entrench their previously acquired belief sets when they encounter new information that might contradict their previous beliefs. However, in most of these cases, it is unclear whether the presented counterevidence was directly contradicting their previous beliefs (as in the adults' studies, e.g., Elio & Pelletier 1997), or simply introduced a new or additional piece of information, which might have been unfamiliar to them. While these studies mainly tested whether children will be more likely to revise their beliefs in light of new information, the question emerges, how children would revise their beliefs about the different causal relations, which piece of belief they would abandon and which they would keep in light of a conflicting evidence.

An interesting study investigating this question was conducted by Van Hoeck and his colleagues (2008, 2012), who examined 7-year-old children's belief revision strategies in counterfactual reasoning tasks. These studies used similar paradigms to the ones discussed in

the studies with adults (Elio & Pelletier 1997) but implemented them in a pretend play context by using a story about heroes and thieves. Consider an example of the statements used:

Premise 1. All heroes wear a white hat (p->q, general premise)

Premise 2. This person is a hero. (p, particular premise)

Conclusion: *He wears a white hat.* (q)

Counterfactual premise: Let's pretend that this person wears a black hat.

In this picture matching task, after children acquired the rules (the general premise), they were presented with the particular premise and a conflicting information involving a counterfactual assumption e.g., let's pretend that hero who previously wore a white hat would now wear a black hat – and at this point, a picture of black hat was placed on top of the picture of the person with the white hat, directly contradicting the previous premises. Children were offered two options as solutions to the counterfactual problem: they could either to select a picture about the hero with a black hat (revising the initial rule e.g., in the following way "the heroes can also wear black hat now") or they could select an option stating that this person is a thief from now on (selecting the picture of a thief, thus rejecting the data, that this person is a hero). The results indicated that 70% of the children selected the option to revise the particular premise (by rejecting the data that this person is hero) when they were asked to solve the inconsistency posed by the counterfactual task. This is consistent with a previous study of Revlin and colleagues (2001) who examined adults' revision strategies in counterfactual reasoning tasks, indicating that contrary to other findings (e.g., Elio, 1997) people tend to solve inconsistencies in a different way when they are presented with counterfactual problems, and are more likely to reject the observed data itself in favor of general rules. While the authors argue that this pattern can be explained by the possibility that revisions of a strongly believed general statement might demand more cognitive resources that the revision of a particular statement, and this might impact children's performance at this age, such explanation is difficult to apply to the adult data. Furthermore, in such pretense scenarios encoding and updating the observed data might differ from an everyday situation: here the data is not an actually observed event, but a stipulated counterfactual event (*Let's pretend that this person wears a black hat*), which may also play a role in participants' revising strategies. Therefore, it is still a question how children deal with conflicting information regarding conditional rules outside pretense?

# 4.2. Aims of the current studies

In the current experiments, we aim to investigate whether children's revision strategies are similar to that of adults' when they receive contradicting evidence, and how they revise beliefs from a third person perspective. In order to explore these questions, we have created scenarios that allowed an easy manipulation of supporting and contradictory evidence. We have chosen to test rules that were either the result of a causal constraint or to the contrary, they were causally underdetermined. Imagine a tube that is designed as an inverted Y and a ball. A small or big ball is dropped into the upper entry point and exits through one of the arms, of which one is wide, and one is narrow. After repeated exposure to these events one may notice that if the ball is big, it always exits on the wide arm, if the ball is small, it will exit on the narrow arm. Note, that in case of the small ball the diameter of the tubes poses no physical constraints the ball, and it has equal prior probability to exit through any of the arms (Unconstrained *trajectory*). This is not the case for the big ball where some apriori constraints apply. It is easy to see that the diameter of the pipe is crucial in this respect. For instance, if one of the arms has a diameter smaller than the ball, this will delimit the possible motion trajectories, acting as a disabling factor, that prevents a big ball to travel through the pipe (Causally determined trajectory). Here we aim to investigate, whether the presence or absence of such physical constraints has an impact on children's belief revision strategies as it is the case for adults (Elio,

1997). The presence of a physical constraint for the big ball, may turn these rules more resistant to revision compared to the case when there is no physical constraint (small ball), and where there rule itself is a simple statistical regularity. See Figure 1 illustrating the absence or presence of physical constrain regarding the ball's trajectory.

Information about the kind of update children prioritize can be collected in the following ways. In our set of experiments, we present participants with video-animations depicting the pipe-system described above. After the participants are exposed to a starting configuration (revealing the position of the narrow/wide arms, the introduction of the small/large ball), children are asked to predict where the ball will fall. During a training period, they repeatedly see that if a big ball is released, it will fall into the box placed under the wide arm while if a small ball is released, it falls into the box under the narrow pipe. In repeated trials, after the prediction is made, they can observe the route of the ball and they also receive additional feedback about the location of the ball signaled by the noise that the box provided when it was shaken by a hand.

The test phase consisted of three trials: in the first, congruent trial the ball always falls into the box under the pipe analogous with its physical size (e.g., the small ball into the box under the narrow pipe), consistent with the previously learnt rule. However, the second and the third trials were incongruent with the previously learnt rule: e.g., after the small ball was released and children could not observe its trajectory, both boxes were shaken, and the box under the wide pipe made the characteristic noise, indicating the presence of the ball. Crucially to our question, after they received this feedback about the ball's final location that was in contradiction with what they experienced earlier, children were asked which ball they think is inside the box, the small or the big one?



*Figure 1.* In training the small ball always exits through the narrow pipe, while the big ball through the wide pipe. However, in one case the ball has no causally determined trajectory (picture on top-small ball) given the absence of physical constrains, whereas in the other case (picture on the bottom) physical constrains are present and the ball is causally determined to exit through the wide pipe.

If children's update strategies are similar to that of adults, contradictory information should lead to rule update in all versions of our experiment, but, overall, more rule update should be documented in the versions where the rules are physically unconstrained as it was also observed in the study by Elio (1997) in adults. However, there might be differences across development about how younger and older children are able to represent possibilities and use this information in their decision making. Representation of possibilities may play an important role in guiding children's update strategies. Based on some previous studies showing that children around the age 3 have difficulties in implementing optimal decisions that require modal representation of alternatives (e.g., Redshaw and Suddendorf, 2016; Suddendorf, Crimston and Redshaw, 2017) we assumed that the expected belief revision strategies will be more evident in the 5-year-old children.

Furthermore, if we find that children's revision strategies are defined by specific patterns in first person reasoning, the question emerges whether similar patterns would prevail when they have to perform such computations from a third person perspective. How would young children update a belief from a third person perspective, and when a third person's perspective contradicts with their own perspective? These questions will be addressed in Experiment 3 and 4.

When a child observes another agent who is faced with conflicting evidence in the above tube task, how would she resolve the conflict from a third person perspective, and when such event is coincides or not with one's own perspective? Would children's update strategies be similar or different in first person and third person reasoning? Given that one has privileged access to one's own knowledge, but others' knowledge can only be inferred (Keysar et. al, 2003), there might be differences in updating one's own vs. updating attributed beliefs. One possibility would be that given that attributing relational contents to others (e.g., 'She believes that if the small ball enters the device, it will fall out from the narrow tube') may more complex than simply attributing a belief about an object at a location, therefore such contents may pose greater challenge to children. Thus, if causal relations are more difficult to represent and possibly to attribute, and thus may be more fragile, children will be more likely to update the rule (regarding the causal relation itself) in a false belief task. Furthermore, co-witnessing events with someone else (i.e., the child and the agent witness together that the big ball falls into the tube) may impact how children revise beliefs. Some studies suggest, that observing events together with others enhances the encoding of episodic information, e.g., gaze cueing results in better memory encoding for a cued target (Gregory & Jackson, 2018) and moreover, the social context, even without ostension enhances encoding objects in memory (Howard & Woodward, 2019). Therefore, it might be the case that if children observe the events with someone else, they will encode the observe data stronger, and will be more likely to update the conditional rule and keep the observed data from another person's perspective. One might argue that the same might apply for co-learning a rule with someone else, but to our knowledge currently there is no evidence suggesting that co-witnessing enhances rule learning. In any case, if both the data and the rule is encoded stronger due to co-witnessing, these two effects may cancel out.

According to another possibility, children might expect an agent to abandon the observed data when the agent faces contradicting information. From a first-person view, the observed data may be more entrenched because it serves as the bases of the beliefs that one can be most certain about. Instead, the learned regularities may simply reflect a hypothesis about the world that can be abandoned in case of a conflicting evidence (Elio & Pelletier, 1997). However, this argument might not apply to the case when children observe an event from someone else's perspective: they might not be certain that an agent has seen and encoded the observed data in the same way as they have seen and encoded it. After all, we do not have direct access to other's perception either. Thus, if they encode the data from another's perspective as less certain, they might reduce their commitment to the observe data, while preserving the rule. As a result, the belief about the data might be revised.

The remaining third theoretical possibility is that there may be no asymmetries regarding the update processes in first person compared to the case when children need to update an attributed belief. One could argue that children may rely on the similar reasoning processes in first person and third person computations, as suggested by findings pointing to common neural substrates recruited in object maintenance and action prediction tasks in first person and third person perspectives (Kampis, Parise, Csibra & Kovács, 2015; Southgate & Vernetti, 2014). In line with this possibility, children might expect another agent to arrive to the same conclusions/updates as they would do.

In order understand children's belief updating strategies, in Experiment 1A and 1B we investigated how 3- and 5-year-old children update their beliefs regarding a rule when this is not justified by a plausible mechanism (i.e., there are no identifiable constrains affecting the experienced motion trajectories). Indeed, if there are no obvious mechanisms responsible for the regularity, the observed statistics cannot be not back up by a specific causal justification. In this case, if children have access to adult-like belief update strategies at this age, in face of contradicting information they will be more likely to update the rule and preserve the data what they observed. In Experiment 2 we further aim to test 5-year-old children belief revision strategies when they acquire rules that describe event outcomes governed by physical constrain (disabling factors) therefore it may be harder to revise despite the conflicting evidence. Similarly to Experiment 1, here we predict that children will be more likely to update the conflicting evidence. Similarly to Experiment 1, here we predict that children will be more likely to update the observed data, but we expect to see a stronger effect when the rule has no physical constrains.

In Experiment 3 we aim to test 5-year-old children's update strategies when they cowitness the events together with an agent and whether the presence of an agent has an impact on their update strategies, especially in the case when children are exposed to conflicting information regarding a conditional rule with physical constrains. We hypothesize, that if cowitnessing events influences how children process certain elements of conditional events, then children's update strategy might be different in a way, that they will be more likely to update the rule itself even in this case (as co-witnessing enhances the encoding of the observed data).

Finally, in Experiment 4 we test 5-year-old children with a false belief version of the task to investigate how children update other's false beliefs. If the representation of conditional rules is more complex, then we hypothesize that children might update someone else's false belief by abandoning the conditional rule.

# 4.3. Experiment 1A and 1B. Strategies of 3 and 5-year-old children's for revising rules that are causally opaque

The amount of predictions children can make only based on their naïve physics is boundless. If there are no immediate physical or other information available to support expectations about the outcome of an event, we nevertheless expect children to extract the relevant regularities. In the first two experiments, we aimed to test children's belief updating strategies in a situation where they are exposed to events predictable only based on statistical evidence. What kind of update follows if their predictions are violated?

Our predictions are the following: if children, similarly to adults (Elio, 1997), revise the conditional rule in such cases, they should indicate that the smaller ball is in the unexpected box, as they revise the rule, that 'if the ball is small then it falls into the box under the narrow pipe', but keep the observed data, i.e., that they saw the small ball in the beginning. However, if children revise the observed data and retain the rule, they should indicate that it is the big ball in the box under the wide pipe in the test phase and reject the perceptual information that they saw the small ball in the beginning.

#### 4.3.1. Methods

#### Stimuli

The movies were designed as 3D animations in Autodesk Maya 2016. They were then exported as movies at 25 fps and further edited with QuickTime Player 7 software and Adobe Premier Pro with real-life videos. Stimuli were presented through an online platform <a href="https://slides.com/">https://slides.com/</a> that parents could open on their home computer, and the presentation itself was controlled by the experimenter from her computer.

# Participants

Participants were recruited from the Central European University's database, and they were tested online. Children were seated in front of their computer or laptop with their parents at their home, and they observed the stimuli from their device, while a female experimenter controlled and led the experiment from her side. Thirty-two children participated in the study in each age range, 17 females in the younger, and 16 females in the older age group. In the younger group, children's age ranged from 3 years 3 months to 3 years 11 months (M=3 years 9 months), whereas in the older age group their age ranged from 5 years 1months to 5 years 11 months (M= 5 years 5 months).

In the younger age group, additional 31 children were tested but excluded due to the following reasons: lack of cooperation (2) impaired language abilities (1) failure to complete at least half of the training trials (6) technical error (1) and incorrect answer to the first incongruent test trial that resulted in the lack of revision of the events based on the incongruent evidence turning our manipulation invalid (21). In the older age group, additional 6 children were tested but were excluded due to their incorrect answer to the first incongruent test trial.

#### Procedure

During the sessions, caregivers were asked not to interact with the child. Both the stimuli and the participants were video recorded for later analysis. First, children received a simple color naming game as a warm-up.

The experiment contained 7 trials in total: 4 training and 3 test trials. Training trials were served as a function to teach children the conditional rule: whenever a small ball is present, it will exit through the narrow arm, whereas whenever a big ball is present, it will exit through the vide pipe. For each participant, the training consisted of 2 transparent training trials

followed by 2 opaque trials. During the training, the direction of the exit was counterbalanced in ABBA order.

#### **Transparent training trials**

The aim of the transparent training trials was to familiarise children with the structure of the pipe and the sequence of events. In the beginning of the video, an occluder opened (2s) revealing the device itself. This was followed by placing two boxes placed under the pipes (8s) by a hand from two directions (left and right, order counterbalanced). Then a mechanical arm was displayed in the upper part holding either a small or a big ball (2 s). At this point, children were asked to predict into which box will the ball fall. To make children's prediction more straightforward, two coloured carpets were placed under each box, a yellow and a green (sides counterbalanced). Children were asked to name the colour of the carpet under the predicted box. Their answers also allowed us to test their understanding about the size of the ball i.e., whenever they witness the big ball in the beginning of the trial, they should predict that it should fall into the box under the wide pipe. After the child answered, the second part of the video started, the arm released the ball which fell into the pipe system and went through either the wide or the narrow pipe and fell into the corresponding box (3 s). At this phase, the route of the ball and its arrival to the corresponding box was visible to the infants. After the ball arrived at the corresponding box, a hand grasped each box and shook it: the box with the ball made a noise, indicating the presence of the ball inside (10 s). To ensure that children pay attention to the screen, all events (e.g., movement of the occluder and the ball) were accompanied by interesting sounds. Children observed two transparent training trials. Illustration of the transparent training trials can be seen on Figure 2.



*Figure 2.* Schematic representation of the transparent training trials (small ball example) in Experiment 1. After the occluder opened (1), a hand placed an empty box under each pipe (2,3) this was followed by the arrival of the small ball transported by a mechanical arm (4) the mechanical arm released a small ball which entered the pipe system and rolled through the narrow pipe (5) a hand shook both boxes (6,7) and the box under the narrow pipe provided the appropriate sound (8).

# **Opaque training trials**

The aim of the opaque trials was to help children forming predictions in absence of visual access to the route of the ball through the pipe. The videos presented here were the same as in the transparent trials, with the following differences: after the ball was released by the mechanical arm its route was not visible as the pipes at these points were opaque. However, to facilitate learning, participants were still able to see which box the ball is falling into. Participants observed 2 opaque training trials. Figure 3 depicts the opaque training trials.



*Figure 3*. Schematic representation of the opaque training trials (small ball example): the sequence was the same as in the transparent training trials, expect that the arm of the tube was now opaque

# **Test trials**

The test trial started with the hand placing the two boxes under the pipes (7 s). This was followed by the occluder (the same used in the beginning of the training trials) covering the end of the pipes, in order to block all the visual access of the possible route of the ball (2). This was followed by the arrival of the mechanical arm with the small ball (2s) and the release and falling of the ball (5s). At this point, children were asked to indicate where ball could have fallen. After they provided an answer, the second part started. The occluder opened (2s) and one hand shook one box (5s) which was followed by the second hand shaking the other (5s). In the test phase, always the shake of the second box provided the sound and thus confirming the final location of the ball, which was either consistent with the expectation, or inconsistent, in which case, the box under the wide pipe provided the sound. Children watched 3 test trials (1 congruent followed by 2 incongruent). Schematic representation of the test trials can be seen on Figure 4.



*Figure 4*. Schematic representation of the test trials. A hand placed an empty box under the narrow and the wide pipe (1-2) an occlude closed, covering the ending of the pipes, and the mechanical arm brought in the ball (3) which then was released (4) this was followed by the opening of the occlude, and the hand shaking both boxes (4-5-6) in the congruent trial, the box under the narrow, whereas in the incongruent trial the box under the wide pipe provided the sound

#### 4.3.2. Results

#### Younger age group

During the training trials, children were required to predict where the ball will fall, and we coded any answer as relevant that specified either the balls' final location (e.g., the box on the green rug, or the box under the narrow pipe) or specifying the balls' motion trajectory (e.g., the small ball will fell through the narrow pipe). Data was coded offline from the recordings, and 30% of the videos were second coded by a research assistant blind to our experimental hypothesis. Agreement was 100% between the two coders across all experiments.

In average, the 3-year-olds included in the final sample were correct in 78% of the training trials, suggesting that they understood the rule. Children's performance on the first congruent test trial, where 21 children (65%) answered correctly (p=.110, binomial) was not
significantly better than chance. We analyzed children's update strategy based on their answers on the first incongruent trial. Children did not show a specific pattern, they revised their beliefs in a random manner: 16 children (50%) updated the conditional rule (by indicating that the small ball is in the box under the wide pipe), and 16 children updated the observed data (by indicating that the big ball is in the box under the wide pipe), a pattern that was not different from chance (p=1). In addition, we calculated binomial Bayes Factors (BF) contrasting the null hypothesis (equal probability of updating the data and the rule) to the alternative hypothesis of higher probability for the expected update strategy (using the default hyperparameters in JASP). The estimated Bayes Factor (BF = 4.6) suggests that the data were moderately in favor of the null hypothesis.

We further analyzed whether first counterevidence they received at the end of the first incongruent test trial would impact children's performance on the second incongruent test trial. If children updated the rule, this might impact their performance in a way that they would indicate that the small ball could have fallen into the box under the wide pipe in the second incongruent test trial. However, 25 children (78%) still indicated, that the small ball fell into the box under the narrow pipe, and their update pattern remained at chance, 16 children (50%) indicating that it must be the small ball under the wide pipe.

## Older age group

Older children's performance in the training trials was 88%, indicating that they successfully understood the conditional rule. Their performance on the first congruent test trial further indicated their understanding of the rule, where 30 (90%) children correctly indicated that the small ball fell into the box under the narrow pipe (p>.001, binomial). In the test phase, children revised the rule more often compared to the observed data: In the first incongruent test trial, 25 children (78%) out of 32 indicated that the small ball is present in the box under the

narrow pipe (p=.002, binomial). Estimated Bayes factor favored the alternative hypothesis (BF=77.2). We analyzed whether the first incongruent evidence impacted older children's performance on making a post diction on the second incongruent test trial, however, 23 children (71%) said that the small ball fell into the box under the narrow pipe (congruently with the previously learnt rule). After receiving incongruent feedback, their update strategy did not change, on the second trial, 26 (83%) children answered that the small ball is in the box under the wide pipe, thus they kept updating the rule. Graphs indicating children's performance in Experiment 1A, and B are shown in Figure 5.

In addition, we have also compared the two groups' performances on the two trials and found that older children compared to younger children are more likely to update the conditional rule both on the first test trial (p=.0360, Fisher's exact test), and the second test trial (p=.016, Fisher's exact test).



*Figure 5*. Children's update strategy in experiment 1A and 1B. The graph represents the percentage on how many children updated the rule across the two age groups (younger on the left, older on the right)

#### 4.3.3. Discussion

In experiment 1 A and B, we tested 3 and 5-year-old children's ability to update their beliefs regarding a rule extracted from events that involved no physical constrains. During the training phase of the experiment one of the rules could be acquired only by extracting the right statistics: children learned that the small ball would always fall into the box under the narrow pipe, whereas the big ball, given its diameter, could only fall into the box under the wide pipe. Older children were more likely to revise the rule when facing information contradicting their previous beliefs, while younger children did not have a preference for a such update strategy, despite that their performance indicated a good understanding of the rule (the participants included in the samples learned the rule we exposed them during the training trials in both revealing a performance beyond 80%). Younger participants updated their beliefs randomly by rejecting either the rule or the observed data: half of the children indicated that it must be the small ball in the box under the narrow pipe, whereas the other half of the children stated that it must be the big ball in the unexpected location. We should note, however, that overall, participants in the younger age-group were less efficient learners: In the final sample, while children's performance on the training trials was high (78%), only 21 children answered correctly to the first congruent trial (65%). Twenty-one participants produced incorrect answer to the first incongruent test trial, hence they were not included in our final sample (incorrect application of the rule preventing a valid experimental manipulation), while there were only 6 participants in the older age group that applied the rules incorrectly.

Thus, when they encountered counterevidence of the previously learnt rule, only the older children showed a specific strategy to solve inconsistency by entrenching the observed data and updating the rule. After they witnessed that the target ball is the small ball, but the feedback revealed a violation of their prediction, suggesting that contrary to their experience,

the ball fall in the box under the large pipe, they were more likely to stick to their perceptual evidence, indicating that the ball in the box under the wide pipe is the small ball. Their choice is indicative of a rule change. Importantly, during the experiment children did not experience any evidence that the small ball can fall into the box under the wide pipe.

The absence of the constraints in Experiments 1A and 1B means that physically both motion trajectories are equally possible, and that there are no identifiable physical principles that guarantee the stability of the experienced statistics. Thus, an important factor supporting rule update in these participants is their ability to represent the possibilities. Participants in our older age group may have access to these modal concepts and take into account the possibilities with a great ease, while this is not the case of the 3-year-olds as show by Redshaw and Suddendorf (2016; Suddendorf, Crimston and Redshaw, 2017) in a task that requires children to perform decisions while taking into account multiple possibilities simultaneously (see Leahy and Carey, 2020 about the development of modal concepts).

# 4.4. Experiment 2. 5-year-old belief revision of rules supported by physical constrain

In Experiment 2, children had to learn the same rules as in Experiment 1, but now they were tested with a version where the supporting causal mechanism were identifiable: e.g., the large ball could travel through the large pipe, but it was prevented to move through the narrow one. Thus, the diameter of the tube served as a physical constrain narrowing down the possible motion trajectories to one. The test phase explored children's belief revision strategies by contrasting the resolutions they provide for the different experimental conditions when they observed the big ball falling into the box under the big (congruent trials) or the narrow pipe (incongruent trials). As the big ball, physically can only fall down through the wide pipe, it is hard to come up with alternative causes or disabling factors that would make it to fall into the

box under the narrow pipe. Therefore, we expected that in this case children will be less likely to update the conditional rule, similarly to adults as in Elio & Pelletier (1997) when they faced a conflicting information regarding a causal rule with few disabling factors. However, we predicted that children at this age in the no constrain group will be more likely to update the conditional rule and retain the observed data in face of contradiction. Figure 6 illustrates the trial sequences of the test phase in Experiment 2.

#### 4.4.1. Methods

#### **Participants**

Similarly to Experiment 1, participants were tested online. The final sample consisted in 32 children (15 females). Children's age ranged from 5 years 1 month to 5 year 11 months (M=5 years 6 months). Nine additional children were tested but were excluded due to their incorrect answers on the first incongruent test trial (N=6), lack of cooperation (N=2) or because they failed to complete at least half of the training trials (1).



*Figure 6.* Trial sequences in the test phase of Experiment 2, demonstrating event outcomes governed by physical constraints. In the congruent condition the box under the wide pipe provided relevant sound indicating the presence of the ball, whereas in the incongruent condition, the box under the narrow pipe made the noise after it was shaken by the hand.

## 4.4.2. Results

5-year-old children's performance on the training trials was 91%, indicating that they successfully learned the rule. Children's performance on the first congruent test trial further confirms their success: 24 (75%) children indicated that the big ball fell into the box under the wide pipe (p=.007, binomial). Children, however, updated their beliefs randomly, 16 out of 32 revised the rule after the first incongruent test trial, which did not differ significantly from chance (p=1, binomial). We contrasted the null hypothesis (equal probability of updating the data and the rule) to the alternative hypothesis for calculating the binomial Bayes Factors (BF). This led to an estimated Bayes Factor that favored the null hypothesis (BF=4.6). Since participants were presented with two incongruent test trials, the counterevidence experienced on the first incongruent trial may have an impact over participants' judgments on the second

test trial. On the second incongruent test trial 27 (84%, p<.000, binomial) children retained the rule and indicated that the big ball fell into the box under the wide pipe. Their update strategy, however, did not differ from what they have shown on the first incongruent trial: 15 children (45%, p=.86, binomial) updated the rule, by indicating that the big ball is in the box under the narrow pipe.

By comparing Experiment 1B and Experiment 2 (Figure 7) we tested whether children's revision strategy depended on presence of a plausible causal explanation and found that children were more likely to update the rule in the no constrain condition of Experiment 1B than in the high constrain condition of Experiment 2 (p=.03, Fisher test).



*Figure 7.* Children's update strategy in experiment 1B and 2. The graph represents the percentage on how many children updated the rule regarding the conditional with no physical constrain (left) and with physical constrain (right)

#### 4.4.3. Discussion

In Experiment 2 (similarly to Experiment 1), we asked children to predict the future location of different balls having different diameters when they slide in a Y-shaped tube, such

that their size could be used as a reliable predictor of their trajectory: if the ball was big, it always exited through the wide arm, if the ball was small, it will exit on the narrow arm. This time, however, (differently from Experiment 1) we tested how 5-year-old children update their beliefs regarding a conditional rule that predicted a motion trajectory, where this trajectory unfolded according to a well-defined physical constrains. This was the case of the large ball. Given its size, it couldn't exit though the arm that had a smaller diameter. This information alone could directly justify the trajectory. Importantly, justifications of this kind were absent in previous research investigating children's ability to revise their beliefs (see for instance Van Hoeck et al. 2012). The main difference is not in the verbal/non-verbal character of the tasks, but the nature of the scenario that invokes few or no disabling factors at all that could prevent them to believe the conditional rule.

As predicted, children in this experiment were less likely to update the rule, as the big ball's route was causally determined. Similarly, to adults, 5-year-old children showed a preference to update the conditional rule to solve inconsistency in Experiment 1, however, this tendency disappeared once the rule was supported by transparent causal factors (i.e., the physical constraint over the motion trajectory). Children were not likely to update the observed data either, they updated their beliefs randomly.

# 4.5. Experiment 3. 5-year-old children belief revision from a third person perspective

In Experiment 3, we tested whether the presence of another agent impacts children revision strategies when they face conflicting information regarding a rule that has few disabling factors. In order to test this question online, we modified the experimental settings in two different ways. In one version, we modified the video animations in a way, that we added an extra digital background containing a TV screen, and an animated bear, who observed the screen (see Figure 8). In this version, children were told that the bear is watching the events with them on this TV screen, and they were required to answer the questions from the perspective of the bear. In the other version, children were simply asked to bring their favorite puppet with them to participate and were told that their puppet is also observing the game, and they were required to answer the questions from their puppet's perspective. These manipulations served as a between group factor in Experiment 3, where we only tested children with the condition where the outcome was determined by the physical constraint, therefore children observed the big ball in the test phase.

#### 4.5.1. Methods

## **Participants**

16 children participated in the screen condition, and 13 children in the puppet condition (16 females). Overall, children's age ranged from 5 years 1 month to 5-years 11 months (M=5 years 5 months). Six children were also tested but they were excluded due to their incorrect answers to the first incongruent test trial.

<image>

**Figure 8.** Schematic representation of Experiment 3 (training trials). The puppet first greets the child (1) and turns back to observe the screen (2) this is followed by the same training trials as in Experiment 1-2 (3-7)

# **Procedure and stimuli**

The procedure was the same as in Experiment 2, besides two things that were modified in order to implement a social context. In one case, the animations were modified in a way that a digital screen was inserted in the original video, and an animated plush puppet was observing the events (screen version). In the other version, we left the original animations, but we asked participants to bring along their favorite plush or puppet, who should sit with them in a way that he is able to see the same screen as the child (puppet version). In both versions, children were asked to answer all the questions from the perspective of the puppet, for instance "What does the puppet think, where will the ball fall now?"

#### 4.5.2. Results

Children's answers showed the same pattern in both conditions; therefore, we combined the two groups. Their performance in the training trials (92%) indicated that they understood the rule which was further supported by their performance on the first incongruent test trial,

CEU eTD Collection

where 25 children (86%) answered correctly that the bear thinks that the big ball fell into the box under the wide pipe (p<.001, binomial). Out of the 29 children, 16 updated the rule in the first incongruent trial (7 in the puppet condition, 9 in the screen condition). This did not differ significantly from chance (p=.7, binomial). Similarly, to the previous experiments, we analysed whether their performance on the second incongruent test trial was impacted by the counter evidence received on the first incongruent trial. Children's performance was slightly lower, only 20 children (68%) answered correctly, but their update strategy on the second test trial did not differ from chance, 16 children (55%) updated the rule (p=.7, binomial).

Bayesian analyses favored the null hypothesis (BF=3.7) Children's performance from Experiment 2 and 3 is shown in Figure 9.



*Figure 9.* Children's update strategy in Experiment 2 and 3. The graph represents the percentage on how many children updated the rule across the two versions (first person update on the left, third person update on the right)

#### 4.5.3. Discussion

In Experiment 3, we aimed to test 5-year-old children belief updating strategies in a scenario where they were required to update the belief from another character's perspective, which was congruent with their own perspective. Similarly to Experiment 2, in this experiment the rule predicting the behavior of the ball was directly determined by physical constrains. We investigated whether the presence of another agent would impact children's updating strategies. We predicted that if co-witnessing enhances the encoding of the observed data, children will be more likely to entrench it, and update the rule itself if they experience inconsistency. Although children's tendency to update the rule in Experiment 3 was slightly higher than in Experiment 2 (55% compared to 50%), this did not differ from chance. It is possible, that co-witnessing enhances not only the encoding of the observed data, but also rule learning, therefore has no specific impact on children's updating strategies. We also assumed that it could be the case that some of the children simply ignored the puppet perspective, and simply answered from their own perspective.

Therefore, we conducted another experiment, where during the test phase, the puppet left the scene and had a perspective that was incongruent with the participant's own perspective. In this case, children were always required to take into account the puppet's perspective as it differed from their own (and we were able to exclude the participants who answered from their own perspective). In Experiment 4, we tested whether children are able to attribute false beliefs to an agent about the conditional rule of the pipe system, and more specifically, we tested how would they update this attributed belief, when the puppet received the conflicting information.

### 4.6. Experiment 4. Revising attributed beliefs in 5 and 6-year-old children

During our social interactions, we may encounter situations in which others experience conflicting information. Some of these conflicts may happen in the context of false beliefs. With Experiment 4 we planned to achieve two goals. First, we aimed to find out whether 5-6-year-old children are able to attribute false belief involving rules that predict the motion trajectory of an object. Second, we aimed to identify children's revision strategies, by asking whether they modify first-person beliefs and attributed beliefs in a similar manner after a conflict is detected. Similarly to Experiment 2 and 3, in the test, participants were supplied with contradictory information about a previously learned physically determined motion trajectory (e.g. a large ball that can move in a pipe with the corresponding diameter). In case of arbitrary rules (e.g., "if the ball is small, will exit on the narrow pipe") children may simply attribute a belief for the puppet about each possible motion paths the small ball can take. As it is physically possible for the small ball to travel both in the narrow and in the large pipe, the box under the narrow or the box under the wide pipe are equally good candidates for its final location.

As in Experiment 3, we introduced an animated character (a bear), who observed the same events as the participant. These events, however, unfolded on a screen that was part of the video animations. The training trials were exactly the same as the screen condition from Experiment 3. Throughout the training we asked children to answer questions from the perspective of this character (e.g., "what does the puppet think, where will the ball fall?"). The false belief manipulation implemented in the test phase had the following characteristics: first, children only observed congruent test trials (the big ball would always arrive at the box under the wide pipe), second, the puppet had a false belief about the outcome. In order to implement this, before the ball fell into the tube, the puppet left the scene (at this point the experimenter also commented that the puppet left, and therefore he is unable to see what is

happening). After the big ball fell into the box, another character (an animated frog) came into the scene and took away the big ball from the box and left the scene. Later the same character brought another object (a cube) and placed it into the box under the narrow pipe. The cube was used to avoid answers corresponding to the reality. Relocating the large ball would not allow us to distinguish between responses based on reality bias and rule update.

After this point, the bear came back to the scene accompanied by the experimenter's comment: "the puppet came back but he didn't see what happen". Then we asked the child: "what does the bear think where is his ball now?" If children successfully attributed a false belief to the puppet, they indicated that he would think that the ball is in the box under the wide pipe. After the false belief test, both the puppet and the child received feedback: a hand shook both boxes, but the box under the narrow pipe indicated the presence of an object by providing a sound. At this point, the child was required to update the belief from the puppet's perspective. To test how children updated the attributed belief, we asked them what does the character think, who did not see what happened, which ball is in the box? As rules may be more difficult to be attributed, we hypothesized that children will be more likely to update the rule from the character's perspective. Therefore, we predicted that if children are more likely to update the rule, they should indicate that the puppet thinks that the big ball is present in the box under the narrow pipe. Figure 10 illustrates the false belief test trials.

#### 4.6.1. Methods

#### **Participants**

32 children participated in experiment (15 females). Their age ranged from 5 year 6 months to 6 years 11 months (M= 6 years 5 months). 12 children were also tested, but they

failed to attribute a false belief to the puppet in the first false belief trial and therefore were excluded from further analyses.

#### **Training trials**

Children participated in the same training as in Experiment 3 (screen-condition).

## **Test trials**

The first test trial was the same congruent test trial as in Experiment 3. The second test trial was a false belief trial. The false belief test trial started in the same manner as in Experiment 3, till the point of the arrival of the mechanical arm with the big ball (2 s) At this point, the puppet left the scene and the experimenter commented: "Look! The bear has left now. He cannot see what is happening."

After this, the big ball has been released and fell into the box under the wide pipe. This was followed by the arrival of a second character, an animated frog who arrived on the side of the box with the big ball (12 s) and who took the big ball out from the box and brought it away, and after 2 seconds reappeared from the other side of the screen and placed a cube into the box under the narrow pipe (12 s). In the meantime, the experimenter commented to the child: "Now look, there is a frog arriving, and he wants to play a trick on the bear, so he brings his ball away, and places a cube into the other box."

After the frog left, the bear came back (2 s) and the experimenter said the followings: "Now the bear is back, but as he was away, he didn't see what happened. What does he think, where is his ball now?"

After the child indicated his response, the occluder opened (2 s) and one hand shook the box under the wide pipe (5 s) which was followed by the second hand shaking the other box under the narrow pipe (5s). In the test phase, it was always the second box shaking that provided

the sound and thus confirming that there is an object inside the box under the narrow pipe. This information always inconsistent with the expectation of the puppet in all trials. At this point the child was asked: "The bear did not see what happened, what does he think, which ball is inside this box?"



*Figure 10.* Schematic representation of the false belief test trials: the puppet turns towards the screen and observes the two empty boxes being placed under each pipe (1-3) he observes the big ball's arrival, but before its release, he leaves the scene (4) the occlude opens and the frog comes in (5-6) he takes the big ball out from the box, takes it away and places a cube into the box under the narrow pipe, and then leaves (7-9) the bear comes back, (at this point the participant is required to answer the false belief question) and observes that the box under the wide pipe is empty, whereas the box under the narrow pipe provides the sound (10-11) the child is being asked what does the puppet think about the size of the ball (12).

# 4.6.2. Results

Children's performance on the training trial was 89% and 29 children (90%, p<.000,

binomial) answered correctly to the first test trial, suggesting that they understood the rule.

In order to test how children updated the attributed belief, we only included participants who successfully attributed false belief to the puppet in the false belief test trial. Overall, children were successful at attributing a false belief to the puppet: upon the puppet's return, 32 out of 44 children predicted that the bear would think that his ball is in the box under the wide pipe and this result was significant (p=.003, binomial). We analyzed the pattern of update based on the 32 children. Children were more likely to revise the rule itself from the puppet's perspective: 18 out of 32 indicated that the puppet thinks that the ball hiding in the box under the narrow pipe is the big ball. Only 8 children indicated the small ball. The remaining 6 children gave an irrelevant answer. These answers can be sorted into different sub-categories. Some children indicated the cube (N=2) others rejected all valid alternatives (e.g., neither the big nor the small ball, N=3). One child came up with a creative solution: "the puppet thinks that it is the small ball because the big ball, which was shown in the beginning, contained a small ball and once it fell, opened up and the small ball rolled into the box under the narrow pipe". Compared to the previous experiments where children had only two options to think what the box might contain, and therefore had to options as an answer to the update question (big ball vs. small ball) here children had three options: the big ball, the small ball or other solutions (e.g., the cube). Therefore, we decided to set the chance level to 33% for the analyses. A binomial test indicated that children were more likely to update the rule from the puppet perspective (p<.001, binomial). Convergingly, the estimated Bayesian Factor favored the alternative hypothesis (BF=34.9). Children's performance is shown in Figure 11.



*Figure 11*: Children's updating strategies elicited by conflicting information in the context of attributed beliefs. The graph represents how many children updated the rule, data, and how many children gave an irrelevant answer to the update question (other)

# 4.6.3. Discussion

In Experiment 4, we tested 5 and 6-year-old children's ability to attribute false beliefs regarding a rule to an agent, and tested belief update from the perspective of the agent. A modification of the scenario used in the previous experiments allowed us to explore this question efficiently. In the test phase of the experiment, false belief was introduced by a location change of the target object: while the character (the bear) observing the events left the scene, and a second character (the frog) entered the scene to take away the big ball from the correct box (under the wide pipe) and then to place a cube into the box under the narrow pipe. Upon the first character's return, children were asked where the bear thinks his ball might be. Most of the children correctly predicted that the bear thinks that his big ball is in the box under the wide pipe, by correctly attributing a false belief. Following the false belief test question, we provided the puppet the incongruent evidence by showing that the box under the wide pipe is empty, instead the box under the narrow pipe is the one that contains an object, as revealed by

sound produced when it was shaken. At this point children were asked what might the character think, which ball is in this box? We specifically used a third type of object, a cube, as a replacement, in order to prevent biased answers.

Most children reacted to the contradictory information with the same update strategy. According to them, "the character thinks that it is the big ball in the box under the narrow pipe". These answers suggest that our participants were more likely to update the character's belief by abandoning the rule. This result was in line with our predictions, that in a scenario where children are required to update an attributed belief, they will be more likely to abandon the rule, even if it was justified by easily detectable physical constrains. This pattern is different from the one we found in Experiment 2 and in Experiment 3. Importantly, children did not expect someone else to arrive to different conclusions when they were facing a conflicting information in Experiment 3.

At the beginning of this chapter, we advanced 3 hypotheses regarding how children might update an attributed belief. People have privileged access to their own beliefs, but not to others' beliefs. Others' beliefs are inferred. Findings from Keysar and colleagues (2003) demonstrate that even adults are biased towards their own knowledge when they are required to reason about other's beliefs; and in some cases, they fail to compute what another person might or might not have seen, and they behave egocentrically during social coordination, relying on others' feedback. These epistemic differences may be responsible for a discrepancy between how participants update their own and an attributed belief.

Our first hypothesis was that simply co-witnessing events with someone will impact children's revision strategies, but this hypothesis has been ruled out in Experiment 3. We further hypothesized, that the as the representation of relational contents might be more difficult, participants will tend to treat attributed beliefs about rules in a more fragile way, which leads them to update it more often compared to the observed data from a third person perspective. The third possibility was that from a third person perspective, participants will be less likely to entrench the observed data, as they might not be as certain about it compared to an observed data from their own perspective, as they don't have direct access to another person's perception.

Our results favor the second possibility: children were more likely to revise the rule in Experiment 4 compared to Experiment 2. In both experiments, the rule to be updated were physically constrained rules, without any possible disabling factors that could suspend them. The only difference was between the two experiments that in Experiment 4, children represented this rule from a third person perspective.

The difference between the revision strategies used in Experiment 2 (first-person beliefs) and Experiment 4 (attributed beliefs) may be informative about the representations involved. First, this finding is congruent with the possibility that attributing rules of this kind poses a greater difficulty for children. Given these difficulties they may be more likely to abandon the rule when facing inconsistency from another person's perspective.

It is also interesting to note, that those children who gave irrelevant answers - despite successfully attributing a false belief to the puppet - were unable to keep track the content of his false belief that finally led to unconventional updates.

# 4.7. General Discussion

Children rapidly learn new information daily, and during this process it is inevitable that new information may contradict with their previously acquired belief sets or else a valid consequence of their beliefs may contradict reality. Children can draw new conclusions integrating what they learned in the past and what they experience in the present revealing facts that were not directly available before. But reasoning may not be impeccable, specially not in children. When inconsistencies emerge, they have to decide what to reject, the stored belief (rules, episodic facts etc. acquired from different sources), the new evidence or the inferential steps that lead to the conclusions that may be now at the origin of the problem. Whereas some previous research testing conflict detection in a deductive framework (Elio & Pelletier, 1994; 1997; Elio, 1995; Politzer & Carles, 2001) showed that adults favor to entrench the observed data, and more likely to abandon the rule when they learn about an information contradicting with their previous beliefs, little is known how children preserve consistency.

The current chapter investigated how 3- and 5-6-year-old children update their own beliefs when they face conflicting information with their previously acquired belief sets. Moreover, we investigated how the update of attributed beliefs might differ from updating beliefs in first person.

The process of inconsistency management may have 3 phases (Johnson-Laird, Girotto and Legrenzi, 2004): the detection of an inconsistency, the revision of beliefs, and the explanation of the inconsistency. We did not plan to directly study the many different ways the inconsistencies can be detected. Instead, we focused on revisions children use. While we did not plan to study children's explanations either, thus we did not elicit actively explanations, occasionally children provided them spontaneously. They served only as some anecdotical evidence for the presence of certain solutions. Future research may investigate children's explanations could gain a better understanding about children's revision strategies when their update shows a random pattern.

In our first set of experiments, we investigated how 3 and 5-year-old children update their beliefs regarding a simple rule acquired in a causally opaque context. The most straightforward description of these rules are with conditionals like: "if the ball is big it always exits on the wide arm" and "if the ball is small, it will exit on the narrow arm". While the motion of the large ball could be explained by the physical properties of the inverted Y pipe, this was not the case for the small ball. When the participants learned to predict the trajectory of the small ball no further justification besides the observed contingencies served as input for the rule. Physically all motion trajectories were equally possible. Our results suggested that 3-year-old children did not follow any obvious strategy, updating their beliefs in a random manner. Instead, 5-year-old children imposed a strategy favored also by adults: they were more likely to abandon the rule in face of a counterevidence and preserve consistency.

Based on these results, we might assume some differences how children represent possibilities across development. As discussed previously, in these experiments, for the small ball, both motion trajectories are equally possible. Thus, being able to represent possibilities is an important factor that can guide children's update strategies. The older age group who reported to have access to modal concepts such as the representation of possibilities had no difficulty to update the rule regarding the small ball's motion trajectories when they faced counterevidence. However, the younger age group showed no specific strategy, and they were equally likely to update the rule and the data as well. Moreover, their performance on the first congruent test trial was lower compared to the older age group's performance. This might be in line with some previous experiments showing that children around the age 3 do not have access to these modal concepts and they are unable to represent possibilities (e.g., Redshaw and Suddendorf, 2016; Suddendorf, Crimston and Redshaw, 2017).

In contrast to Experiment 1, five-year-old children tested in Experiment 2 did not reveal a systematic updating. When the rule is shaped by well-defined, immediately detectable constraints (this was the case of the large ball, that could move only in the large pipe, given its physical properties) this new sample of 5-year-old children hesitated to abandon it in face of a conflicting information. In this case, the rule capturing the motion path (i.e., "if the ball is big, it always exits on the wide arm") was causally more determined (a large solid object cannot fit in a small tube, exactly because its solidity). The stimuli and the instruction were designed in a way, that it did not help invoking disabling factors that could suspend the rule itself (e.g., that the material the ball was made of, in fact is not rigid). This is in line with the findings from Elio (1997) about adults, where participants were less likely to abandon a rule that had no disabling factors, compared to rules with several disabling factors. We received spontaneous explanation from some of the participants indicating that they did search for factors that may allow the suspending the initial justification of the rule they learned during the familiarization. After children faced the conflicting evidence that the box under the narrow pipe that indicates the ball's presence, some of the participants who stated that it must be the small ball in the box under the narrow pipe commented, that "it was the big ball in the beginning, but it shrank when it entered the pipe so now it is small". Based on these comments, we cannot be certain that those children who indicated the small in the unexpected box, truly updated the observed data: they might have simply thought that they saw the big ball in the beginning, and something might have happened that resulted in a smaller ball in the box under the narrow pipe. Possibly, in order to conserve all information they acquired, they complete their beliefs, and add hypothetical attributes to the objects that can turn their understanding of the scene coherent again.

In these experiments the conflict that triggered the revision was between the predicted location of the target object and the object's final location. The experiments assumed correct predictions based on the learned rule and the provided evidence, the correctness of these predictions served also as inclusion criteria for the participants. A conflict can be detected through logical contradiction. If new evidence (i.e., the premise that captures this evidence) is in conflict with the earlier beliefs or the earlier conclusion, there is a contradiction. But logically a contradiction implies any conclusion. Hence, while allows detecting inconsistencies, logic does not force an update.

In Experiment 3 and 4 we tested how children update beliefs from a third person perspective. We hypothesized, that as co-witnessing an event might enhance the encoding of the observed data, children will be more likely to update the rule when they are witnessing the events with another agent, and when they are required to do so from a third person perspective. Moreover, as people have limited access to others' beliefs, there might be certain asymmetries regarding how one update a belief from their own, compared to someone else's perspective. Specifically, if a relational content such as rule are harder to represent and are harder to be attributed, children might treat them as more fragile and will be more likely to update it when they experience conflicting evidence from someone else's perspective.

Results from Experiment 3 showed that simple co-witnessing does not impact children's update strategy. However, as shown in Experiment 4, once children were required to take into account an agent's perspective, as it was different from their own, and they attributed a false belief to the agent, their update strategy changed: they were more likely to abandon the rule and preserve the observed data.

These results from Experiment 4 are in line with our initial hypothesis based on the theoretical possibility from Keysar (2003), suggesting that given the limited access to others' beliefs compared to our owns, we might expect someone else to arrive to a different conclusion as we do. That is, as in Experiment 2 children did not show a specific update strategy when they were required to update a belief from their own perspective, their strategy was different when they were required to do so from a third person perspective in Experiment 4, and they expected the agent to update the rule itself to resolve inconsistency. This might be explained by that attributing relational content is more complex, children are more likely to abandon the rule regarding the relational content, when they update such belief from someone else's perspective.

Future research may also investigate the stability of the update strategies. For instance, when children have direct perceptual evidence about the data, but they learn the rule from a

communicative context (e.g., the rule is communicated by another agent, but is never observed by the participant) or the other way around: the data is communicated by someone else, while they learn about the rule by observing it directly. Some attempts with adults indicate that participants are more likely to abandon a rule that was communicated by an authority than the one that has been self-experienced (Schmeltzer & Markovits, 2005) and this tendency increases when more variability is introduced in the counterevidence (Markovits & Schmeltzer, 2007).

In summary, older, but not younger children tend to revise their beliefs in a similar manner to adults; when there are no physical constrains regarding the conditional rule, they are more likely to abandon the rule itself, and entrench the observed data. Interestingly, this pattern changes when children are required to revise an attributed belief, where they are more likely to update the rule even in the case of physical constrains.

# **Chapter 5. General Discussion**

Humans possess highly sophisticated abilities to reason about others' minds. During our everyday social interactions, we face various scenarios when we need to predict, explain and understand each other's actions, to cooperate and communicate efficiently. However, the inferences we make to understand others' minds can unfold in dynamic and complex situations, when we are required to continuously track each other's beliefs, which may be influenced by various factors. We can assume, that a neurotypical adult would face no difficulty to understand what beliefs others might form based on the available evidence, attributing a wide range of belief contents to others. However, little is known how flexibly children can deal with different scenarios that involve understanding others' minds.

The current thesis aimed to investigate 3-6-year-old children's abilities to represent beliefs about relational contents, and their belief updating strategies regarding relational contents from a first and a third person perspective.

The first empirical chapter (Chapter 2) aimed to answer the question whether 3-5-yearold children are able to attribute beliefs to others about the efficiency of a tool.

As previous experiments targeting early Theory of Mind abilities tested children's ability to attribute beliefs about the location or the identity of objects, the experiments discussed in Chapter 2 implemented a scenario in which children were required to integrate an agent's understanding about the efficiency of the available tools into their belief reasoning.

First, children were familiarized with a puppet whose goal was to obtain his desired objects (colorful balls) from a long tube with the help of two different sticks that varied in length, and thus in their efficiency. With the help of the short stick, the puppet was only able to obtain the balls on the top, whereas with a longer stick, he was able to obtain the remaining balls too, those that remained on the bottom of the tube. In a crucial manipulation, in the absence of the puppet, the experimenter changed the efficiency of the sticks in two different ways, by either breaking the long stick into two and reattaching the ending to the previously inefficient short stick (reality bias version, Experiment 1,3), or simply taking it away (thus leaving two inefficient sticks in two boxes, no reality bias version, Experiment 2, 5 and 6). Our results showed that while 5-year-olds, as expected, tended to correctly predict the puppet's behavior based on his belief in the version with reality bias, 4 and 3-year-old children only succeeded in the no-reality bias version. These results lead to two important conclusions. First, such findings are in line with previous research showing that when processing demands are reduced, even younger children can succeed on explicit false belief tasks (e.g., Rubio-Fernandez & Geurts, 2013,2016). Second, most importantly to our research questions, in contrast to the previously used explicit tasks, our studies involved a scenario that required the integration of additional inferences on top of inferring others' beliefs. In these experiments children also needed to reason about the efficiency of the sticks and integrate these inferences in their belief reasoning to predict someone's action. These findings suggest that children from early on are able to flexibly apply their belief reasoning abilities in various scenarios that go beyond representing beliefs about object locations.

As discussed in the introduction, attributing beliefs about efficiency might build on more complex representations than attributing beliefs about the location of objects. Whereas representing object locations requires children to represent the spatial coordinates of objects, representing the efficiency of objects involves not only representing *where* an object is but also *how* an object is being represented. Furthermore, children need to integrate their causal knowledge underlying tool efficiency in order to predict how an agent will behave based on his beliefs in order to fulfill his desired goal.

The finding that already 3-year-olds are successful in such tasks does not seem to be in line with some of the signature limits proposed by the two-system account. This account predicts that children younger than the age of three, given that they rely on belief-like states and not proper beliefs, would succeed on location change tasks based on registrations, but not on more complex ToM tasks. In contrast to this proposal, our results suggest that children from early on possess abilities that allows them to track others' beliefs in scenarios that go beyond the representation of object locations.

These findings contribute to our understanding of the flexibility of children's belief tracking abilities. Based on these results, we can assume that already by the age of 3, children can integrate information about tool efficiency and another agent's causal knowledge about the efficiency of tools in their false belief reasoning. These research findings seem to be more in line with the proposals arguing for a single Theory of Mind system that develops early and that is able to operate efficiently in various scenarios, similarly to the adult Theory of Mind system (Leslie, 1987; Baillargeon et al. 2010; Kovács, 2016), and becomes more efficient with the development of other cognitive abilities (Carruthers, 2016).

Chapter 3 aimed to answer the question whether young children are able to represent false beliefs based on causal inferences.

Earlier research showed that children are highly skilled about making sense of others' actions when they need to consider information that is available to them through perceptual evidence. However, most of these studies investigating infants' and young children's mental state reasoning abilities tended to focus on simple scenarios, and it is unclear whether children can reason about others' action when a scenario involves not simply direct perceptual evidence, but additional inferences drawn from the perceptual information available to them. The study aimed to explore whether children are able to take into account an agent's belief and the causal inferences she may have performed in order to arrive to these beliefs. Through three experiments, we tested children with a touch screen task where they were required to predict an agent's action who had a false or true belief regarding the causal structure of a ball dispenser

machine. In the first experiment, 3,5 and 5-year-old children's task was to predict where an agent would search her balls from the dispenser's machine, if an obstacle prevents the balls to roll out on one of the two sides of the machine. In the test phase, the obstacle's location was changed, and this change was either witnessed or not by the agent, therefore she either had a true or false belief about the causal structure of the device. Although children performance was relatively high during the training and the true belief trials, they did not manage to predict the character's behavior based on her false belief, suggesting that they were successful at understanding the causal structure of the device which allow them to predict the location of the balls, but they did not take the agent's perspective into account when she had a false belief. Next, we hypothesized that children would be better motivated to track the agent's perspective if they are required to do so from the beginning of the training trials. We tested this hypothesis in Experiment 2, where children participated in a Go No-Go task. In some of the training trials, the same obstacles were used as in the test phase in Experiment 1, and children were required to predict the character's action only in case she was able to see the obstacle from the beginning (as it was not covered by the occluder - Go trials) whereas they were asked not to act when the obstacle was covered by the occluder from the agent's perspective (No-Go trials). Children's overall accuracy on the Go and No-Go trials was relatively high in familiarization, so as their accuracy on predicting the agent's action on the Go and the true belief test trials. However, their performance was significantly below chance in the false belief trials, indicating that they did not take into account the agent's false belief regarding the causal structure of the device. These results indicated that despite the modification of the paradigm, the requirement to track the perspective of the person from the beginning of the task did not help children to take into account the agent's perspective in the false belief test trials. Next, we aimed to verify whether it is our task design (involving lots of repetitions and learning trials) that causes children to fail in the false belief trials Experiment 1 and 2, or their inability to track false beliefs about causal events. Therefore, in Experiment 3, we conducted a location change study, by modifying the design of Experiment 1 and 2. As children above 4 years of age are reported to be able to pass the standard false belief task involving a simple location change scenario, we hypothesized that if children would pass the location change task in Experiment 3, then it would indicate that they may lack the ability to represent false beliefs about causal events, and this causes their failure in Experiment 1 and 2, not our task design. However, if their performance remains at or below chance in Experiment 3 involving a location change scenario, then this would indicate that it is our task design that caused children's poor performance in Experiments 1 and 2, and not their lack of competence. Therefore, in Experiment 3 we modified the experiment into a location change task to test 4-5-year-old children. To our surprise, while our participants' performance was high in the training and true belief trials, yet, their performance was below chance in the false belief trials, in which they simply tracked the location of the ball and predicted the character's goal directed actions while ignoring her perspective. This result shows that its rather the design of the task that impacted children's low performance.

The findings from Chapter 3 do not provide proof for the proposal that children at the age of 3-5 would be able to track false beliefs about causal events. However, one could also argue that the design and the methods that we have used might be responsible for children's poor performance through the three experiments. Future research implementing a simpler visual design and involving less training trials may further investigate at what age children become successful in attributing beliefs about causal events.

The goal of the third study presented in the third empirical chapter (Chapter 4) was so investigate how do children update their own beliefs and beliefs attributed to others.

Children rapidly gain new knowledge every day, and it is likely that sometimes the newly acquired information contradicts their previous beliefs. However, the cognitive system aims at preserving consistency. When facing inconsistent information, one needs to decide what

to keep and what to revise from the previously acquired belief sets. While some theories postulate that rule-like relations that have a greater explanatory power about the world enjoy priority for entrenchment, others state that the observed data are the ones that we can be more certain about, and rules are merely hypothesis about the world, and can be updated easily.

Previous research indicates that adults tend to favor revising the rule-like relations and retain the observed data when facing inconsistent evidence (Elio & Pelletier, 1994; 1997; Elio, 1995; Politzer & Carles, 2001). However, their tendency to revise the rule decreases when the cause-effect relations captured by the rule have only few or no immediately identifiable disabling factors that could suspend them (e.g., laws of physics). While there is ample experimental evidence about adults' belief reasoning strategies in different contexts, little is known whether children update their beliefs in a similar manner. The scope of this chapter was to examine what strategies children use when they update their beliefs in case of conflicting information, and whether they expect others to arrive to similar conclusions. That is, we tested how children update their beliefs regarding different rules, and how they update beliefs from a third person perspective.

First, we tested 3- and 5-year-old children belief updating strategies regarding a simple rule acquired in a causally opaque context (Experiment 1A, B). First, they were asked to predict the motion trajectory of balls with different diameters when they were thrown into a pipe system, which had the shape of an inverted Y, and one of the arms was narrow and the other was wide. Children had to learn that the ball with the smaller diameter will always exit through the narrow arm, and that the ball with the larger diameter could only exit through the wide arm. While the motion trajectory of the big ball was physically determined, this was not the case for the small ball. In the latter case, children simply acquired the rule based on the observed contingencies.

In the test phase children were shown conflicting evidence (that the small ball arrived at location under the wide arm), and we found that 3-year-old children did not follow any obvious strategy, about half of them updated the rule and half the observed data. Instead, 5year-old children followed the strategy reported to be used by adults too: they were more likely to abandon the rule in face of conflicting information and retain the observed data in order to preserve consistency. These findings suggest, that already 5-year-old children favor perceptual evidence, probably because this is what they can be certain about, whereas they treat rule like relations hypothetical and easier to be updated in case of counterevidence. The difference between the two age groups might be explained by the proposal that younger children may lack access to modal concepts and cannot represent different possibilities (Leahy & Carey, 2020), and thus the present case cannot represent that the small ball could in theory exit either via the wide tube or the narrow tube. An important factor supporting rule update in 5-year-olds in this experiment was that participants should understand that in case of the small ball, both trajectories are physically equally possible. However, if younger children cannot form such representations, the lack of rule update as a predominant strategy is not that surprising.

Next, we tested 5-year-old children in a scenario where the rule that had to be updated involved a motion trajectory that was shaped by an immediately detectable physical constrain (Experiment 2). In this experiment children received counterevidence regarding the large ball's final location. The findings of this experiment revealed that in this case 5-year-old children hesitated to abandon the recently learned rule, as it was causally more determined and there were no disabling factors that could suspend the rule itself. These results with 5-year-old children are in line with the results from the studies conducted by Elio (1997); when facing information that is conflicting with their previously acquired beliefs, adults generally prefer to revise the rule itself and retain the observed data, but only in cases where there are factors that would easily disable the rule itself.

The last two experiments from Chapter 4 investigated how children would update beliefs from a third person perspective. When 5-year-old children observed the events together with an agent (Experiment 3), and they had to answer how the character would update his own beliefs regarding causally more constrained rules (about the large ball, as in Experiment 2), the pattern of children's answers did not differ from the previous experiment where they were required to answer to the same questions in first person (no preference for the rule update). Their strategy changed in a false belief scenario, however (Experiment 4). When their perspective differed from the agent's perspective, and when they took into account the agent's false belief about the final location of the ball, most of the children expected the agent to update the rule, when he faced counterevidence. What theories can explain children's revision strategies regarding attributed beliefs? Seemingly, there are certain asymmetries regarding how children update their own, vs. someone else's beliefs to preserve consistency. The attribution of relational contents might involve complex processes compared to the attribution of episodic information. These representations can be fragile, and therefore children (and maybe adults as well) could be more likely to abandon it when they are required to update an attributed belief about rules. Interestingly, as our findings also indicate that when children update a belief from someone else's perspective, what enjoys priority for entrenchment is the observed data. As some previous theories suggest (Elio, 1997) that the reason behind the entrenchment of the observed data can be explained in a way that the observations are one can be most certain about, these results raise the question whether this certainty remains even in a third person perspective. Our study involved learning new information via direct visual experience about the world, but often information is acquired through communication from social agents. Thus, future research could explore whether different update strategies are used in communicative contexts. Studies with adults (e.g., Schmeltzer & Markovits, 2005) revealed that people are more likely to abandon a rule when it was communicated by an authority than when it was based in self

experience. How would children's update strategies be different for instance, when they have direct perceptual evidence about the observed data, but the rule would be communicated by another person? And how would update strategies differ in a case where the data would be communicated by someone else, but they would learn the rules by self-experience? Future work in these directions would significantly advance our knowledge about children's ability to represent and update their beliefs and clarify how they learn about the surroundings in various social and non-social contexts.

# Conclusions

The current thesis aimed to enhance our understanding of children's belief reasoning abilities in the following ways. First, we found that children already by age 3 are able to reason about others' beliefs not only regarding simple scenarios, for instance involving the change of location of objects, but even in situations that require the reasoning about causal relations such as efficiency of tools. Contrary to some of the signature limits proposed by the two-system accounts (Apperly & Butterfil, 2009; Butterfil & Apperly, 2013), these results show that young children may have sophisticated abilities to represent others' beliefs, and these abilities are not limited to track registrations between the agent and the object at a location. Second, when learning about the world, 5-year-old children seem to flexibly revise and reorganize their beliefs similarly to adults when they face contradictions.

These findings enhance our knowledge about the flexibility and possible limits of children's belief reasoning abilities, supporting theories and previous findings about children's early developing Theory of Mind system.

# References

- Albiach-Serrano, A., Sebastián-Enesco, C., Seed, A., Colmenares, F., & Call, J. (2015). Comparing humans and nonhuman great apes in the broken cloth problem: Is their knowledge causal or perceptual? *Journal of experimental child psychology*, 139, 174-189.
- Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belieflike states? *Psychological review*, 116(4), 953.
- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in cognitive sciences*, 14(3), 110-118.
- Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. Annual review of psychology, 67, 159-186.
- Bauer, P. J., Schwade, J. A., Wewerka, S. S., & Delaney, K. (1999). Planning ahead: Goaldirected problem solving by 2-year-olds. *Developmental psychology*, 35(5), 1321.
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? Cognition, 21(1), 37-46.
- Bennett, J. (1978). Some remarks about concepts. Behavioral and Brain Sciences, 1(4), 557-560.
- Birch, S. A., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental-state attribution. *Psychological Science*, *14*(3), 283-286
- Bonawitz, E., Fischer, A., & Schulz, L. (2012). Teaching 3.5-year-olds to revise their beliefs given ambiguous evidence. *Journal of Cognition and Development*, *13*(2), 266-280.
- Bonawitz, E. B., Ferranti, D., Saxe, R., Gopnik, A., Meltzoff, A. N., Woodward, J., & Schulz, L. E. (2010). Just do it? Investigating the gap between prediction and action in toddlers' causal inferences. *Cognition*, 115(1), 104-117.
- Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition*, 104(2), 287-314.
- Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*(2), 337-342.
- Buttelmann, D., Over, H., Carpenter, M., & Tomasello, M. (2014). Eighteen-month-olds understand false beliefs in an unexpected-contents task. *Journal of experimental child psychology*, 119, 120-126.

- Buttelmann, F., Suhrke, J., & Buttelmann, D. (2015). What you get is what you believe: Eighteen-month-olds demonstrate belief understanding in an unexpected-identity task. *Journal of experimental child psychology*, *131*, 94-103.
- Buttelmann, F., & Kovács, A. M. (2019). 14-Month-olds anticipate others' actions based on their belief about an object's identity. *Infancy*, *24*(5), 738-751.
- Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, 28(5), 606-637.
- Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child development*, 69(3), 672-691.
- Carpenter, M., Call, J., & Tomasello, M. (2002). A new false belief test for 36-montholds. *British Journal of Developmental Psychology*, 20(3), 393-420.
- Carruthers, P. (2013). Mindreading in infancy. Mind & Language, 28(2), 141-172.
- Carruthers, P. (2016). Two systems for mindreading? *Review of Philosophy and Psychology*, 7(1), 141-162.
- Choi, Y.J., & Luo, Y. (2015). 13-month-olds' understanding of social interactions. *Psychological science*, *26*(3), 274-283.
- Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, 107(2), 705-717.
- Cesana-Arlotti, N., Martín, A., Téglás, E., Vorobyova, L., Cetnarski, R., & Bonatti, L. L. (2018). Precursors of logical reasoning in preverbal human infants. *Science*, 359(6381), 1263-1266.
- Cesana-Arlotti, N., Kovács, Á. M., & Téglás, E. (2020). Infants recruit logic to learn about the social world. *Nature communications*, *11*(1), 1-9.
- Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & cognition*, 19(3), 274-282.
- Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition*, 23(5), 646-658.
- Dennett, D. C. (1978). Beliefs about beliefs [P&W, SR&B].
- Evans, T. A., & Westergaard, G. C. (2004). Discrimination of functionally appropriate and inappropriate throwing tools by captive tufted capuchins (Cebus apella). *Animal Cognition*, 7(4), 255-262.
- Elio, R., & Pelletier, F. J. (1994, November). On relevance in non-monotonic reasoning: Some empirical studies. In *Relevance: American Association for Artificial Intelligence 1994 Fall Symposium Series* (pp. 64-67).
- Elio, R. (1997). What to believe when inferences are contradicted: The impact of knowledge type and inference rule. In *Proceedings of the nineteenth annual conference of the cognitive science society* (pp. 211-216).
- Elio, R., & Pelletier, F. J. (1997). Belief change as propositional update. *Cognitive Science*, *21*(4), 419-460.
- Elsabbagh, M., Fernandes, J., Webb, S. J., Dawson, G., Charman, T., Johnson, M. H., & British Autism Study of Infant Siblings Team. (2013). Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood. *Biological psychiatry*, 74(3), 189-194.
- Fagard, J., Rat-Fischer, L., & O'Regan, J. K. (2014). The emergence of use of a rake-like tool: a longitudinal study in human infants. *Frontiers in psychology*, 5.
- Fizke, E., Butterfill, S., van de Loo, L., Reindl, E., & Rakoczy, H. (2017). Are there signature limits in early theory of mind? *Journal of Experimental Child Psychology*, *162*, 209-224.
- Forgács, B., Gervain, J., Parise, E., Csibra, G., Gergely, G., Baross, J., & Király, I. (2020).
  Electrophysiological investigation of infants' understanding of understanding. *Developmental Cognitive Neuroscience*, 43, 100783.
- Ganea, P. A., Shutts, K., Spelke, E. S., & DeLoache, J. S. (2007). Thinking of things unseen: Infants' use of language to update mental representations. *Psychological science*, 18(8), 734-739.
- Ganea, P. A., & Harris, P. L. (2010). Not doing what you are told: Early perseverative errors in updating mental representations via language. *Child development*, *81*(2), 457-463.
- Ghrear, S., Fung, K., Haddock, T., & Birch, S. A. (2021). Only familiar information is a "curse": Children's ability to predict what their peers know. *Child development*, *92*(1), 54-75.

- Gregory, S. E., & Jackson, M. C. (2021). Increased perceptual distraction and task demand enhances gaze and non-biological cuing effects. *Quarterly Journal of Experimental Psychology*, 74(2), 221-240.
- Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive science*, *25*(4), 565-610.
- Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain sciences*, *16*(1), 1-14.
- Harman, G. (1978). Studying the chimpanzee's theory of mind. *Behavioral and Brain Sciences*, 1(4), 576-577.
- He, Z., Bolz, M., & Baillargeon, R. (2011). 2.5-year-olds succeed at a verbal anticipatory-looking false-belief task. *British Journal of Developmental Psychology*, *30*(1), 14-29.
- Howard, L. H., & Woodward, A. L. (2019). Human actions support infant memory. Journal of Cognition and Development, 20(5), 772-789.
- Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review*, 111(3), 640.
- Kampis, D., Karman, P., Csibra, G., Southgate, V., & Hernik, M. (2021). A two-lab direct replication attempt of Southgate, Senju and Csibra (2007). *Royal Society open science*, 8(8), 210190.
- Kampis, D., Parise, E., Csibra, G., & Kovács, Á. M. (2015). Neural signatures for sustaining object representations attributed to others in preverbal human infants. *Proceedings of the Royal Society B: Biological Sciences*, 282(1819), 20151683.
- Kampis, D., & Kovács, Á. M. (2021). Seeing the world from others' perspective: 14-month-olds show altercentric modulation effects by others' beliefs. *Open Mind*, 1-19.
- Katsuno, H., & Mendelzon, A. O. (1991). Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52(3), 263-294.
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89(1), 25-41.
- Király, I., Jovanovic, B., Prinz, W., Aschersleben, G., & Gergely, G. (2003). The early origins of goal attribution in infancy. *Consciousness and cognition*, 12(4), 752-769.

- Kloo, D., & Perner, J. (2008). Training theory of mind and executive control: A tool for improving school achievement? *Mind, Brain, and Education, 2*(3), 122-127.
- Knudsen, B., & Liszkowski, U. (2012a). 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy*, 17(6), 672-691.
- Knudsen, B., & Liszkowski, U. (2012b). Eighteen-and 24-month-old infants correct others in anticipation of action mistakes. *Developmental science*, *15*(1), 113-122.
- Kotovsky, L., & Baillargeon, R. (1994). Calibration-based reasoning about collision events in 11-month-old infants. *Cognition*, *51*(2), 107-129.
- Kovács, A. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330(6012), 1830-1834.
- Kovács, Á. M. (2015). Belief files in theory of mind reasoning. *Review of Philosophy and Psychology*, 7(2), 509-527.
- Kovács, Á. M., Téglás, E., & Csibra, G. (2021). Can infants adopt underspecified contents into attributed beliefs? Representational prerequisites of theory of mind. *Cognition*, 213, 104640.
- Kammermeier, M., & Paulus, M. (2018). Do action-based tasks evidence false-belief understanding in young children? *Cognitive Development*, 46, 31-39.
- Kulke, L., & Rakoczy, H. (2018). Implicit Theory of Mind–An overview of current replications and non-replications. *Data in brief*, *16*, 101-104.
- Kulke, L., von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. *Psychological science*, 29(6), 888-900.
- Kushnir, T., & Gopnik, A. (2007). Conditional probability versus spatial contiguity in causal learning: Preschoolers use new contingency evidence to overcome prior spatial assumptions. *Developmental psychology*, 43(1), 186.
- Leahy, B. P., & Carey, S. E. (2020). The acquisition of modal concepts. *Trends in Cognitive Sciences*, 24(1), 65-78.
- Leslie, A. M. (1984). Spatiotemporal continuity and the perception of causality in infants. *Perception*, 13(3), 287-305.

- Leslie, A. M. (1987). Pretense and representation: The origins of" theory of mind.". *Psychological review*, 94(4), 412.
- Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, 25(3), 265-288.
- Leslie, A. M., Friedman, O., & German, T. P. (2004). 'Theory of mind' as a mechanism of selective attention. *Trends in Cognitive Sciences*, 12(8), 528-533.
- Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in cognitive sciences*, 8(12), 528-533.
- Leslie, A. M., German, T. P., & Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive psychology*, 50(1), 45-85.
- Limongelli, L., Boysen, S. T., & Visalberghi, E. (1995). Comprehension of cause effect relations in a tool-using task by chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, *109*(1), 18.
- Luchkina, E., Sommerville, J. A., & Sobel, D. M. (2018). More than just making it go: Toddlers effectively integrate causal efficacy and intentionality in selecting an appropriate causal intervention. Cognitive Development, 45, 48-56.
- Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science*, 16(8), 601-608.
- Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, *105*(3), 489-512.
- Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science*, 24(3), 305-311.
- Low, J., Drummond, W., Walmsley, A., & Wang, B. (2014). Representing how rabbits quack and competitors act: Limits on preschoolers' efficient ability to track perspective. *Child development*, 85(4), 1519-1534.
- Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives*, 10(3), 184-189.

- Macris, D. M., & Sobel, D. M. (2017). The role of evidence diversity and explanation in 4-and 5-year-olds' resolution of counterevidence. *Journal of Cognition and Development*, 18(3), 358-374.
- Markovits, H., & Potvin, F. (2001). Suppression of valid inferences and knowledge structures: The curious effect of producing alternative antecedents on reasoning with causal conditionals. *Memory & Cognition*, 29(5), 736-744.
- Markovits, H., & Schmeltzer, C. (2007). What makes people revise their beliefs following contradictory anecdotal evidence? The role of systemic variability and direct experience. *Cognitive Science*, *31*(3), 535-547.
- Mascalzoni, E., Regolin, L., Vallortigara, G., & Simion, F. (2013). The cradle of causal reasoning: Newborns' preference for physical causality. *Developmental science*, 16(3), 327-335.
- Meltzoff, A. N., Waismeyer, A., & Gopnik, A. (2012). Learning about causes from people: observational causal learning in 24-month-old infants. *Developmental psychology*, 48(5), 1215.
- Miscione, J. L., Marvin, R. S., O'Brien, R. G., & Greenberg, M. T. (1978). A developmental study of preschool children's understanding of the words "know" and "guess". *Child Development*, 1107-1113.
- Mitchell, P., & Lacohée, H. (1991). Children's early understanding of false belief. *Cognition*, 39(2), 107-127.
- Mody, S., & Carey, S. (2016). The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition*, *154*, 40-48.
- Moll, H., Carpenter, M., & Tomasello, M. (2007). Fourteen-month-olds know what others experience only in joint engagement. *Developmental Science*, *10*(6), 826-835.
- Moll, H., Khalulyan, A., & Moffett, L. (2017). 2.5-year-olds express suspense when others approach reality with false expectations. *Child Development*, 88(1), 114-122.
- Muentener, P., & Bonawitz, E. The Development of Causal Reasoning. *The Oxford Handbook* of Causal Reasoning.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *science*, 308(5719), 255-258.

- Oktay-Gür, N., Schulz, A., & Rakoczy, H. (2018). Children exhibit different performance patterns in explicit and implicit theory of mind tasks. *Cognition*, *173*, 60-74.
- Perner, J. (2010). Who took the cog out of cognitive science? Mentalism in an era of anticognitivism. In *Cognition and neuropsychology* (pp. 255-276). Psychology Press.
- Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British journal of developmental psychology*, 5(2), 125-137.
- Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, 308(5719), 214-216.
- Perner, J., & Roessler, J. (2012). From infants to children's appreciation of belief. *Trends in cognitive sciences*, 16(10), 519-525.
- Politzer, G., & Carles, L. (2001). Belief revision and uncertain reasoning. *Thinking & Reasoning*, 7(3), 217-234.
- Poulin-Dubois, D., Rakoczy, H., Burnside, K., Crivello, C., Dörrenberg, S., Edwards, K., ... & Ruffman, T. (2018). Do infants understand false beliefs? We don't know yet–A commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cognitive Development*, 48, 302-315.
- Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child development*, *61*(4), 973-982.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral* and brain sciences, 1(4), 515-526.
- Recanati, F. (2012). Mental files. Oxford University Press.
- Redshaw, J., & Suddendorf, T. (2016). Children's and apes' preparatory responses to two mutually exclusive possibilities. *Current Biology*, *26*(13), 1758-1762.
- Revlin, R., Calvillo, D. P., & Ballard, S. (2005). Counterfactual reasoning: Resolving inconsistency before your eyes. *Psychologica Belgica*, 45(1), 47-56.
- Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological science*, *24*(1), 27-33.
- Rubio-Fernández, P., & Geurts, B. (2016). Don't mention the marble! The role of attentional processes in false-belief tasks. *Review of Philosophy and Psychology*, 7(4), 835-850.

- Rubio-Fernández, P. (2017). Can we forget what we know in a false-belief task? An investigation of the true-belief default. *Cognitive Science*, *41*(1), 218-241.
- Rubio-Fernández, P., Jara-Ettinger, J., & Gibson, E. (2017). Can processing demands explain toddlers' performance in false-belief tasks? *Proceedings of the National Academy of Sciences*, 114(19), E3750-E3750.
- Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental review*, 34(3), 265-293.
- Sally, D., & Hill, E. (2006). The development of interpersonal strategy: Autism, theory-of-mind, cooperation and fairness. *Journal of economic psychology*, *27*(1), 73-97.
- Salter, G., & Breheny, R. (2019). Removing shared information improves 3-and 4-year-olds' performance on a change-of-location explicit false belief task. *Journal of Experimental Child Psychology*, 187, 104665.
- Schmeltzer, C., & Markovits, H. (2005). Belief revision, self-construction and systemic certainty. *Current psychology letters. Behaviour, brain & cognition*, (17, Vol. 3, 2005).
- Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in cognitive sciences*, 4(8), 299-309.
- Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental* psychology, 40(2), 162.
- Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: preschoolers engage in more exploratory play when evidence is confounded. *Developmental psychology*, 43(4), 1045.
- Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental science*, 10(3), 322-332.
- Schulz, L. E., Bonawitz, E. B., & Griffiths, T. L. (2007). Can being scared cause tummy aches? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental psychology*, 43(5), 1124
- Shultz, T. (1982). Rules of causal attribution. Monographs of the Society for Research in Child Development, 47, 1–51.

- Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child development*, *80*(4), 1172-1196.
- Scott, R. M., He, Z., Baillargeon, R., & Cummins, D. (2012). False-belief understanding in 2.5year-olds: Evidence from two novel verbal spontaneous-response tasks. *Developmental science*, 15(2), 181-193.
- Scott, R. M., Richman, J. C., & Baillargeon, R. (2015). Infants understand deceptive intentions to implant false beliefs about identity: New evidence for early mentalistic reasoning. *Cognitive psychology*, 82, 32-56.
- Scott, R. M., & Roby, E. (2015). Processing demands impact 3-year-olds' performance in a spontaneous-response task: New evidence for the processing-load account of early falsebelief understanding. *PloS one*, *10*(11), e0142405.
- Scott, R. M. (2017). Surprise! 20-month-old infants understand the emotional consequences of false beliefs. *Cognition*, *159*, 33-47.
- Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*.
- Scott, R. M., Roby, E., & Setoh, P. (2020). 2.5-year-olds succeed in identity and location elicitedresponse false-belief tasks with adequate response practice. *Journal of Experimental Child Psychology*, 198, 104890.
- Scott, R. M., Roby, E., & Baillargeon, R. (2022). How Sophisticated Is Infants Theory of Mind?. The cambridge handbook of cognitive development, 242.
- Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences*, 113(47), 13360-13365.
- Sloman, S. A., & Lagnado, D. (2015). Causality in thought. *Annual review of psychology*, 66, 223-247.
- Smith-Flores, A. S., & Feigenson, L. (2021). Preschoolers represent others' false beliefs about emotions. *Cognitive Development*, 59, 101081.
- Song, H. J., Onishi, K. H., Baillargeon, R., & Fisher, C. (2008). Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition*, 109(3), 295-315.

- Song, H. J., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental psychology*, 44(6), 1789.
- Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological science*, *18*(7), 587-592.
- Southgate, V., & Begus, K. (2013). Motor activation during the prediction of nonexecutable actions in infants. *Psychological Science*, *24*(6), 828-835.
- Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*, *130*(1), 1-10.
- Suddendorf T, Crimston J, Redshaw J. 2017 Preparatory responses to socially determined, mutually exclusive possibilities in chimpanzees and children. *Biol. Lett.* 13, 20170170. (doi:10.1098/rsbl.2017.0170)
- Surian, L., & Leslie, A. M. (1999). Competence and performance in false belief understanding: A comparison of autistic and normal 3-year-old children. *British Journal of Developmental Psychology*, 17(1), 141-155.
- Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology*, 30(1), 30-44.
- Thibault, S., Py, R., Gervasi, A. M., Salemme, R., Koun, E., Lövden, M., & Brozzoli, C. (2021). Tool use and language share syntactic processes and neural patterns in the basal ganglia. *Science*, 374(6569), eabe0874.
- Ting, F., He, Z., & Baillargeon, R. (2021). Five-month-old infants attribute inferences based on general knowledge to agents. *Journal of Experimental Child Psychology*, 208, 105126.
- Van Hoeck, N., Dieussaert, K., & Revlin, R. (2008). Children's Counterfactual Reasoning Strategy in Belief Contravening Problems. In 30th Annual Conference of the Cognitive Science Society (pp. 1574-1579).
- Van Hoeck, N., Revlin, R., Dieussaert, K., & Schaeken, W. (2012). The development of counterfactual reasoning in belief revision. *Psychologica Belgica*, 52(4), 407.
- Waldmann, M. R., & Hagmayer, Y. Causal Reasoning. In *The Oxford Handbook of Cognitive Psychology*.

- 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.
- 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.
- Wimmer, H., & Weichbold, V. (1994). Children's theory of mind: Fodor's heuristics examined. *Cognition*, 53(1), 45-57.
- Zelazo, P. D., & Müller, U. (2002). Executive Function in Typical and Atypical Development. *Blackwell Handbook of Childhood Cognitive Development*, 445-469.