MINIMAL MINDREADERS LET SPINNING DOGS LIE:

NEW EVIDENCE FOR THE DUAL-PROCESS ACCOUNT OF HUMAN MINDREADING

By

KATHERYN EDWARDS

A thesis

submitted to

Victoria University of Wellington

in fulfilment of the requirements for the degree of

Doctor of Philosophy

Victoria University of Wellington

2019

Abstract

Five experiments investigated evidence for a dual-process account of mindreading (Apperly, 2010). This account is motivated by two puzzles: First, why is it that three-year-olds fail standard false-belief tests when looking patterns infer that infants are sensitive to others' false beliefs? Secondly, why is adult mindreading sometimes slow and effortful, and at other times fast and effortless? The seemingly contradictory observations may be explained by drawing upon two relatively distinct mindreading abilities: 'Efficient' processing supports precocious infant performances in non-verbal tasks and fast-paced social interaction in adults, while the later developing 'flexible' processing permits full blown understanding of beliefs and facilitates correct verbal responding in standard false-belief tests. Evidence for this theory can be sought by exploiting the idea that there are 'signature limits' to the type of information that can be efficiently processed.

One conjecture is that representations underpinning efficient belief-tracking relate agents to objects, leading to the prediction that efficient processing cannot handle false-beliefs involving identity. Experiments 1 and 2 used a novel action-prediction paradigm to determine if adults' reaction-time patterns differed between two false-belief tasks, one involving a standard change-of-location scenario, and one which also incorporated an identity component. The findings revealed equivalent flexible processing across both tasks. However, there were distinct reaction-time profiles between the tasks such that efficient belief-tracking was only observed in the change-of-location task. The absence of efficient processing in the task incorporating an identity component supports the conjecture that efficient belief-tracking is limited to relational, rather than propositional attitudes.

A second conjecture is that representations underpinning efficient belief-tracking either do not specify agents' locations or do not specify objects' orientations. This leads to the prediction that efficient belief-tracking alone will not yield expectations about agents' perspectives. In a novel object-detection paradigm, Experiments 3 to 5 tested the extent to which adults efficiently tracked the belief of a passive bystander in two closely-matched but conceptually distinct tasks. In a task involving homogenous objects, reaction times were involuntarily influenced by the presence of the bystander. By contrast, in a second task in which the object could be differently perceived depending on where the agent was located in relation to that object, the presence of the agent did not influence adults' response times, supporting the second conjecture.

Acknowledgements

I would like to express my gratitude to all those who gave their unrelenting support and assistance during my candidature. I am deeply indebted to my supervisor, Associate Professor Jason Low. Without his guidance and unwavering support this thesis would not have been possible. It has been a privilege to work under his tutelage, and I am immensely grateful for unfettered access to his considerable knowledge, experience and wisdom. I would also like to thank Dr. Stephen Butterfill for his dynamic theorizing and insightful comments which both challenged and consolidated my thinking.

I wish to thank my lab colleagues, past and present - Chris Maymon, Radhika Patel-Cornish, Claudie Peloquin, Schyana Sivanantham, Cong Fan, Pieter Six and Gio Zani for their friendship and understanding over the years. And I am very grateful to Dr.'s Anne Macaskill, Maree Hunt, and Reneeta Morgan for their advice and guidance in times of need. Of course, my experiments would not have left the ground without the technical help of Sam Mohd Akhir, Al Abenoja and the late Doug Drysdale.

Finally, I turn to my family. Without their unconditional love I would not have had the time, strength or confidence to see this through. I owe so much to my father, Dick Edwards, who died just weeks before my candidature began; I miss you Dad, but thanks to my theory of mind we get to talk every single day. I'm deeply grateful to my wonderful mother, Nesta Edwards, who always encouraged me to push myself and face the challenges I'd sooner run away from, and to my bigger and better sister, Sian Ledbury, whose hard work and generosity has always been an inspiration. I owe much to David and Carole Edwards for the restorative meals and bottles of red wine. But the biggest thank you's go to those who had to put up with the mood-swings, doom-laden monologues and panic attacks on a daily basis: Thank you, Holly and Elijah, for always believing in me and for smoothing out the peaks and troughs with your wit and wisdom; and thank you Adam, for letting me escape into your alternative universe, and for giving me another perspective (yes, sometimes Lego IS all that matters). Last but certainly not least there's Paul: my punch bag, my muse, my life-support system. Always there with a cwtch and kind word. You above all, had the hardest task, and my love and gratitude know no bounds. Here's to our next adventure together.

Table of Contents

Table of Figures

List of Tables

Structure of Thesis

This thesis is comprised of an introductory chapter, which presents conflicting evidence regarding the emergence of mindreading, and outlines several theories which attempt to elucidate this developmental puzzle (Chapter 1). The following two chapters (Chapters 2 & 3) detail two experiments that use a novel action-prediction paradigm to tease apart competing theories. Chapter 4 introduces another puzzle, wherein adults' mindreading is sometimes automatic, and other times deliberate. The three subsequent chapters (Chapters 4, 6 & 7) describe three experiments which employ a novel object-detection paradigm to determine the boundaries of automatic mindreading. Finally, Chapter 8 brings the two sets of findings together to shed light on the nature and development of human mindreading.

This thesis contains sections from published articles and a book chapter (in press) for which I was first author. It also expands upon these publications by making reference to more recently published research. Chapter 1 contains sections from the following:

Edwards, K., & Low, J. (2017). Reaction time profiles of adults' action prediction reveal two mindreading systems. *Cognition, 160*, 1-16. doi.org/10.1016/j.cognition.2016.12.004

Edwards et al. (in press). False belief understanding: On cognitive development, cognitive competence & cognitive systems. *The Wiley-Blackwell Encyclopedia of Child and Adolescent Development.*

Chapters 2 and 3 describe two experiments from **Edwards & Low (2017)**.

Chapters 4 contains sections from the published article:

Edwards, K., & Low, J. (2019). Leve1 2 perspective-taking distinguishes automatic and non-automatic belief-tracking. *Cognition, 193*. doi.org/10.1016/j.cognition.2019.104017

Chapters 5, 6 and 7 describe three experiments from **Edwards & Low (2019)**.

Statement of Authorship and Copyright

I am the primary (lead) author on the co-authored experimental articles presented in this thesis. I collected and managed the data, designed the research question, conducted and interpreted the analyses and wrote the first drafts. Jason Low was involved in discussions about the research question and provided critical feedback and revisions.

With regard to copyrighted published material I confirm that the publisher of *Cognition*, Elsevier, allows for the two articles (Edwards & Low, 2017; Edwards & Low, 2019) to be included in the author's thesis. I also confirm that Wiley has granted me permission to reproduce content from a review article to be published in The Wiley-Blackwell Encyclopedia of Child and Adolescent Development, of which I am the primary author.

CHAPTER 1. *General Introduction*

This chapter contains content written by Katheryn Edwards from the following published article and book chapter (in press):

Edwards, K., & Low, J. (2017). Reaction time profiles of adults' action prediction reveal two mindreading systems. *Cognition, 160*, 1-16. doi.org/10.1016/j.cognition.2016.12.004

Edwards, K. et al. (in press). False belief understanding: On cognitive development, cognitive competence & cognitive systems. *The Wiley-Blackwell Encyclopedia of Child and Adolescent Development.*

1.1. *Introduction*

The capacity to ascribe beliefs to others allows humans to predict and make sense of others' actions. While researchers agree on the importance of this mindreading (or theory of mind) ability in everyday life (see Sabbagh & Bowman for a recent review, 2018), there is much debate over its nature and development. Questions remain as to whether young children's mindreading competencies have been underestimated and whether there might be different kinds of cognitive processes and representations that guide humans' ability to track and ascribe others' beliefs. The current chapter presents an overview of research investigating the emergence of mindreading in early childhood. From this review a hotly debated developmental puzzle emerges: Why do infants pass non-verbal belief-reasoning tasks and then go on to fail verbal false-belief tasks at age 3 years? The chapter introduces three main theories seeking to explain the conflicting findings among infants and young children, and provides a rationale for testing between different mindreading accounts using a novel action-anticipation task with an *adult* sample.

1.2. *False-belief reasoning*

Modern mindreading literature emerged from the dual efforts of philosophers and psychologists keen to address how humans come to understand their own, and others', mental states. In 1978, a seminal paper by Premack and Woodruff - "Does the Chimpanzee Have a Theory of Mind?" - coined the now familiar phrase and triggered a new wave of mindreading research. The authors claimed that chimpanzees' capacity to predict the behaviour of a human actor suggested that they were able to theorize about the invisible contents of others' minds. However, the philosopher Dennett (1978) proposed that for an animal to exhibit a theory of mind it must demonstrate some understanding that other minds can see, know, want, or believe something different from its own. For developmental psychologists such reasoning prompted novel methodologies designed to measure children's predictions about the behaviour of agents with inaccurate representations of reality. These false-belief tasks came to be seen as the litmus test of children's ability to appreciate others' mental states, with successful performances reflecting a conceptual shift in mindreading cognition.

1.2.1. Conceptual change

In a pivotal false-belief paradigm referred to as an unexpected-transfer (or change-of-

location) task, Wimmer and Perner (1983) presented children a puppet show in which Maxi stored his chocolate in a cupboard and then left the scene. During Maxi's absence, the chocolate was moved into a drawer by his mother. Children were then asked where Maxi would look for the chocolate on his return. In studies using this paradigm, children older than 4 years typically predict that Maxi will erroneously look in the cupboard, suggesting that they appreciate the representational nature of belief - that is, how Maxi's misrepresentation of reality would manifest in his behaviour. Children younger than 4 years typically predict that Maxi will look in the drawer where the chocolate really is, demonstrating an inability to attribute false beliefs to others. Younger age children also fail unexpected-contents false-belief tasks (e.g., Perner, Leekam, & Wimmer, 1987). Here, children are typically asked what they think is inside a container that appears to hold a particular kind of object (e.g., a crackers box). After the true contents are revealed to them (e.g., marbles), they are asked what a naive agent would think was in the box. Younger children incorrectly predict that the agent would think there were marbles in the box, whereas most 4- to 5-year-olds appreciate that the newcomer would hold a false belief that there were crackers inside. Decades of research on standard false-belief tasks requiring direct verbal reasoning, indicate that a full-blown theory-of-mind emerges in humans from about 4 years of age. The age effect is robust for these verbal – or 'explicit' - tasks; it does not matter whether the task measures someone's false belief about an object's location, content, or properties, or whether the task measures attribution of someone else's false belief or the child's own previous false belief (Wellman, Cross, & Watson, 2001).

Conceptually sophisticated attributions of belief also involve handling *aspectuality* as well as misrepresentation. The term aspectuality refers to the way beliefs always represent referents only under some description and not under others. For example, if Maxi believes that Mr. Hyde moved his chocolate, it does not follow that Maxi also believes Dr. Jekyll moved his chocolate, because Maxi may think that the two descriptions "Mr. Hyde" and "Dr. Jekyll" represent numerically distinct individuals. Studies show that children's understanding of belief is conceptually unified. When 4-year-olds start to pass tasks requiring an ascription of someone's false belief about an object's location, they also start to pass tasks requiring an appreciation of how someone's false belief about an object's identity can lead the person to think that there are more (or fewer) objects or individuals in the world than there really are (Oktay-Gür, Schulz, & Rakoczy, 2018; Rakoczy, Bergfeld, Schwarz, & Fizke, 2015).

1.2.2. Socio-cultural variation

It is important to point out that, against this backdrop of a common findings, there are sociocultural variations in theory of mind development (Wellman, 2014). For example, while some cross-cultural research provides evidence for a universal onset of false-belief understanding (e.g., Avis & Harris, 1991; Oberle, 2009), there is other research showing onset delays. One study reported no difference in the performances of 3- and 5-year old Samoan children in a change-of-location false-belief task – and only a 68% pass rate in the 12 to 14 years age group (Mayer & Träuble, 2013). More recently, another study showed that children from Vanuatu failed to show false-belief understanding (according to standard testing) until 7 to 9 years of age (Dixson, Komugabe-Dixson, Dixson, & Low, 2018). These findings emphasize the effect of environmental factors on the onset of false-belief understanding; for example, delays may arise in cultures or social groups where the discussion of private mental states is rare, or where the sharing of practical know-how is valued above the sharing of beliefs.

Variations are also revealed in young children's progressive understanding of the constructs underlying full blown theory of mind. To elucidate, Wellman and Liu's (2004) Theory of Mind Scale captured the discovery that, in the advancement to false-belief-reasoning (FB), children first appreciate that people: may have diverse desires (DD); may have diverse beliefs (DB); and may not know that something is true (knowledge access; KA). Preschoolers in the U.S., Canada, Australia, and Germany progress through these stages in a reliable order of difficulty: DD→DB→KA→FB (an understanding of hidden emotion follows FB, but is not discussed here) (Kristen, Thoermer, Hofer, Aschersleben, & Sodian, 2006; Peterson, Wellman, & Liu, 2005; Wellman & Liu, 2004). However, other research has shown that Chinese, Iranian and Turkish preschoolers develop an understanding of KA before DB (DD→KA →DB →FB) (Selcuk, Brink, Ekerim, & Wellman, 2018; Shahaeian, Peterson, Slaughter, & Wellman, 2011; Wellman, Fang, Liu, Zhu, & Liu, 2006; Wellman, Fang, & Peterson, 2011). A compelling explanation for this is that China, Iran and Turkey have collectivist orientations which tend to value access to knowledge about shared norms, and have a lower tolerance for independent belief learning (Wellman, 2018).

To summarise, children's learning about the workings of the human mind undergoes protracted development. There are multiple components that support children's successful performance on false-belief tasks, including advancements in understanding representations,

counterfactual reasoning skills, executive function skills, richness of vocabulary, syntactic skills, and exposure to complex social and conversational interactions that introduce mental states on an interpersonal level. The classical view is that advances in language, executive function and participation in complex social interactions help children learn about subjective mental representations (Low & Perner, 2012; Perner, 1991; Wellman, 2014; Wellman, Cross, & Watson, 2001). The emergence of theory of mind abilities should be assessed against a broad socio-cultural framework, so that when we see the linear developmental trajectory in representational understanding of belief from below to above chance (Wellman et al., 2001) we do so with an understanding that theory of mind development is a fitting example of "universalism without uniformity" (p.507, Shweder & Sullivan, 1993).

1.2.3. Early cognitive competence

While the general consensus is that children's explicit false-belief understanding emerges around 4 years of age, research measuring certain nonverbal (also referred to as implicit) responses suggests that children *younger* than 4 years may show implicit sensitivity to others' false beliefs. Clements and Perner (1994) discovered that 3-year-olds correctly looked in anticipation of a character searching in the false-belief-based location despite giving incorrect verbal predictions. Young children (3- to 4 years-of-age) in small-scale societies also show correct gaze anticipations whilst giving incorrect verbal predictions in standard change-of-location false-belief tasks (e.g., Wang, Hadi, & Low, 2015). These looking responses seem to be unconscious: in a replication of Clements and Perner's study, Ruffman, Garnham, Import, and Connolly (2001) showed that the presence of precocious visual orienting behaviours does not shake children's confidence in their incorrect verbal judgments. To measure confidence, they asked children (3 to 5 years-of-age) to bet highly-valued counters on where a character would go to find an object. The authors found that 94% of the younger children (mean age 3.40 years) who displayed accurate anticipatory looking, were certain of their incorrect verbal predictions despite wagering all of their counters on the outcome. Ruffman et al.'s findings challenge Zelazo, Frye, and Rapus (1996) who claim that looking behaviour in false-belief tasks might reveal conscious awareness. It is challenging to make firm conclusions about the nature of the knowledge guiding young children's visual orienting, and Clements and Perner caution that implicit understanding observed in terms of pure action is unlikely to be based on an abstract understanding of belief. However, recent studies have widened the gap between the age when children demonstrate false-belief understanding on tasks that require direct judgments and the age when children demonstrate

false-belief sensitivity on tasks that measure indirect behaviour, raising the possibility that young children have been underestimated in their cognitive competency.

In a groundbreaking study, Onishi and Baillargeon (2005) used a violation-of-expectancy (VOE) paradigm that capitalized on the tendency for prelinguistic infants to look longer at events that they find surprising. Infants (15-month-olds) watched familiarisation scenarios in which an agent placed a watermelon toy into a green box (trials 1 and 2), and then reached into the green box (trial 3). Infants then experienced one of four belief induction trials: in the TB-green condition there was no movement of the toy (it remained in the green box, where the infant saw it being placed in the familiarisation trials); in the FB-green condition the agent was absent when the toy moved from the green to the yellow box; in the TB-yellow condition the agent was present when the toy moved via self-propelled action into a yellow box; and in the FB-yellow condition the agent witness the toy move into the yellow box but was absent when the toy subsequently moved back to the green box. After a pause in which nothing happened, infants experienced one of two test trials: the agent either reached into the green box or reached into the yellow box. The infants in the TB-green and FB-green conditions looked longer if the agent reached into the yellow box, while the opposite pattern was revealed in the TB-yellow and FB-yellow conditions. The authors concluded that the infants were surprised when the agent did not act according to her true or false belief about the toy's location. They argued that by minimizing task demands they were able to show that 15-month-olds "realise that others act on the basis of their beliefs and that these beliefs are representations that may or may not mirror reality" (p.257). VOE methods have since been used to suggest that infants as young as 7-months old can attribute false beliefs to others (e.g., (e.g., He, Bolz, & Baillargeon, 2011; Kovács, Téglás, & Endress, 2010; Scott & Baillargeon, 2009).

Onishi and Baillargeon's (2005) work inspired a wave of studies using other non-traditional techniques to investigate early mindreading competence. For example, Southgate, Senju, and Csibra (2007) tested a sample of 25-month-old's using a non-verbal anticipatory-looking (AL) procedure. They presented the toddlers with a change-of-location task in which an agent could retrieve an object from one of two boxes by reaching through one of two windows. In one false-belief condition (FB1) the agent saw that the object was placed in Box A and then transferred to the Box B, but they did not the object's subsequent removal from the scene. In a second false-belief condition (FB2) the agent saw the object being placed in

Box A but did not witness its movement to Box B or its removal from the scene. When the agent was positioned in readiness to retrieve the object, toddlers looked first and longest to the correct locations in the FB1 and FB2 conditions (Box B and Box A respectively). As the two different conditions eliminated non-mental state explanations of looking behaviour (e.g., looking at the first or last location of the object, or looking at the last place the actor or puppet attended to) Southgate and colleagues concluded that the findings strongly suggest that 2-year-olds can attribute false beliefs to others. In support of their conclusion, evidence from numerous AL studies suggest that children younger than 3 years can predict another's future actions based on their belief content (Luo & Baillargeon, 2007; Senju, Southgate, Snape, Leonard, & Csibra, 2011; Surian, Caldi, & Sperber, 2007).

As well as VOE and AL methods infant laboratories have reported early emerging mindreading capabilities using preferential-looking, anticipatory-pointing, emotional-response tasks (Knudsen & Liszkowski, 2012; Moll, Khalulyan, & Moffett, 2016; Scott, He, Baillargeon, & Cummins, 2012) and neural measures (Kovács, Kühn, Gergely, Csibra, & Brass, 2014; Southgate & Vernetti, 2014). Not only have these studies suggested that infants and toddlers are capable of representing other's beliefs, there is also evidence to suggest that young children can reason about complex causal interactions between false beliefs and other mental states. For example, 18-month-olds seem to even consider others' false beliefs about an object's content, property, or identity when interpreting their actions and when forming expectations about their behaviours (Scott, Richman, & Baillargeon, 2015).

Other researchers document early false-belief reasoning by taking advantage of toddlers' propensity to help others (Buttelmann, Carpenter, & Tomasello, 2009; Buttelmann, Over, Carpenter, & Tomasello, 2014; Southgate, Chevallier, & Csibra, 2010). In Buttelmann and colleagues' study, 18-month-olds watched an agent place a desired toy in Box A. In the true-belief condition, but not the false-belief condition, the agent saw the toy being transferred into Box B. In both conditions the agent then attempted to open Box A but failed to do so. Infants in the false-belief condition helped the agent by opening Box B. They reasoned that the agent mistakenly thought the toy was still inside Box A. By contrast, infants in the true-belief condition helped by opening Box A because, argued the authors, the infants reasoned that the agent knew where the toy had been moved to, so the agent must be looking in Box A for another reason. These findings suggest that infants' helping may be guided by a sensitivity to other people's mistaken beliefs.

1.2.4. Replication issues

Since Onishi and Baillargeon's (2005) study, over 30 published papers, using 11 different methods, offer evidence to suggest that infants' and toddlers' understanding of belief is an abstract one (Baillargeon, Buttelmann, & Southgate, 2018; Scott & Baillargeon, 2017; Scott, Roby, & Baillargeon, in press). This impressive body of evidence has encouraged other researchers to consider the wider theoretical implications of a precocious theory mind, and to undertake research of their own. However, some attempts by researchers to devise follow-up studies have been hampered by their inability to replicate the original findings (Sabbagh & Paulus, 2018). This prompted the journal Cognitive Development to release a non-replication Special Issue (2018, vol. 46) which documents a number of unsuccessful attempts to reproduce the findings of notable VOE, AL and helping-behaviour tasks (see also Kulke & Rakoczy, 2018, for qualitative survey of null findings using implicit theory of mind paradigms).

In order to make sense of the failed VOE replications (e.g., Dörrenberg, Rakoczy, & Liszkowski, 2018; Powell, Hobbs, Bardis, Carey, & Saxe, 2018; see also Poulin-Dubois, Polonia, & Yott, 2013; Yott & Poulin-Dubois, 2016) some researchers point to methodological deviations, suggesting that even minor procedural changes can produce negative results (Baillargeon et al., 2018; Rubio-Fernández, 2018). For example, Buttelmann et al. criticise Powell and colleagues' conceptual replication of Onishi and Baillargeon (2005) on the grounds that it may not have given their infants enough time to form an expectation of the agent's future behaviour; unlike the original study, they did separate belief-induction and test trials with a pause, preventing the processing of novel information (e.g., a self-propelled watermelon).

Non-replications of AL paradigms (e.g., Burnside, Ruel, Azar, & Poulin-Dubois, 2017; Dörrenberg et al., 2018; Grosse Wiesmann, Friederici, Disla, Steinbeis, & Singer, 2018; Kulke, Reiß, Krist, & Rakoczy, 2018; Schuwerk, Priewasser, Sodian, & Perner, 2018) are less likely to be challenged on the grounds of procedural differences, especially when original stimuli have been used. One possible explanation for the erratic pattern of findings, and the absence of correct looking-behaviour in true-belief as well as false-belief conditions is a lack of consistency in participant motivation; perhaps participants need to be highly engaged by the agent, in order for them to successfully predict future behaviours (Baillargeon et al., 2018; Scott et al., in press).

Attempts to reproduce toddlers' helping behaviour have also produced partial or null findings (Crivello & Poulin-Dubois, 2018; Fizke, Butterfill, van de Loo, Reindl, & Rakoczy, 2017; Oktay-Gür et al., 2018; Poulin-Dubois & Yott, 2017; Powell et al., 2018; Priewasser, Rafetseder, Gargitter, & Perner, 2018). Priewasser and colleagues (2018) successfully replicated Buttelmann et al.'s (2009) false-belief condition, but in the true-belief condition toddlers were just as likely to assist by opening either box (rather than opening the empty box). Success in the false-belief condition, but not in the true-belief condition, suggests that toddlers were tracking the agent's goal <retrieve desired toy> rather than tracking the agent's belief. In their follow up study the experimenter tried to open a *third* box (box C) on her return. Priewasser et al. argued that if toddlers are sensitive to beliefs, they should help the experimenter open box C in both conditions. Instead they found that toddlers tended to open the box containing the toy irrespective of whether the experimenter had a true or false belief. Crivello and Poulin-Dubois (2018) ran a conceptual replication of the Buttelmann et al. (2009) task. In an attempt to reduce the high attrition rates associated with helping-behaviour studies, the experiment was undertaken at a table rather than on the floor thereby reducing the distance between the toddlers and the boxes. Despite losing fewer toddlers to fussiness or other complications, and tripling the original study's sample size, they found that their toddlers did not perform above chance in either condition. Baillargeon et al. (2018) draw upon a number of factors to explicate the contrary findings, such as differences in set up, procedure, populations tested, statistical power and familiarity with experimenter.

Where do we go from here? It could be argued that the replication failures directly challenge the existence of false-belief understanding in infants and young children, or rather that they reflect the ephemeral nature of early mindreading abilities. What *is* clear is that by age 5 years, children have a demonstrable understanding that others can interpret the world differently from themselves, and from reality. Putting aside the ongoing debate over replication matters, there is still the broader question of cognitive processes: why does early sensitivity not manifest itself in performances on standard false-belief tests?

1.3. *Cognitive processes for tracking and ascribing belief*

Any account of human mindreading must be able to elucidate the contradictory findings in developmental research. A heated debate about the exact nature of psychological reasoning is currently ongoing between proponents of three main perspectives: an early mindreading account, a deflationary account and a dual-process account. The following subsections

provide a brief description of each perspective.

1.3.1. Early mindreading account

Advocates of an early mindreading account (e.g., Baillargeon et al., 2010; Carruthers, 2013, 2015, 2016; Leslie, 1994) claim that infants have a single, abstract psychological reasoning system. According to this "rich" or "mentalist" viewpoint, a fully representational theory of mind is online by the second year of life - some even suggest it is innate (e.g., Baron-Cohen, 1995; Leslie, 1987; Scott & Baillargeon, 2009) or emergent in the first few months of life (e.g., Luo, 2011; Sodian, 2011). Crucially, this single system operates during the course of a human's lifetime and, as Carruthers (2013) points out, "while the operations of this system probably become more streamlined and efficient with age, its representational capacities do not alter in any fundamental way" (p. 142). The abstract mentalistic competencies of infants and young children are underestimated by standard false-belief tasks because these rely on direct (verbal) measures. Infants and toddlers pass VOE, AL and naturalistic helping tasks because these indirect tests only tap the belief representational system, which is operational before the second year of life. Direct false-belief tasks, in contrast, tap the belief representational system as well as response selection and response inhibition skills. The additional cognitive demands imposed by having to select a particular verbal response and to inhibit the temptation to report reality (i.e., the reality bias) make standard false-belief tasks difficult for young children. Essentially, despite having access to a sophisticated mindreading system, 3-year-olds fail because they lack the necessary language, knowledge and executive function to respond explicitly in standard belief testing.

Studies show that performance on direct false-belief tasks is correlated with performance on a range of independent measures of inhibitory control, such as the day/night task, in which children are asked to say "night" when presented with the sun on a white card and "day" when presented with the moon and stars on a black card (Carlson & Moses, 2001; Carlson, Moses, & Claxton, 2004; Flynn, 2007; Perner & Lang, 1999). The claim is that it is only when children have developed executive function skills will they be able to pass direct false-belief tasks. There are, however, several qualifications to such an explanation.

First, Wellman (2014) proposed that 18-month-olds do not need to engage in belief attributions to solve indirect tasks; they can solve the tasks by just reasoning about desire-

awareness. For example, in Buttelmann et al.'s (2009) study, in the false-belief condition, infants helped the agent achieve his desire by opening Box B (which contained the toy); in the true-belief condition, infants helped the agent achieve his alternative desire by opening Box A (which was empty).

Second, older children's understanding of belief is conceptually unified and generalizes across different mindreading scenarios (Low & Watts, 2013; Rakoczy et al., 2015). If responses on indirect tasks tap an early developing and abstract understanding of mental states, then individual infants should also show generalization in their reasoning across tasks. Within the same infant, however, there is little evidence of responses being coherent across different mental-state tasks or being coherent across different contexts of belief induction (Poulin-Dubois & Yott, 2017; Thoermer, Sodian, Vuori, Perst, & Kristen, 2012).

Third, the assumption that conceptual competency is masked on tasks measuring direct judgments of false belief because young children have not yet developed response inhibition and response selection skills must also be carefully considered. For example, Call and Tomasello (1999) found that 4-year-olds performed no better in a false-belief task in which the reality bias was removed (i.e. the children did not know the true location of an object). Moreover, studies show that in the case of diverse desires (e.g., judging that two persons have different desires about the same object) response inhibition and response selection skills are also involved, and yet 18-month-olds perform perfectly well (Repacholi & Gopnik, 1997; Wellman & Liu, 2004). Researchers have also discovered that conflict control is uniformly related to measures of mindreading that impose either high or low executive demands and that better executive functioning does not necessarily translate to better standard false-belief task performances (Carlson, Claxton, & Moses, 2013). Developments in executive function skills may instead help children pick up new information for building concepts about belief. Overall, the relationship between mindreading and executive functioning is much deeper and more complex, and the standard false-belief task cannot be treated as being more or less an executive functioning exercise (Devine & Hughes, 2014; Wellman, 2014).

1.3.2. Deflationary accounts

Further challenges emerge from deflationary accounts in which infant success is construed as the result of low-level processes (Heyes, 2014a, 2014b; Perner, 2010; Ruffman, 2014; Ruffman, Taumoepeau, & Perkins, 2012). Such a stance encourages a leaner, more

cautious interpretation of non-verbal responding on the grounds that impressive (false-belief) task performances do not necessitate mental state representation.

Heyes (2014a) interprets infant success in terms of low-level novelty wherein looking behaviour reflects the extent to which the perceived stimuli are novel with respect to previously encoded events. Specifically, infants look longer at events that they perceive or imagine incorporate novel colours, shapes, and/or movements. Thus, the familiarisation and belief induction phases (vital to most false-belief studies) are effective in influencing infant behaviour, not because they manipulate the belief content that infants ascribe to agents, but because they manipulate what infants believe about experimental props and agent actions. For example, in Heyes' low-level novelty explanation of Onishi and Baillargeon's (2005) findings, infants looked longer when the agent reached for yellow in the TB-green and FB-green conditions because the yellow-reach event was more perceptually novel than the green-reach event. To remind the reader, the agent only ever reached toward the green box in the familiarisation trials. In the TB-yellow and FB-yellow conditions infants *did* look longer at the green-reach outcomes, despite the influence of the familiarisation trials. Heyes argues that in the TB-yellow and FB-yellow inductions the movement of the toy toward the yellow box was perceptually akin to the test event (in which there is a movement of the agent toward the yellow box) so that the novelty of the yellow-reach outcome was attenuated. The infants also saw the toy move towards the yellow box in the FB-green condition, but the encoding of this event was disrupted by the salient return of the agent before the test trial. In response Scott and Baillargeon (2014) argue that a novelty-based explanation of infant looking time infers that false-belief studies have taken place "in a vacuum" (p.60). Such an interpretation overlooks the breadth of experimental research on infants' psychological reasoning providing overwhelming evidence that infants represent psychological events as 'actions of agents on objects', and not just a configuration of 'colour, shapes and movements' (Baillargeon, Scott, & Bian, 2016).

Offering an alternative deflationary account, Ruffman (2014) also illuminates the 'interpretational ambiguity' that emerges from infant studies. Typical false-belief scenarios either require infants to anticipate an agent's behaviour based on his or her perceptions, or to react to an agent's behaviour based on his or her perceptions. Early mindreading advocates claim that infants pass these (non-verbal) false-belief tasks by linking agents' perceptions to their mental states and subsequent behaviours. By contrast, Ruffman and colleagues (2014;

Ruffman et al., 2012) suggest that intervening mental states are not required: infants may predict future actions, or react to past actions, by linking agents' perceptions directly to agents' behaviours. Accordingly, infants rely on past experiences, rather than mental state ascription; they observe others' actions, and use their capacity for statistical learning to encode and categorize others' behaviours. Statistical learning mechanisms generate rules about how people behave and allow future action to be predicted without having to represent the mental state justifying the agent's action. For example, the infants in Onishi and Baillargeon's (2005) VOE study may have based their expectations on the behaviour rule that "people will search for objects where they last saw them." Likewise, the infants in Southgate et al.'s (2007) AL study may have assumed that "people will search for an object at an initial location, only when they have not seen it being moved somewhere else"; and in Buttelmann et al.'s (2009) false-belief condition (where the agent tries in vain to open Box A) toddlers' helping behaviour (retrieving the object from Box B) reflects their understanding that people try to retrieve things from where they last saw them and generally do not stop till they do. The toddlers' helping behaviour differs in the true-belief condition (i.e., they help the agent to open Box A) because of their experience-based knowledge of searching behaviour: people do not typically search for an object in a place where (they are aware) an object is not located. In the true-belief condition then, the toddler is not helping the agent to retrieve the toy, but is helping them to open Box A for another (unknown) reason.

In a recent study, Wellman, Kushnir, Xu, and Brink (2016) present evidence to support the idea that infants use statistical information to make sense of the psychological world. In their VOE experiment, infants watched habituation trials in which an agent picked out blue balls from a transparent box that contained red and blue balls. In the 'minority condition' 80% of the balls in the box were red, while in the 'majority condition' 80% of the balls were blue. In the test trials infants either saw the same agent choose a red ball or a blue ball. The authors found significantly longer looking at the 'choose red' test trial in the 'minority condition' but not in the 'majority condition', suggesting that infants we able to work out the agent's preference based on statistical information. A possible behaviour rule explanation – that people who pick out blue balls will continue to do so – was eliminated as there was no difference in looking time between 'choose red' and 'choose blue' test. Rather, Wellman et al. conclude that "by observing agents repeatedly violating physical probabilities in their intentional actions, infants begin to posit unobservable causal psychological variables", such as desires, wishes and preferences (p. 674).

Evidence that non-human animals' are able to impute mental states (e.g., Call & Tomasello, 2008; Krupenye, Kano, Hirata, Call, & Tomasello, 2016; Whiten, 2013) is also questioned. Theoretically, in infant and non-human animal studies there is always room for a non-mental state explanation wherein there is no way to empirically test between behaviour-reading and mind-reading (termed 'the logical problem' by Lurz, 2011). Mental states come with behavioural correlates, so how can we be sure that infants and non-human animals are using mental state concepts rather than learnt associations when predicting others' behaviours? A recent study (Kano, Krupenye, Hirata, Tomonaga, & Call, 2019) sought to circumvent the logical problem by utilizing Heyes' (1998) goggles task, which is designed to differentiate mental state attribution from behaviour reading by testing whether participants can project their own visual experiences (with various opaque and transparent barriers) onto others. In line with previous research undertaken with 18-month-old infants (Senju et al., 2011), Kano and colleagues found that great apes succeeded in a goggles version of an anticipatory false-belief test, supporting the claim that infants and non-humans can attribute mental states to others. However, some researchers have questioned whether the experience-projection method succeeds in differentiating mindreading from behaviour reading. One criticism is that representing another's line of sight in not equivalent to representing a mental state of seeing (Csibra, 1998; Lurz & Krachun, 2019) and another claims that non-mentalistic solutions are not eliminated as participants can still apply rules around what others tend to do when faced with barriers that may or may not prevent line-of sight (Scarf & Ruffman, 2017).

Scott (2014) argues that a behavioural-rule or statistical learning account does not generate predictions about future behaviours in novel scenarios. It provides only post hoc explanations for positive findings in specific tasks, rather than a coherent explanation for infant behaviour in general. Despite this, following his systematic assessment of infant false-belief studies, Ruffman (2014) concludes that infant performances can be explained by (domain general) statistical learning combined with an innate or early developing curiosity for eyes, faces and biological movement. The transition to adult-like mental state reasoning is then facilitated by language development and vital inputs from the social environment. For example, there is a considerable body of research showing that maternal mental state language is associated with mental state talk in children and with children's performances in theory of mind tasks (Ruffman, Puri, Galloway, Su, & Taumoepeau, 2018; Ruffman, Slade, & Crowe, 2002; Ruffman et al., 2012; Taumoepeau & Ruffman, 2006, 2008).

Could it be, as claimed by Carruthers (2018) that there are far too many experiments demonstrating a sophisticated grasp of belief-reasoning, using disparate measure, procedures and age groups for this to be a credible conclusion? According to Leslie (1987), it is hard to conceive how perceptual evidence alone could ever allow a child (or even adult) to dream up the idea of unobservable mental states. However, deflationary interpretations continue to strongly oppose the idea of a modular, representational mechanism for infant mindreading.

1.3.3. Dual-process account

We reach a point where neither the competence-masking of the early mindreading account, or the low-level processing of the deflationary account, fully illuminates why young children show sensitivity to beliefs in some tasks but not others. A different solution is offered by the dual-process account (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013; Low, Apperly, Butterfill, & Rakoczy, 2016) which suggests that the apparent contradictions are resolved by supposing that human beings have two mindreading processes with different characteristics.

Efficient (or minimal) mindreading is evolutionarily and ontogenetically ancient, operates quickly, and is largely automatic and relatively independent of domain-general resources, allowing infants, children, and adults to rapidly track others' belief-like states. Efficient mindreading typically guides responses that occur independently of a participant's task and motives, supporting anticipatory looking and other spontaneous behaviours. Fast-paced mindreading, however, comes at a cost; there are limits on the kinds of information that can be processed.

Flexible mindreading emerges later, when developments in language and executive functioning help children learn and form abstract concepts about belief. It is important to remind the reader that the term 'flexible mindreading' refers to a fully developed competence in belief-reasoning, or "the ability to use all cognitively-available facts to ascribe any belief that the subject can, themselves, entertain" (p. 964, Apperly & Butterfill, 2009). Such high-level processing supports belief ascriptions over a wide range of content, covering matters of misrepresentation, aspectuality and non-existence. Flexible mindreading is recruited by tasks that require declarative expressions of, or deliberation about, beliefs—for example, verbally indicating and justifying a protagonist's likely behaviour in a false-belief task (Low et al., 2016). Such flexibility supports belief ascriptions over a wide range of content, covering

matters of misrepresentation, aspectuality and non-existence. However, providing a more fine-grained picture of others' beliefs as such is costly, placing great demands on central cognitive resources.

It must be noted here that Apperly and Butterfill's (2009) account describes a 'two-system' account, while the current thesis largely uses the term 'dual-process'. Arguably, the two terms could be used interchangeably throughout the document, however adopting 'dual-process' reflects a cautious recognition that dual system theories predominantly claim the mind is physically separated into distinct systems or mechanisms (Frankish, 2010). The current research adopts the position that flexible mindreading develops as a relatively separate process, but that this does not preclude the possibility that, while efficient mindreading remains relatively distinct from flexible mindreading, there may an exchange of information to some extent over development (Apperly, 2010). Thus mindreading may be achieved via a *relatively* encapsulated and cognitively efficient process, "provided this process itself does not become increasingly dependent on knowledge, memory, or executive function" (p.229, Butterfill & Apperly, 2016).

The two mindreading processes are distinguished by the type of model of the mind that each relies on. Flexible mindreading uses a canonical model of the mind that supports sophisticated and abstract representations of belief. A canonical model considers the aspectuality of beliefs, so that although Mr. Hyde is Dr. Jekyll, Maxi's belief that Mr. Hyde snuck into the room to hide his chocolate is distinct from his belief that Dr. Jekyll was there. Such flexible reasoning would support attributions of others' false beliefs about identity in the numerical sense, such as when Maxi believes that Mr. Hyde is not Dr. Jekyll. By contrast, the efficient system uses a minimal model of the mind – (also referred to in this thesis as minimal mindreading) that is set to track belief-like states, called *registrations*. A registration is an encountering relationship that persists even when the object is no longer in the agent's field: "one stands in the registering relation to the object and location if one encountered it at that location and if one has not encountered it somewhere else" (p.962, Apperly & Butterfill, 2009). Registration is therefore belief-like in that it has a correctness condition which may or may not obtain but it falls short of being a proper propositional attitude in that it does not consider how a particular state of affairs is represented to the other. If Maxi has a belief-like relational attitude to his chocolate and its position in the drawer, and if he did not encounter his chocolate being moved to the cupboard, then he has an incorrect registration of the

chocolate's whereabouts.

Tracking others' registrations as a proxy for their beliefs can guide infants', children's, and adults' expectations or anticipations of others' actions in a useful but limited range of situations. Given that a registration is a relation between an agent, object, and location, the contents of belief-like states are not aspectual (i.e., they do not distinguish the different guises under which objects and situations are represented). If Maxi registered Mr. Hyde moving the chocolate, and if Mr. Hyde is Dr. Jekyll, then Maxi also registered Dr. Jekyll moving the chocolate. Accordingly, if tracking registrations is limited to efficient processing of what others see but not how they see it, then it is theoretically possible to detect *signature limits*. A signature limit of a system is a pattern of behaviour that the system exhibits which is both defective, given what the system is set up to handle, and peculiar to that system. For example, young children (and adults in fast-paced social interactions) can rely on their minimal mindreading system to perform well in simplified, unexpected transfer tasks, but it will not allow for success in false-belief tasks in which an agent has a false belief about numerical identity. Identifying signature limits can therefore provide evidence concerning which systems underlie performance on different tasks and in different types of subjects (e.g., Carey, 2009).

Low and Watts (2013) tested the dual-process account by investigating whether efficient mindreading, as compared to flexible mindreading, would be subject to certain signature limits. Specifically, they asked whether children and adults could efficiently process another's false belief about an object's numerical identity (a hallmark of appreciating how beliefs are essentially aspectual). Low and Watts reasoned that, without using a canonical model of the mind (propositional attitude representation), the operation of efficient mindreading processes would fail to take into consideration the particular way in which a single object was construed from different perspectives. They hypothesized that children and adults would exhibit correct anticipatory looking in a standard object-location false-belief task (in which a false belief about an object's location was induced), but not in an object-identity false-belief task (in which a false belief about an object's identity was induced). In the object-identity task, familiarisation trials introduced participants to an agent who always selected blue and not red toys (the agent's colour preference was counterbalanced). In these trials, participants saw two boxes in front of a screen in which two windows had been cut. Behind the screen the participant could see an agent with full visual access to the boxes, one

of which contained a blue object, the other a red object. When the lights around the windows flashed the agent always reached through one of them to retrieve a blue object.

The object-identity test trial involved a single object, a symmetrical dog-robot toy (painted blue on one side and red on the other). The toy was placed into Box A (red side forward to begin with) before the agent (facing participants) entered the room (see Figure 1-1, row A). Then the agent saw the toy move from Box A to Box B, appearing red to the participant and blue to the agent. On reaching Box B, the toy spun around, showing its dual aspect to the participant only (there was a recessed window on the side of Box B that only participants could see). It then moved from Box B back to Box A, this time with its red side facing the agent and its blue side facing the participant. If the participants were able to efficiently track beliefs as such, they should have inferred that the agent believed there were two objects (a red dog in Box A and a blue dog in Box B).

All age groups (3-, 4-year-olds and adults) showed incorrect gaze anticipation (looking first and longer at the full box containing the object itself). The same participants also completed a standard object-location false-belief task: individuals across age groups showed correct gaze anticipants (looking first and longer at the empty box). Participants' direct reasoning was not subject to signature limits – the accuracy of participants' verbal predictions increased with age (showing above chance performance from age 4 years onwards). Evidence that, from age 4 years, participants were able to effortlessly track a false belief about an object's location, but not about its identity, supports the premise that the humans utilize efficient mindreading capabilities that are limited in scope. Moreover, Low and Watts' pattern of findings have been replicated in studies testing diverse ages, populations and paradigms (e.g., Fizke et al., 2017; Low, Drummond, Walmsley, & Wang, 2014; Mozuraitis, Chambers, & Daneman, 2015).

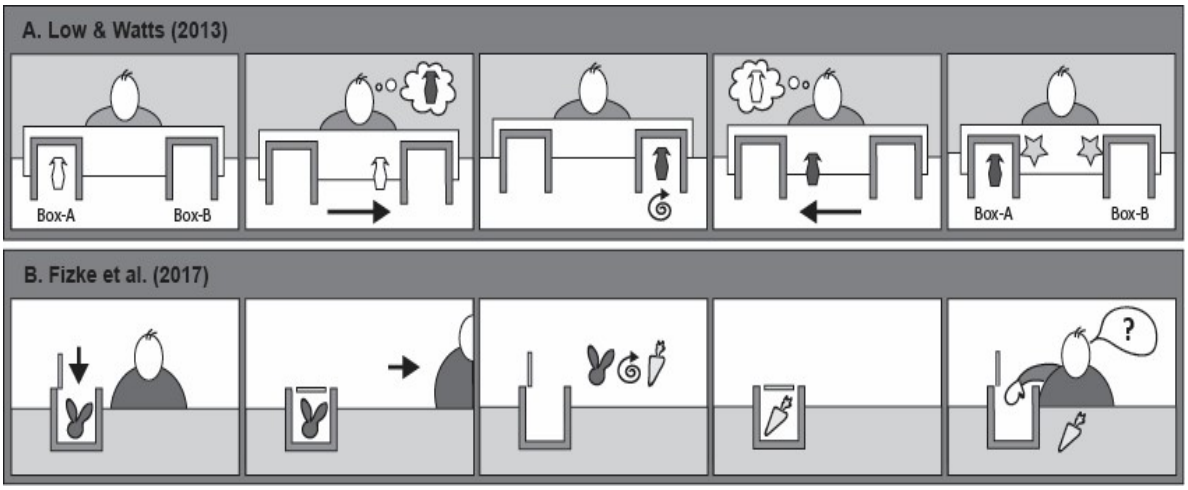

Figure 1-1 Key events in two object-identity tasks in which the participant, but not the agent, is aware of the dual identity of an object.

In a task measuring looking behaviour (A: Low & Watts, 2013), a toy-dog travels from Box-A to Box-B, appearing blue to the agent and red to the participant. The object spins around and travels back to Box-A with its blue aspect facing the participant and its red aspect facing the agent. Flashing lights then alert the participant that the agent will execute a reaching action into one of the boxes. Children and adults looked incorrectly to Box-A. In a task measuring helping behaviour (B: Fizke et al., 2017), an agent places a single bunny in a box and then leaves. In the agent's absence, the bunny (a reversible toy) is transformed into a carrot and placed back in the box. The agent returns, retrieves the carrot and continues to search the box in a quizzical manner. Toddlers helped to search in the box rather than pass the agent the carrot/bunny.

However, there are challenges to the theorizing and empirical findings of Apperly, Butterfill, Low and their colleagues. Scott and Baillargeon (2009) claimed that 18-month-olds *could* attribute false beliefs about an object's numerical identity. In their VOE task, there were two toy penguins, one that could be pulled apart (2-piece penguin) and another that could not (1-piece penguin). The infants watched as an agent placed a key in the bottom half of the 2-piece penguin and then reassembled it. The two penguins now looked identical. In the test trials, when the agent was absent, an experimenter stacked the 2-piece penguin and placed it under a transparent box. She then placed an opaque box over the 1-piece toy. When the agent came back with a key the infants looked reliably longer when the agent chose the transparent box (the unexpected outcome), as opposed to the opaque box (the expected outcome). The pattern of looking responses was interpreted as evidence that infants can engage in a complex chain of causal reasoning. For example, the infants deduced that the agent wanted to put the key into the two-piece penguin and would search for it under the opaque box because the agent had a false belief that the doll under the transparent box was

not the two-piece penguin. Thus, infants looked longer when the agent reached for the transparent box because their causal expectations of her actions were violated.

How substantiated is this claim? Proponents of a deflationary account point out that the agent never orients towards the intact penguin in the familiarisation phase, so the infants' surprise is perhaps due to the first occurrence of this event in the false-belief task (Heyes, 2014a; Ruffman, 2014). In addition, the task itself may not tap into false beliefs about identity in the strict numerical sense (e.g., when someone falsely believes there are two distinct objects in the world when there is, in fact, only one object; Low, Apperly, Butterfill, and Rakoczy, 2016). After all, the presence of two objects (rather than a single object) suggests that infants are simply reasoning about types of objects, irrespective of identity (Butterfill & Apperly, 2013; Low et al., 2016). The agent was aware of two types of penguins, and the task may simply measure infants' ability to track someone's false belief about the location of one type of penguin. Even then, Zawidzki (2013) cautioned that there are no strong grounds to conclude that the agent should form a false belief that the one-piece toy penguin is under the transparent box: Since the agent is aware that the two-piece doll can be assembled to look like the one-piece doll, the agent ought to believe during the test trial that the two-piece doll might be under either of the boxes.

Given that VOE studies with infants and toddlers are fraught with interpretive problems, Low and Edwards (2018) were keen to observe how a mature mindreader would interpret the proceedings of different types of VOE tasks. They asked adult participants to watch videos in which events from established VOE tasks were closely replicated. The authors revealed that while adults considered the event sequences of Onishi and Baillargeon's (2005) object-location task to be meaningful, this was not the case for Scott and Baillargeon's (2009) penguin task. Adults found the event sequences of the latter study difficult to interpret. Specifically, the majority of adults judged the unexpected outcome as being expected: it was deemed reasonable for the agent to reach for the transparent box because at least there was visible penguin to be retrieved. Even those who expected the agent to reach for the transparent box justified this by calling upon object types rather than identity in the strict numerical sense.

In response, Baillargeon et al. (2018) argued that adults' reactions to paradigms designed for infants are not appropriate or informative. However, Poulin-Dubois et al. (2018) maintain that adult responding is a valid way to establish construct validity, especially when

infants are faced with scenarios that are far more complicated than standard false-belief tasks. It is problematic to suggest that the same complex series of events that baffles adults, is understood by infants. More generally, gathering data from samples that differ by age or culture has long been used as a way to provide converging evidence of a particular psychological phenomenon. Critically, "similarities and/or differences in response profiles that persist despite differences in experiences can shed light how tasks are being interpreted" (p. 304, Poulin-Dubois et al., 2018). Nonetheless, Scott, Richman, & Baillargeon (2015) maintain that infants' psychological reasoning system is conceptually rich and abstract, and report evidence that 18-month-olds can even reason about one person's intention to implant in another person a false belief about object identity. This claim will be further addressed in the General Discussion.

Another potential criticism of Low and Watts' (2013) object-identity task is that it might have placed greater demands on working memory than their object-location task. This raises the possibility that signature limits may be an artifact of additional processing costs, rather than reflecting constraints stemming from some (minimal) model of the mind that the efficient mindreading system uses. Carruthers (2015, 2016) suggests that different performances between the object-location and object-identity tasks might be instead due to non-mental content: only the object-identity task required spatial rotation to represent another's perspective. However, signature limits on children's efficient mindreading abilities were also revealed in a study by Fizke et al. (2017) in which performance demands between object-identity and object-location false-belief tasks were matched as closely as possible. Crucially, in their spontaneous helping paradigm an appreciation of the agent's false belief regarding an object's identity did not require mental rotation. Toddlers in an object-location task watched as an agent placed two toys (a bunny and a carrot) into a box (see Figure 1-1, row B). Then, an experimenter removed the bunny and hid it under a tissue, either in the presence of the agent (true-belief condition) or in her absence (false-belief condition). The agent subsequently returned to the box, retrieved the carrot, and then continued to search in the box. In this final phase, children's behaviour differed between conditions; in the true-belief condition they helped the agent continue to search in the box but in the false-belief condition they revealed the location of the bunny. Toddlers in the object-identity task saw an agent place a single bunny into a box. The experimenter revealed that the toy was reversible in the presence (true-belief condition) or absence (false-belief condition) of the agent; she took the bunny out of the box, turned it into a carrot, and then placed it back in the box. In the

true-belief condition, the agent saw the toy as one object with two aspects, whereas in the false-belief condition the agent was led to believe that the bunny and the carrot were two distinct objects. The agent, in the final phase, reached into the box, pulled out a carrot, and continued searching. In contrast with the object-location task, the toddlers' behaviour did not differ between conditions; they helped to search in the box irrespective of the agent's belief. Responding in the object-location task was in keeping with Buttelmann et al.'s (2009) findings; however, the inclusion of an identity component met the predictions of a two-systems account: Toddlers' efficient belief-tracking systems can track false beliefs about the location of objects, but they cannot track false beliefs about identity where two objects appear as one or one object appears as two.

1.4. *The importance of testing adults*

Mindreading has been extensively studied in developmental psychology, but there is comparatively little empirical work exploring theory of mind past the age of 6 or 7 years, "as if there were nothing more to mindreading than the ability to pass tests for the minimal possession of key mindreading concepts" (p.2, Apperly, 2013). However, a full account of human mindreading must also explicate the idiosyncrasies, errors and imperfections of mature mindreaders. For example, why do adults - with their fully developed grasp of belief concepts - sometimes demonstrate egocentric behaviour in theory of mind tasks (e.g., Apperly, Carroll, Samson, Humphreys, & Moffitt, 2010; Birch & Bloom, 2007; Keysar, Lin, & Barr, 2003; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010)? As demonstrated in other areas of cognitive research (number cognition, language, reasoning, etc.), gathering converging data across different age groups and populations is essential when generating models of cognition. Investigating variation or harmony in response profiles of adults with differing past experiences may allow for a more fine-grained appreciation of the mental models that influence human behaviour (Dixson et al., 2018; Hinten, Labuschagne, Boden, & Scarf, 2018; Poulin-Dubois et al., 2018; Xu, Carey, & Welch, 1996). Furthermore, as Low and Edwards (2018) argue, adult responding in mindreading scenarios has the potential to shed light on the precocious performances of infants and toddlers. Afterall, when deliberating the logic of infant false-belief tasks, it is not sufficient to focus on tracking the infant's belief of an agent's belief; we "must also take into account what the researchers may or may not believe about what the infants may or may not believe about the agent's beliefs" (p.648; Heyes, 2014a). As discussed in the previous section, Low and Edwards's findings

showed that testing adults' reactions to infant-based studies can provide a wider context in which to interpret infants' performances.

Arguably the most important reason for testing adults in the current context is that detecting signature limits in a mature mindreader would provide more convincing evidence of a dual-process account than detecting them in younger age groups (where signature limits may be explained in terms of overwhelming task demands). But how should one go about assessing adults' mindreading capabilities? Traditional methods, such as standard false-belief tests, are redundant in older age groups, with ceiling effects typically reached after the age of 4 or 5 years. As a result, researchers have devised other ways to determine variation in older participants: some employ more complex tasks (e.g., Happé, 1994), others seek to exploit our heuristics and biases (e.g., Birch & Bloom, 2007). But, using different methodologies for different age populations is problematic when seeking to establish continuities or discontinuities between early developing and mature mindreading (Apperly, 2010). To mitigate this, another approach is to measure adult responding in tasks that are conceptually comparable to those used to test young children. This method has been effective in teasing apart the component processes of mindreading such as encoding, storing and using information about others' mental states (Apperly, Back, Samson, & France, 2008; Apperly et al., 2010; German & Hehman, 2006) – and it is this methodological approach that is adopted in the current research.

1.5. *The current research*

Currently, there are several limitations in what is known about the development of belief understanding. The problem is that, whilst measuring indirect behaviour has led to impressive advances in the theory of mind field, looking time responses alone cannot definitively answer the question of whether efficient belief-tracking is underpinned by a canonical understanding of belief, statistical learning or a minimal understanding of belief-like states (Fizke et al., 2013; Schneider et al., 2014; Schneider & Low, 2016). Whereas looking time signposts competency, it does not illuminate the underlying cognitive processes (Haith, 1998). One of the downsides of using a method originally designed to answer perceptual and sensory questions is that researchers must be prepared to defend their high level cognitive interpretations against perceptual ones (Heyes, 2014a). Furthermore, as highlighted in Section 1.4, impressive advances have been made through investigating children's differential responding in false-belief tasks, but that is not the only way to study the

development of belief reasoning. We can make sense of some of the conflicting findings outlined in this chapter by considering the nature of the mature mindreading system that children grow into (Apperly, 2010). For centuries philosophers have questioned whether the mind is unitary, but it is only over the past few decades that dual-process accounts have been developed in psychology to explore social cognition (as well as learning, reasoning, and decision-making) (Frankish, 2010). Since its publication in 2009, Apperly and Butterfill's dual-process account of mindreading has attracted considerable attention, both in terms of theoretical debate and empirical research. Acknowledging its influence on the ongoing exploration of mindreading processes, the overarching aim of current thesis is to shed light on the nature and development of mindreading by testing the dual-process account using two novel paradigms. Furthermore, in recognizing the problems associated with a reliance on looking behaviour tasks and infant samples, the current tasks measure reaction times in adult samples. The current body of work tests the dual-process account with the aim of illuminating the relationship between early and mature mindreading. By exposing adults to two paradigms that are conceptually analogous to those devised for the study of early mindreading the current thesis can be more confident when making inferences about the cognitive continuities and discontinuities between early and mature mindreading.

1.5.1. The action prediction paradigm: Experiments 1 and 2

One conjecture of the dual-process account is that *representations underpinning efficient belief-tracking relate agents to objects*, leading to the prediction that efficient processing cannot handle false-beliefs involving identity. Chapters 2 and 3 set out the methodology and results of two experiments undertaken to investigate whether this signature limit is evident in an adult sample. The experimental procedure was motivated by the need to design a novel task for adults that was conceptually related to developmental procedures, with the potential to uncover the component processes underlying theory of mind.

1.5.2. The object-detection paradigm: Experiments 3, 4 and 5

Chapter 4 outlines research that has been undertaken to address a second mindreading puzzle: Why is it that adult mindreading is sometimes automatic and sometimes not automatic? In testing the dual-system account using a novel object-detection paradigm (Chapters 5, 6 and 7), the current thesis offers a second conjecture: *representations underpinning efficient belief-tracking either do not specify agents' locations or do not specify objects' orientations*. This leads to the prediction that efficient belief-tracking alone will not

yield expectations about agents' perspectives. If this is the case then the signature limit of efficient mindreading is not defined by drawing upon numerical identity alone.

CHAPTER 2. *Experiment 1*

This chapter contains the methodology and results, written by Katheryn Edwards, from an experiment contained in a published article with the following citation:

Edwards, K., & Low, J. (2017). Reaction time profiles of adults' action prediction reveal two mindreading systems. *Cognition*, *160*, 1-16. doi.org/10.1016/j.cognition.2016.12.004

2.1. *Introduction*

The evidence for the extent to which representing beliefs about an object's identity is a signature limit of the efficient mindreading system is mixed. This is partly due to the current emphasis on infant studies, and the focus upon looking time data. The goal of Experiment 1 was to provide new and converging behavioural data from an adult sample to tease apart the dual-process account from the early mindreading account. To achieve this goal a novel action prediction paradigm was devised which drew upon the work of Southgate and Vernetti (2014) and Low & Watts (2013).

2.1.1. Action prediction

Southgate and Vernetti (2014) investigated infants' and adults' sensitivity to the relationship between an agent's beliefs and subsequent actions. In their unexpected-transfer scenario, 6-month-olds and adults passively observed video presentations in which an agent is induced to have a false belief about the presence of a desired ball. In A+O- trials the agent sees the ball jump into a box directly in front of her. A curtain then drops, preventing her from seeing the ball leave the scene. When the curtain is raised there is a pause before the agent acts in accordance with her belief: she reaches for the ball because she has a false belief that it is in the box. In the A-O+ trials the agent sees the ball jump out of the box and leave the scene. Then the curtain drops, preventing her from seeing it return to the box. When the curtain is raised the agent acts in accordance with her false belief that the ball is not present and does not reach into the box. The authors predicted that their participants would anticipate the agent's reaching action in the A+O- trails but not in the A-O+ trials.

Notably, anticipation was measured by motor cortex activity. Southgate and Vernetti (2014) exploited the finding that the motor cortex is recruited, not only when one is observing another's action, but also when one is generating a prediction of that action (e.g., Cross, Stadler, Parkinson, Schütz-Bosbach, & Prinz, 2013; Kilner, Vargas, Duval, Blakemore, & Sirigu, 2004; Southgate, Johnson, El Karoui, & Csibra, 2010). In their electroencephalography (EEG) study they used a decrease in alpha activity over the sensorimotor cortex as a proxy for motor activation. In doing so, the presence, or not, of alpha suppression informed whether a participant anticipated an agent's movement, based on their appreciation of the agent's 'belief'. The authors found alpha suppression in the 'pause' phase of the A+O- trials (where the agent falsely believed the ball was present), but not in the

'pause' phase of A-O+ trials, (where the agent falsely believed the ball was absent). They subsequently suggested that both adults and infants are able to predict the actions of others based on their beliefs. Southgate and Vernetti's study illustrates the importance of using adults when investigating infant cognition. The authors did so to ensure that their EEG measure of action anticipation produced the expected result in mature mindreaders. In addition, they also measured adults' looking behaviour to validate the use of alpha suppression as a measure of action prediction. Here, they found that adults made more first looks to the agent's hand in A+O- trials compared to A-O+ trials, supporting their neural findings.

Southgate and Vernetti's (2014) findings are frequently cited, by proponents of the early mindreading account, as evidence that infants as young as 6-months-old have an understanding of false belief (e.g., Carruthers, 2016; Scott, 2017; Scott et al., in press.). However, their findings are also consistent with a dual-process account. According to this viewpoint, by tracking agents' belief-like states, or registrations, infants are equipped to accurately predict others' actions in a change-of-location task. Adding an identity component to the task could potentially tease the two accounts apart, as the dual-process account proposes that infants (as minimal mindreaders) fail to appreciate that agents can represent the same object in different ways. Given the potential processing-demand confounds that infant testing incurs, it is preferable to explore this idea using an adult sample (see section 1.4). Adults action predictions may be compared in cases that do - and do not - involve ascribing false beliefs about an object's identity. Furthermore, by careful experimental design it is possible to achieve this by contrasting two aspects of a single response (its speed and accuracy), rather than considering two different responses to a false belief scenario (e.g., anticipatory looking and verbal response).

2.2. *The current study*

The specific aim of the current study was to determine whether adults would react more quickly in situations when they anticipated a particular response from an actor, compared to when they anticipated no response. A simple procedure was devised whereby participants had to select whether they thought someone with a false belief would or would not reach for a box to retrieve a desired or undesired object. To aid understanding, the experimental design and hypotheses are described with an assumption that the actor desires blue objects, but does not desire red objects (the actor's colour preference was counterbalanced in the experiment).

Of particular interest was how the type of object used would affect response times. To determine this, adults' performances were compared in two different action prediction tasks. In the standard unexpected transfer task (henceforth referred to as the 'Location' task) the object seen by the participant and actor was either a fully blue or a fully red ball. In a second task (termed the 'Identity' task) an object specifically designed to investigate the identity component outlined above was used; this was a single, dual aspect dog-robot, which appeared blue if viewed from one side and red if viewed from the other side (see Figure 2-1). Of the two dependent variables, error rates (gauging accuracy) served as a measure of flexible mindreading, whilst response times reflected the extent to which mindreading is affected by efficient processing.

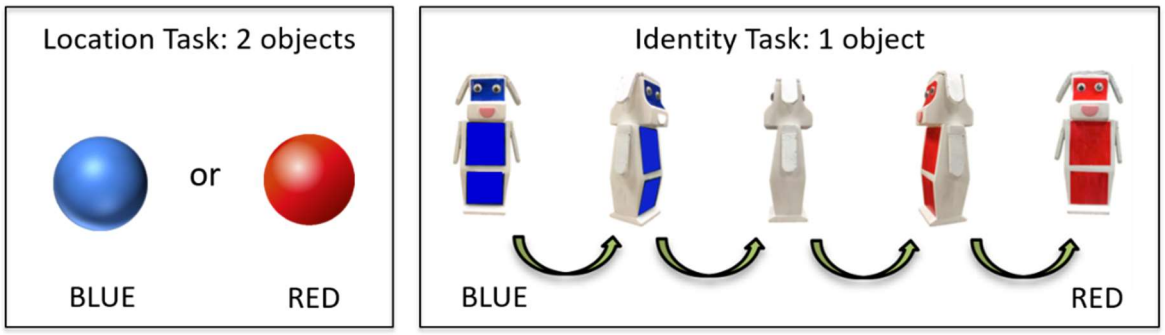

Figure 2-1 Experiment 1: Objects featured in the Location and Identity tasks.

The current study exploits the idea that deployment of motor preparatory mechanisms facilitates the processing of actions (Thillay et al., 2016). The anticipatory activity exhibited by motor cortical neurons translates to a preparatory state in which the motor cortex is primed for optimal processing (Confais, Kilavik, Ponce-Alvarez, & Riehle, 2012). Thus, the justification for measuring reaction times derives from the robust evidence that motor activity occurs *prior* to observing a movement (Kilner et al., 2004) and that by pre-activating cortical areas, motor preparation mechanisms will lead to speeded response times (Bidet-Caulet et al., 2012). To maximize this effect, participants' responses had to correspond to the right-handed reaching movement of the agent; participants were required to use their right hand only to reach for a response key in every trial.

The present research preserves the rationale of Southgate and Vernetti (2014)'s action prediction paradigm but transforms it into a reaction time study. The central feature of this

modified procedure was an 'identity' component that allowed us to investigate the existence of signature limits on adults' efficient belief reasoning when differing perspectives lead to different experiences of the same object. Two hypotheses were tested based on the dual-process account. Hypothesis 1 was that, in the Location task, participants would be fastest to respond when the actor falsely believed that a desired (blue) object was in the box (the AD+ condition). This is referred to as the 'Location Hypothesis'. By contrast, Hypothesis 2 was that participants in an Identity task would be fastest to respond when the actor falsely believed that an undesired (red) object was in the box (the AU+ condition). Henceforth, this will be referred to as the 'Identity Hypothesis'. These predictions are compared with those of an early mindreading account in Figure 2-2. According to a dual-process account, in the AD+ condition of the Location task (Figure 2-2a), participants' efficient processing tracks the actor's registration (or belief-like state) that the preferred ball is in the box, even though it is no longer there. Motor cortex activation is triggered because the actor's goal-directed action is to retrieve it; this then facilitates the fastest responding in the AD+ condition. An early mindreading interpretation (Figure 2-2b) is that motor cortex activity is generated by a single, possibly innate, mindreading system that tracks mental states – in this case, the actor's false belief that the desired ball is present. Both mindreading accounts would predict fastest responding in the AD+ condition of the Location task.

In the AD+ condition of the Identity task, the accounts offer contradictory predictions. According to the dual-process view (Figure 2-2c), there is no anticipatory motor cortex activation as minimal mindreading processes erroneously track that the actor last registered an unwanted (red) object in the box. The signature limit is revealed as a failure to take into account the way in which the actor perceives the object. An early mindreading account (Figure 2-2d) does predict response facilitation in the AD+ condition because of the sophisticated representational capacities of the single-system, which tracks the agent's false belief that the desired (blue) object is the box. An early mindreading account, but not a dual-process account, would predict fastest responding in this condition for the Identity task.

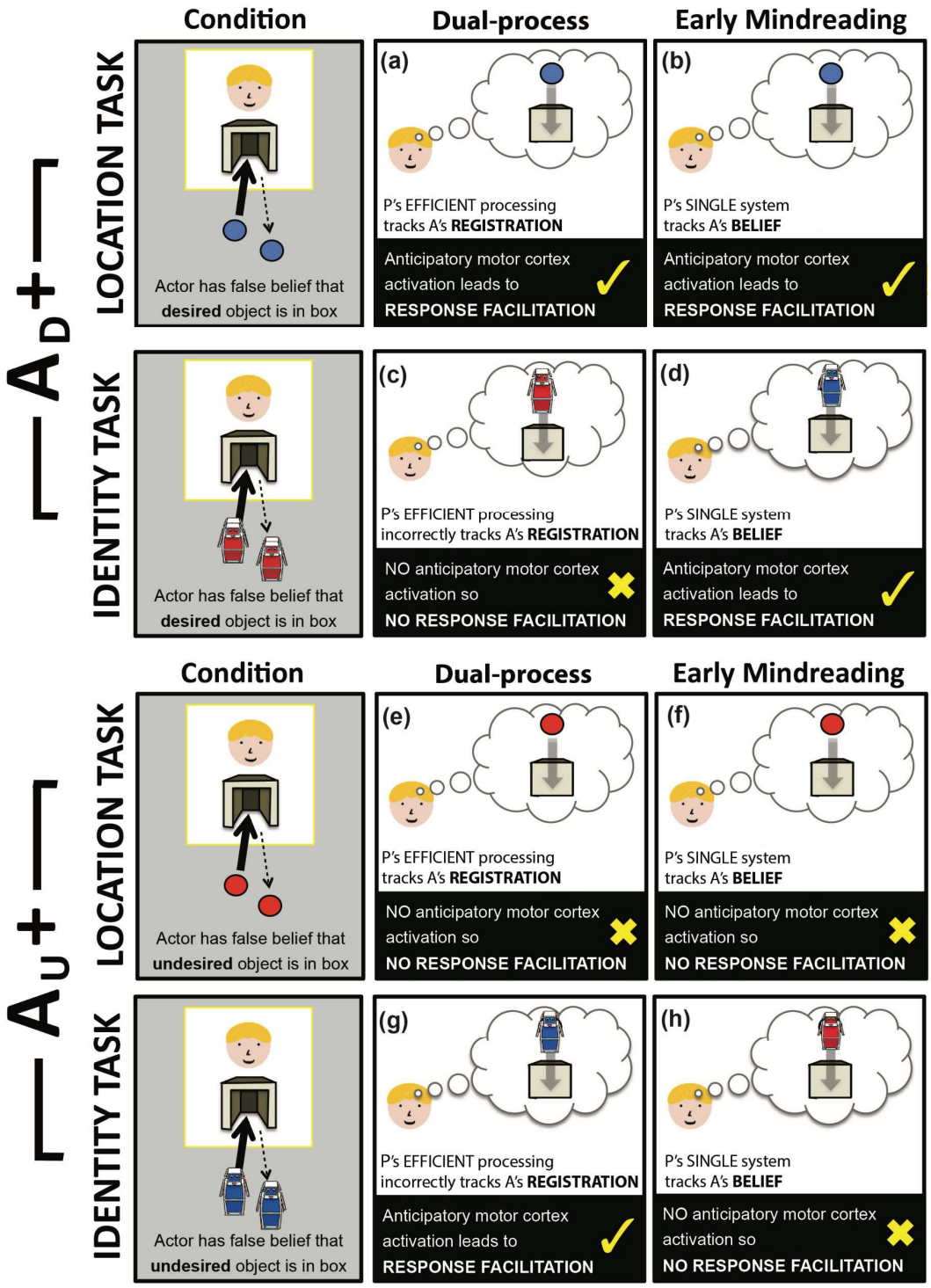

Figure 2-2 Experiment 1: A schematic representation of processes underlying the Location and Identity hypotheses.

The predictions from a dual-process account are compared with an early mindreading account. The solid black arrows in the 'Condition' panels indicate the path of the object witnessed by the actor. The dashed arrows show the path of the object when the actor's view was occluded. In this example, the agent desires blue objects and ignores red objects. Note: P = Participant; A = Actor.

In the AU+ condition, the actor has a false belief that the red object is in the box. In the Location task, both dual-system and early mindreading accounts would predict no response facilitation; anticipatory motor cortex activation would not occur as the participant tracks the actor's registration (Figure 2-2e) or belief (Figure 2-2f) that an undesired object is present. As indicated above, the accounts diverge when forecasting outcomes in the Identity task. A dual-process viewpoint predicts fastest responding in the AU+ condition (Figure 2-2g); this is a seemingly inappropriate result given the actor's goal-directed action towards a blue object. The rationale behind the prediction is that motor cortex activation is triggered when minimal mindreading processes erroneously track that the agent last registered a blue object in the box. This contradicts the early mindreading prediction that there would be no response facilitation in this condition (Figure 2-2h), and that speediest responding will occur in the AD+ condition for both tasks.

To summarize, the current investigation marries methodological ideas from a looking time study (Low & Watts, 2013) with an electroencephalogram study (Southgate & Vernetti, 2014) to yield a new behavioural task that allows us to accurately measure the extent to which representing beliefs about an object's identity is a signature limit of efficient mindreading processes. The dependent variables are error rates and reaction times. Error rates reveal participants' accuracy levels, thereby serving as an explicit measure of belief reasoning. The reaction time measure reflects the extent to which mental state processing is affected by the efficient tracking of belief-like states. In sum, anticipatory motor activation is used as a proxy for action anticipation, which in turn is indexed by facilitated response time. Dissociations in reaction time patterns within location and identity tasks would converge with previous evidence that suggest adult humans have not one but two processes for tracking and ascribing beliefs.

2.3. *Method*

2.3.1. Participants

Participants were 40 right-handed adults (19 females and 21 males) who were recruited from the Victoria University of Wellington campus and local businesses in exchange for a coffee voucher. To determine our sample size we used Southgate and Vernetti's (2014) behavioural findings as a guide. They found that adults made significantly more first looks to the agent's hand in A+O- trials than in A-O+ trials. An a priori analysis using G*Power

(Faul, Erdfelder, Lang, & Buchner, 2007) (input parameters: $\alpha = .05$, power =.8) determined that a sample size of at least 19 participants was required to detect the standardised effect size (r = .65). The standardised effect size was calculated using the formula, $r^2 = t^2 / (t^2 + df)$, where t = reported t-test statistic of the difference between percentage of first looks to A+O- versus A-O+ trials = 2.74, and df = 10. Participants had an average age of 32.7 years (Range 18 to 63). All participants signed informed consent forms before participating and were debriefed orally at the end of the session. The University Human Ethics Committee granted ethical approval prior to commencement.

2.3.2. Design

To test the hypotheses, a 2 (Task: Location, Identity) x 4 (False-Belief Condition: A_D+, A_U+, A_D-, A_U-) within-subjects experimental design was employed. The design concentrated on false-belief reasoning and did not include true-belief conditions for two main reasons: First, Southgate and Vernetti's (2014) method was directly followed; this allowed for the generation of opposing action predictions using only two different false-belief scenarios (see also Krupenye, Kano, Hirata, Call, & Tomasello, 2016; Southgate et al., 2007). Second, the primary prediction was that the relative ease of A_D+ and A_U+ would reverse across different false-belief scenarios. As in Low and Watts (2013), even without true belief data, it is possible to illuminate the cognitive processes underlying different mindreading abilities by zeroing in on dissociations between object-location and object-identity performances.

Participants experienced four familiarisation trials. Half of them saw trials in which an actor had a preference for blue objects and the other half were familiarised with an actor's preference for red objects. They then progressed to the test trials in which they experienced one block of the Location task trials and one block of Identity task trials (order counterbalanced). The instructions were manipulated so that half of the participants were directed to focus on an actor's behaviour and the other half were instructed to focus on her mental state. These instructions were presented prior to the familiarisation videos, and once again before the test phase commenced. Given that this did not affect participants' behaviour the data in these conditions was collapsed.

2.3.3. Stimuli: Familiarisation

Each participant watched four familiarisation videos. Two of the videos featured a blue object and two featured a red object. In the Blue Colour Preference condition each video

began with an actor seated at a table. On the table directly in front of the actor was a lidded box, with an opening that faced the participant. An object (a toy car or toy duck) appeared in the foreground and moved towards the box. When the object was blue the actor smiled and exclaimed, "Yay!" (Figure 2-3a). The object eventually entered the box and was no longer visible to the actor. The actor then lifted her right hand from the table and opened the box's lid to retrieve the object. The final frame showed a smiling actor holding the desired object aloft (Figure 2-3b). If the object was red, the actor frowned and uttered, "Yuck!" as it appeared and moved towards the box (Figure 2-3c). The actor did not retrieve it when it entered the box, instead remaining motionless until the final frame (Figure 2-3d). The four videos in the Red Colour Preference condition showed the same events except that the actions of the actor were reversed; she retrieved the red objects and never the blue. The video dimensions were 19.5cm x 16.5cm and the total duration for four videos (including 1000ms fixation crosses separating each one) was 1 minute 40 seconds. The aim of this phase was to familiarise participants with the actor's colour preference and goal: she desires blue (or red) objects and will act to obtain them, and she does not desire red (or blue) objects and thus will not act. Following the familiarisation phase the participants either proceeded to the Location task or to the Identity task (order counterbalanced).

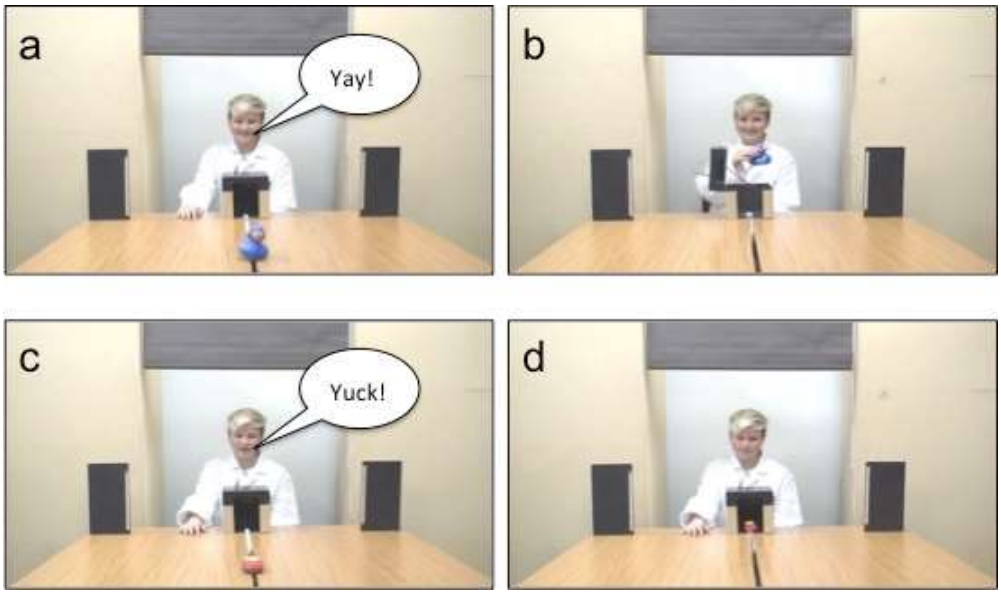

Figure 2-3 Experiment 1: 'Blue Preference' familiarisation video stills.

An actor exclaims, "Yay!" as a blue object appears and moves towards a box (a). When the object enters the box, she opens the lid and retrieves it (b). When a red object appears the actor says, "Yuck!" (c), and remains motionless when it enters the box (d). The actor's behaviour is reversed in the 'Preference Red' familiarisation videos.

2.3.4. Stimuli: Test Phase

Each test trial consisted of a sequence of ten video stills featuring the same actor and setting of the familiarisation trials. The stills (19.5cm x 16.5cm) were presented in chronological order and showed the induction of a false belief in an actor, achieved by changing the location of an object when the actor's view was occluded. The challenge for the participant was to quickly and accurately select the most appropriate outcome of the sequence (from a choice of two) based on the familiarisation phase. A complete test trial comprised a fixation frame (1000ms), followed by ten video stills (each 700ms). The tenth still had a yellow border, to facilitate anticipation of the outcome phase. At the end of each trial the participant was presented with a choice of two images, side by side (each 8.4cm x 6.4cm), in which the actor was either reaching or not reaching for the box (see Figure 2-4).

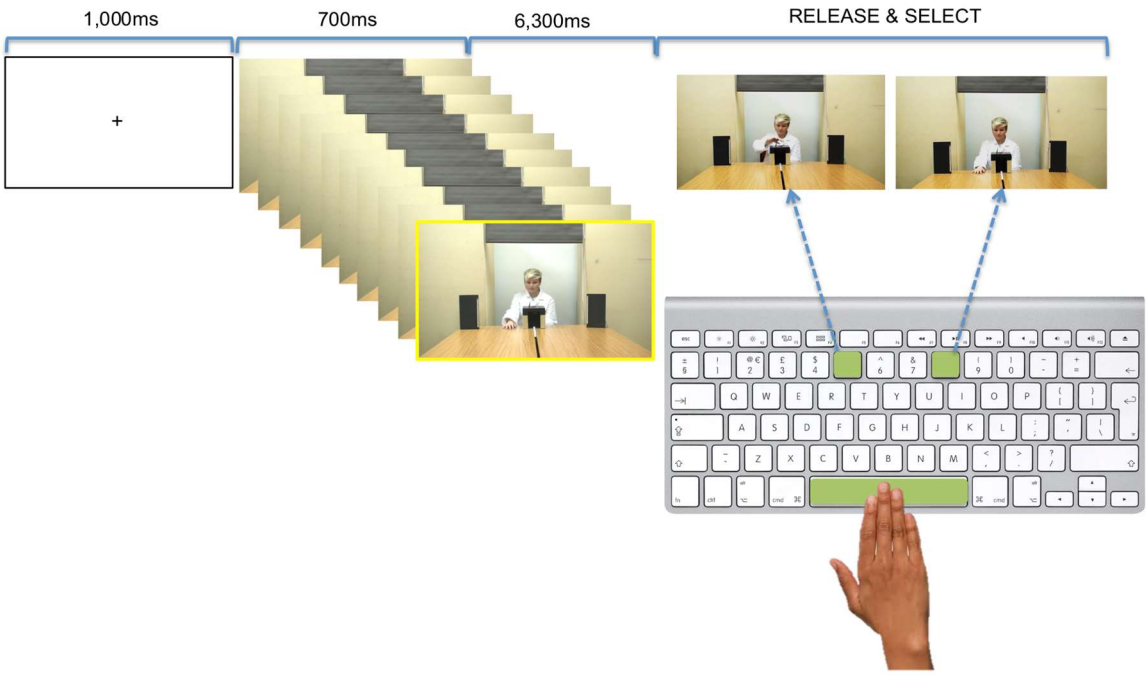

Figure 2-4 Experiment 1: A schematic diagram showing the timeline of a typical test trial in the Location and Identity tasks.

2.3.5. Stimuli: Conditions

The participants experienced four conditions in each task. For ease of understanding the conditions are described in detail below. In each case, the actor desires blue, not red, objects. In two of the conditions the actor falsely believes a blue object is present (AD+) or absent (AD-). In the other two conditions the actor falsely believes that a red object is present (AU+)

or absent (AU-). The critical manipulation between tasks is the type of object used. In the Location task the object is either a blue or a red ball, whereas in the Identity task there is a single object that is blue on one side and red on the other (see Figure 2-1). The dual aspect nature of this object is revealed to the participants in a 20-second video clip in which the object appears on the table and turns 180 degrees anticlockwise four times while the actor sits behind a blind.

AD+ Condition: Figure 2-5 shows how the actor (A) is induced to falsely believe that a preferred blue object is in the box (D+). In frames 1-4 she sees that the object emerges in the foreground and then enters the box. In frame 5 a blind is lowered, so that in frames 6 to 8 the actor does not see the object leave the box. Following the final frame, signaled by a yellow border, the participant must choose the most likely event from a choice of two pictures. In the Location task (Figure 2-5a) the participant and the actor both see the movements of a blue ball, whereas in the Identity task (Figure 2-5b) the actor sees a blue dog whilst the participant sees a red dog. Low error rates in outcome selections were expected for both tasks; participants should ascribe that the actor falsely believes a desirable object to be in the empty box. In the Location task, it was predicted that participants would react quickest in this condition because responding would be implicitly and efficiently facilitated by fast-paced tracking of registrations, instigating motor cortex activity which occurs in anticipation of another's action. In the Identity task it was postulated that response times would not be facilitated; efficient processes would mistakenly track the relation between the actor and the red, undesired, object, which would not trigger anticipatory activity in the motor cortex.

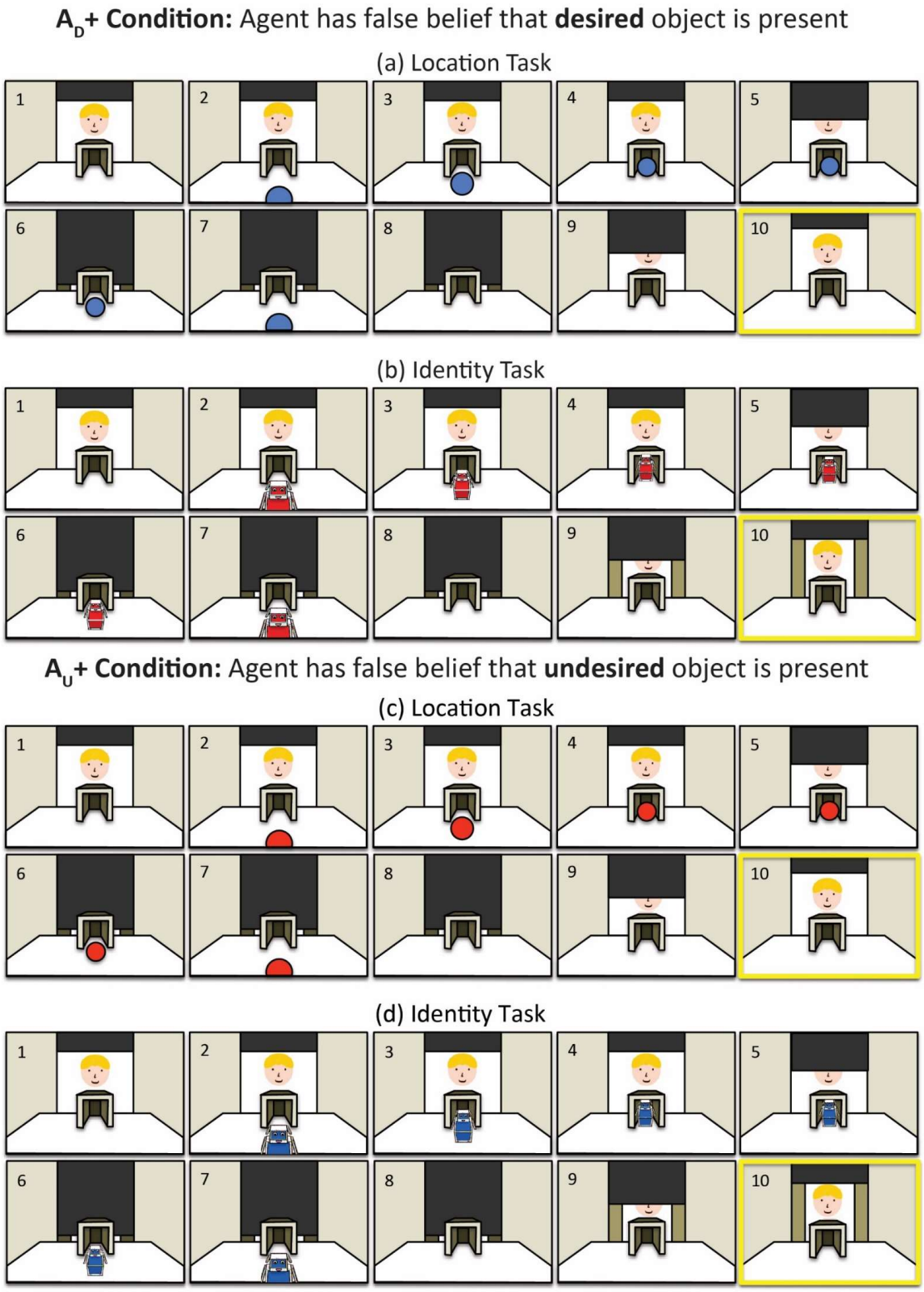

Figure 2-5 Experiment 1: A schematic depiction of the sequence of stills for the A_D+ and A_U+ conditions in the Location and Identity tasks.

In the Location task, both participant and actor see a blue ball, whereas in the Identity task the dual-aspect object requires that the participant sees a red-object while the actor sees a blue object.

AU+ Condition: In the AU+ condition, the actor watches a red object enter the box and but does not see it leave when her view is masked by a blind (Figures 2-5c & 2-5d). In both the Location and Identity tasks it was expected that participants would accurately select the outcome in which the actor does not reach for the box; the actor falsely believes that the object is present but does not wish to retrieve an undesired toy. It was also predicted that there would be no efficient response facilitation in the Location task. However, in the Identity task it was postulated that participants' responses would be implicitly and efficiently facilitated, leading to fastest reaction times in the AU+ condition. The research rationale was that efficient mindreading processes would incorrectly track the actor's registration of a blue, not red, object; it would fail to take into account how the actor perceives the dual aspect dog and would trigger motor cortex activity in anticipation of a reach response. Flexible mindreading ultimately overrides the efficient response facilitation by reasoning that the actor believes there are two different dog-robots across the multiple trials (one blue and one red), just as there are two different blue and red objects in the familiarisation trials. She believes that the red one is in the box so will not reach for the box. The crucial reaction time prediction rests on the existence, or not, of a signature limit in the ability to predict action in others based on the subjective nature of their beliefs.

AD- Condition and AU- conditions: In these conditions the actor is led to believe that either a desired or undesired object is absent (see Figure 2-6). Frames 1-4 show how the object emerges from the box and leaves the scene. The blind is then lowered and the object returns to the box, invoking a false belief in the actor (frames 5-8). It was expected that participants would explicitly and flexibly select the accurate 'no reach' outcome whether or not the object was preferred, based on the actor's false belief. There would be no reason to expect motor cortex activity during the anticipatory phase in these conditions; minimal mindreading processes would implicitly and efficiently track the actor's registration that the object is not in the box.

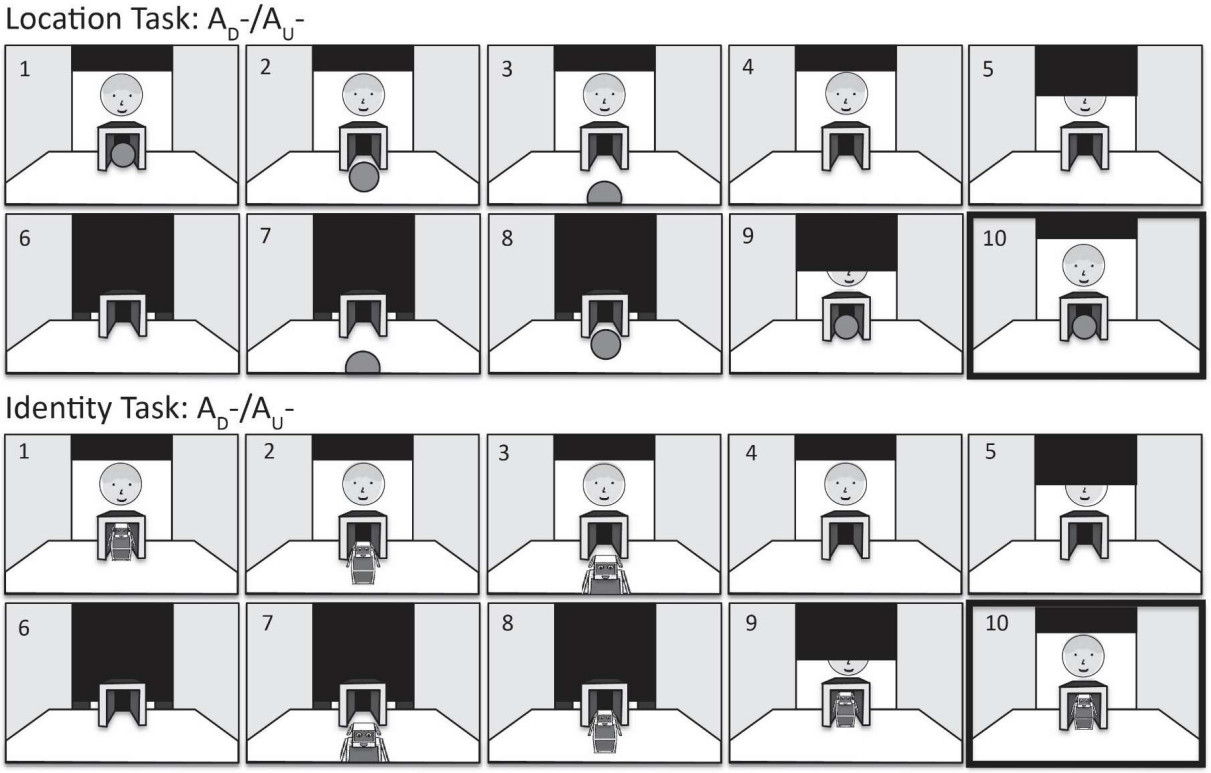

Figure 2-6 Experiment 1: Sequence of stills for the A_D- and A_U- conditions in the Location and Identity tasks.

2.3.6. Procedure

Participants, wearing headphones, sat at a Dell Latitude E5440 laptop (31cm x 17.5cm screen). All stimuli presentation and instructions to the participant were entirely developed and run using E-Prime 2.0. Participants were guided through the task phase via on-screen directions. General instructions, available to all participants, explained the format of the test trials and provided the correct procedure for responding. These procedural instructions were identical for both tasks: "You will see a series of images, one after the other. These are 'stills' taken from videos, like the ones you just watched. The last image in the series will have a yellow border, like this…then you will see two images. Your task is to select the image that best concludes the series as QUICKLY and ACCURATELY as possible." Each trial started with an instruction to press and hold the spacebar with the right hand. It was stressed that the spacebar should not be released until the two images appeared. When ready, the participant was told to click on the "5" key for the left-side image or the "8" key for the right-side image. Participants then proceeded to the trials in both tasks; the only difference was that the Identity task first presented a short clip of a rotating object before continuing (see Figure 2-1). For

each task, 40 sequences were presented in a pseudorandom order; comprising five cycles of four different conditions, each with a counterbalanced left or right outcome image. Thus, participants experienced 80 trials in total. No performance feedback was given after individual test trials to minimize trial time and distraction. On completion of the two tasks participants were debriefed and their data collected.

To address the potential variability of untrained performances (Sternberg, 2004), a training phase exposed participants to 8 practice trials with feedback. These were undertaken before each block of experimental trials and comprised each of the four trial types paired with counterbalanced outcomes (reach outcome on right versus left side). To ensure that training was effective an accuracy threshold of 80% was set. This required that participants had to select the correct answer in 7 out of 8 trials before they could move on to the experimental phase. If this threshold was not met the participants were required to repeat the training block.

2.4. *Results*

Statistical analysis was only undertaken on correct responses, in which the participant selected a response that was consistent with the actor's false belief. Error rates are reported separately in Section 2.4.2. Outliers were excluded from the analysis of response times on the basis of being 3 standard deviations away from the mean response time (between 1% and 2% of individual responses across the four conditions of the Location task and between 1.5% and 2.5% across the four conditions of the Identity task). Initial analysis revealed no colour preference or task order effects. Furthermore, there was no difference in mean response times between the first and second half of trials in each condition. Tests for normality revealed a positive skew in reaction times and error rates. A logarithmic transformation of this data was performed before proceeding with further statistical analysis. Transformed and untransformed means for response times are presented in Tables 2-1 and 2-2, and error proportions in Tables 2-3 and 2-4. Greenhouse Geisser corrections were used whenever the assumption of sphericity was violated (that is, when the Mauchly's test statistic was significant).

Table 2-1

Experiment 1: Logarithmically Transformed Mean Response Times

Task	Condition	*m*	*sd*
Location	A_D+	3.02	0.16
	A_U+	3.16	0.14
	A_D-	3.21	0.14
	A_U-	3.15	0.12
Identity	A_D+	3.16	0.13
	A_U+	3.06	0.12
	A_D-	3.19	0.15
	A_U-	3.20	0.15

Note. N=40

Table 2-2

Experiment 1: Mean Response Times (in milliseconds)

Task	Condition	*m*	*sd*
Location	A_D+	1132.48	474.47
	A_U+	1521.27	480.53
	A_D-	1714.05	553.94
	A_U-	1480.98	421.63
Identity	A_D+	1507.90	496.72
	A_U+	1195.90	416.77
	A_D-	1660.18	601.72
	A_U-	1692.30	650.25

Note. N=40

2.4.1. Response Times

The dual-process and early mindreading theories both predict that reaction times will be fastest when participants expect the agent to reach for the desired blue object based on her false belief (see Figure 2-2). The Location Hypothesis was confirmed when it was found that reaction times were at least 348ms faster in the A_D+ condition than the other conditions. The theories were detangled in the Identity task which revealed that participants responded at least 312ms faster in the A_U+ condition than in all other conditions, as predicted in the Identity hypothesis. Notably, an early mindreading approach would instead predict faster responding in the A_D+ condition.

The critical predictions were tested in a 2 (Task: Location, Identity) x 2 (False-Belief Condition: A_D+, A_U+) ANOVA. In these trials the agent falsely believed that an object was in the box. A significant Task x False-Belief Condition interaction, F (1,39) = 26.53, $p < .001$, $\eta p^2 = .41$, confirmed a selective response time facilitation effect (see Figure 2-7a). Participants were faster to respond when they expected the agent to reach for the desired (blue) single-aspect object in the Location task, but in the Identity condition they were faster to respond when the agent falsely believed that the *undesired* (red) object was in the box.

All conditions were investigated in a 2 x 4 repeated measures ANOVA with Task (Location, Identity) and False-Belief Condition (A_D+, A_U+, A_D-, A_U-) as within-subjects factors. There was no main effect of Task, but there was a main effect of False-Belief Condition, F (2.19, 85.23) = 32.73, $p < .001$, $\eta p^2 = .46$, and an interaction between Task and False-Belief Condition, F (1.86, 72.69) = 19.66, $p < .001$, $\eta p^2 = .34$. To investigate the interaction further the data was separated by Task.

Location Task: As predicted in the Location Hypothesis, participants performed fastest in the scenario where the actor falsely believed the desired object was in the box (see Figure 2-7b). A repeated measures ANOVA revealed a main effect of False-Belief Condition, F (1.49, 57.97) = 32.18, $p < .001$, $\eta p^2 = .45$. Following Bonferroni-corrected pairwise comparisons it was determined that the mean response time for the A_D+ condition was faster than that of the A_U+ condition, t (39) = 5.04, $p < .001$, the A_D- condition, t (39) = 7.19, $p < .001$, and the A_U- condition, t (39) = 5.20, $p < .001$. Response times were significantly longer in the A_D- condition than in the A_U+, t (39) = 4.25, $p = .001$, and A_U-, t (39) = 5.80, $p < .001$, conditions. There was no difference in mean reaction times between the A_U+ and A_U- conditions.

Identity Task: The Identity Hypothesis was supported, in that participants were fastest to respond in the condition in which the actor had a false belief that an undesired object was in the box. Again, there was a main effect of False-Belief Condition, F (1.97, 76.99) = 18.71, $p < .001$, $\eta p^2 = .32$. Bonferroni-adjusted pairwise comparisons showed that response times in the A_U+ condition were significantly faster than in the A_D+ condition, t (39) = 3.96, $p = .002$, the A_D- condition, t (39) = 5.40, $p < .001$ and the A_U- condition, t (39) = 5.07, $p < .001$. There were no other differences (see Figure 2-7b).

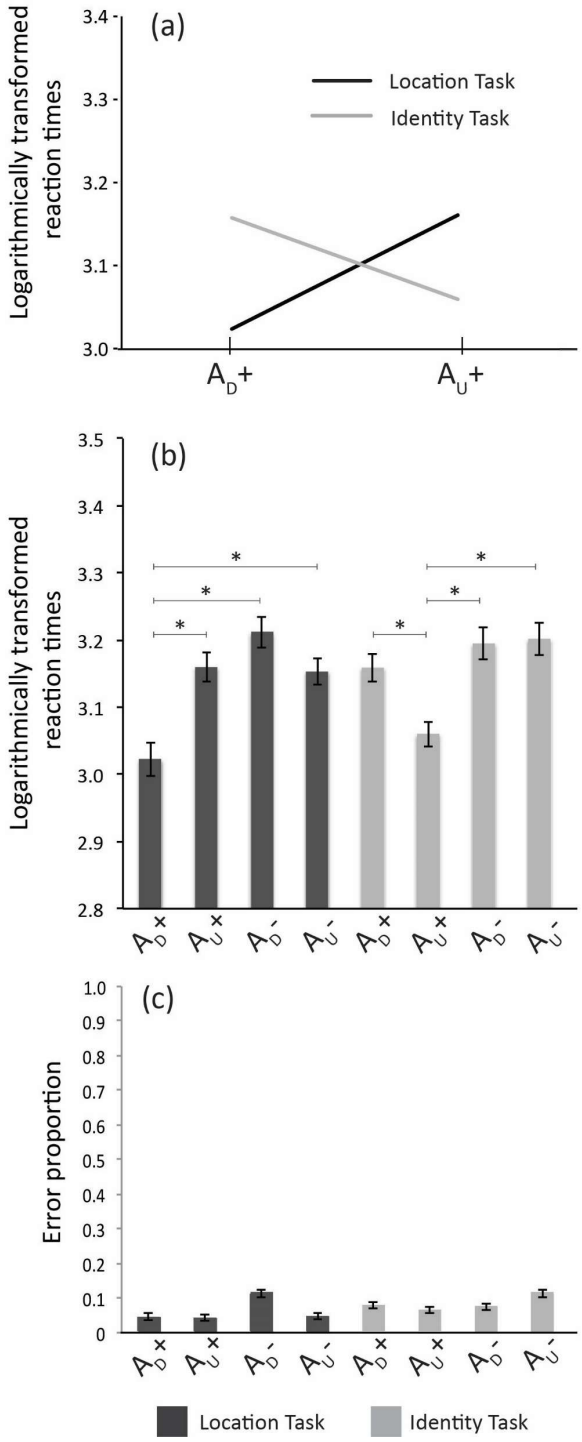

Figure 2-7 Experiment 1: Line chart and bar charts of reaction times and error proportions for the Location and Identity tasks.

*The Task x False-Belief Condition interaction (a) support the Location and Identity Hypotheses for Experiment 1. Bar charts show the logarithmically transformed response times (b) and mean error proportions (c) for the Location and Identity tasks. Error bars represent the standard error of the mean. Note: N=40; * significance level, $p < .01$.*

2.4.2. Errors

Error rates served as a measurement of explicit belief reasoning; overall, participants displayed high performance levels during the training and test trials as revealed by low mean error proportions. There was no evidence of speed-accuracy tradeoffs in the critical A_D+/A_U+ conditions; lower response times for the A_D+ condition in the Location task were not accompanied by significantly greater errors in this condition. Similarly, such a reverse pattern was not found in the Identity task; there was faster responding in the A_U+ condition, but no difference in mean error proportions across conditions. For the practice trials, 95% of the participants, who first experienced the Location task, and 93% of those starting with the Identity task, required just one practice block (of 8 trials) before proceeding to the test trials. The remaining two participants in the Location task, and three in the Identity task, required two practice blocks before moving on to the experimental trials. All participants were ready to proceed to trials after a single block of practice trials in their second task. In the test trials, the overall error rates were low (6% and 9% in the Location and Identity tasks respectively; see Figure 2-7c for mean proportion of errors in each condition). Tests for normality revealed that the error data was positively skewed. To account for this, all analyses of variance were performed on logarithmically transformed data.

In keeping with the reaction time analysis, the initial examination was hypothesis-driven: a 2 x 2 ANOVA between Task (Location, Identity) and False-Belief Condition (A_D+, A_U+). Contrasting with reaction time analysis there was no Task x False-Belief Condition interaction, and no main effect of condition. However, a main effect of Task, F (1, 39) = 10.38, p = .003, ηp^2 = .21, revealed that the proportion of errors was lower in the Location (logarithmically transformed M = .02) than the Identity (M = .03) task. This main effect was also found in the subsequent 2 x 4 repeated measures ANOVA with Task (Location, Identity) and Condition (A_D+, A_U+, A_D-, A_U-), F (1, 39) = 10.38, $p < .001$, ηp^2 = .3, with Identity errors (M = .034) being greater than Location errors (M = .025). There was also a main effect of False-Belief Condition, F (2.60, 101.22) = 2.92, p = .003, ηp^2 = 2.92 and an interaction between Task and False-Belief Condition, F (2.50, 97.67) = 4.09, p = .013, ηp^2 = .01. To examine this further, each task was considered separately.

Location Task: A repeated measures ANOVA determined that mean error proportions differed between the four conditions, F (2.36, 92.06) = 4.92, p = .006, ηp^2 = .11. Pairwise comparisons with Bonferroni corrections revealed that participants made more errors in the

A_D- condition than in the A_D+ condition, t (39) = 3.00, p = .045, or A_U+ condition (see Tables 3-3 and 3-4 for descriptive statistics). It was noted that participants were significantly slower and more error-prone in the A_D- condition. Whilst not the focus of the current predictions this phenomenon may indicate an approach bias, where the presence of the blue ball in the box in the final frame influences the participant's 'reach/no reach' decision.

Identity Task: An analysis of variance revealed no significant difference in mean error proportions across conditions, F (2.75, 107.24) = 2.03, p = .119.

Table 2-3

Experiment 1: Logarithmically Transformed Mean Error Proportions

Task	Condition	*m*	*sd*
Location	A_D+	.02	.03
	A_U+	.02	.03
	A_D-	.04	.05
	A_U-	.02	.03
Identity	A_D+	.03	.03
	A_U+	.03	.03
	A_D-	.03	.04
	A_U-	.05	.04

Note. N=40

Table 2-4

Experiment 1: Mean Error Proportions

Task	Condition	*m*	*sd*
Location	A_D+	.05	.07
	A_U+	.04	.07
	A_D-	.12	.16
	A_U-	.05	.09
Identity	A_D+	.08	.09
	A_U+	.07	.08
	A_D-	.08	.09
	A_U-	.12	.11

Note. N=40

2.5. *Summary*

Response times were compared across four conditions in two separate tasks. Central to the research predictions, the Task x False-Belief Condition interaction for response times revealed that performance was dependent on task. Supporting the Location Hypothesis, participants in the Location task were faster to respond when the actor had a false belief that the desired object was in the box (A_D+ condition). This result concurs with the findings of Southgate and Vernetti (2014). Utilizing their paradigm, the current study went a step further by revealing that adults behave differently when tracking beliefs involving identity. As predicted in the Identity Hypothesis, they were faster to respond in the A_U+ condition, where the actor falsely believed that an undesired object was in the box. One explanation for this behavioural distinction derives from the dual-process account: efficient mind-reading is able to track an actor's registration of an object's *location* but it cannot process how an object's *identity* is represented by the actor.

Consider participant performance in the Location task: in the A_D+ condition, the participant and actor see a desired blue ball enter the box. Then the participant, but not the actor, witnesses the ball leaving the box. Flexible mindreading ascribes that the actor will retrieve the ball because she likes blue things and she *thinks* it is in the box. The crux of the findings, however, is revealed in the implicit measure. According to the dual-process theory, as these events unfold the participant's efficient mindreading processes track the actor's registrations of the changing environment. At the onset of the anticipatory (yellow border) period, efficient mindreading processes record that the actor registered the preferred object in the box. It is proposed that the faster responses for the A_D+ condition in the Location are the result of implicit and efficient processes (tracking of registrations, not belief states) that lead to activation of the motor cortex in anticipation of a reach response from the actor.

Support for the dual-process approach is provided by the response time findings in the Identity task. Consider a participant's experience in the same A_D+ condition. Here, the dual identity dog-robot enters the box and then leaves while the actor's view is masked (Figure 2-5b). Flexible processing allows flexible reasoning (e.g. "I saw a red dog enter the box and then leave, but she thinks a blue one is there, so she'll reach for the box"), but the same pattern of faster reactions in the A_D+ condition was not found because of a signature limit operating upon efficient mindreading. Efficient, but inflexible mindreading is not set up to process how others perceive an object and thus tracks the *location* of the dog-robot but not

how it appears to the agent. In this condition the motor cortex is not activated in anticipation of a reach to the box because efficient mindreading tracks, efficiently but incorrectly, that the actor registered a red dog-robot in the box and thus will remain motionless.

Response times from the A_U+ condition provide further evidence of dual processes at work. In the Location task, the participant and actor see the undesired red ball enter the box, but only the participant sees the ball leaving the box. Efficient mindreading tracks the actor's registrations and flexible mindreading ascribes the actor's beliefs regarding the location of the unwanted object. Unsurprisingly, participants respond correctly that the actor will remain motionless, and they do so significantly more slowly in this condition than the A_D+ condition, indicating that there was no response facilitation due to anticipatory motor cortex activation.

It is the performance of participants in the A_U+ condition of the Identity task that provides key evidence that adult humans possess more than one mindreading mechanism. In this condition, the dual aspect dog-robot enters the box (red side facing the actor), and then exits while her view is occluded (Figure 2-5d). Participants correctly judged that the actor would not reach for the box and, remarkably, they were faster to do so. Significantly faster response times in this condition compared to the other conditions cannot be explained by an early mindreading approach (or by applying behaviour rules). The explanation given here is that while flexible mindreading can explicitly reason that the actor will not reach into the box because she believes it to contain a red object, efficient mindreading fails to account for the way in which the actor identifies the dog (as a red, not blue object) and continues to track the relationship between actor, location and *blue* object. As a result, the A_U+ response times in the Identity condition are facilitated by motor activation, as they are in the A_D+ condition of the Location task. It is noteworthy that there is no statistical difference in response times between the A_D+ (Location) and A_U+ (Identity) conditions, both of which, it is argued here, are accelerated due to efficient processing and the follow-on effects of anticipatory motor cortex activity. Crucially, the main findings were replicated in a second experiment (Chapter 3) in which the task instructions were slightly modified.

CHAPTER 3. *Experiment 2*

This chapter contains the methodology and results, written by Katheryn Edwards, from an experiment contained in a published article with the following citation:

Edwards, K., & Low, J. (2017). Reaction time profiles of adults' action prediction reveal two mindreading systems. *Cognition*, *160*, 1-16. doi.org/10.1016/j.cognition.2016.12.004

3.1. *Introduction*

In Experiment 1, no behavioural effect was revealed when participants were asked to either focus on the actor's mental state or on her behaviour. As both these instructions required the participant to attend to the actor in some way, Experiment 2 sought to determine if an instruction that directed attention away from the agent would influence the overall pattern of participants' performances as compared to Experiment 1.

3.2. *Method*

3.2.1. Participants

Participants were 20 students from Victoria University of Wellington who participated in partial fulfillment of a course requirement. The sample size was adequate to detect the standardized effect size highlighted in Section 2.3.1. The sample included 16 females and 4 males with a mean age of 18.5 years (Range 18 to 20). Consent and ethical approval were granted as for Experiment 1.

3.2.2. Design and procedure

The design and procedure were identical to that of Experiment 1, except for a minor change to the task instructions; rather than being asked to focus on the actor's mental state or behaviour prior to the familiarisation videos and test phase, participants were instructed to focus on the object's location. Seventeen participants proceeded to the test phase after one block of practice trials, the remaining three required two blocks.

3.3. *Results*

Participants' explicit belief-reasoning was highly accurate as shown by the error data. Implicit mindreading differed according to task, revealed by the False-Belief Condition x Task interaction in response times. The crucial finding was that for False-Belief Conditions A_D+ and A_U+, reaction times were reversed; in the Location task participants were significantly fastest to respond when the actor falsely believed that a desired-colour object was in the box whereas in the Identity task they responded most rapidly when the actor falsely believed that an undesired-colour object was in the box. Faster response times in these conditions were not the result of speed-accuracy tradeoffs.

As in Experiment 1, incorrect responses and outliers were excluded from the analysis of response times. Outliers represented between 1% and 3.5% of individual responses across the four conditions of the Location task and between 0.5% and 2.5% across the four conditions of the Identity task. Response times and error proportions were positively skewed so analyses of variance were performed on logarithmically transformed data. Means and standard deviations for both logarithmically transformed and non-transformed data are presented in Tables 3-1 and 3-2. Error rates are analyzed separately in section 3.3.2. ANOVA's revealed that neither preference or task order affected performance, and mean response times did not differ between the first and second half of trials in each condition.

Table 3-1

Experiment 2: Logarithmically Transformed Mean Response Times

Task	Condition	*m*	*sd*
Location	A_D+	3.30	0.09
	A_U+	3.34	0.09
	A_D-	3.35	0.12
	A_U-	3.35	0.12
Identity	A_D+	3.37	0.08
	A_U+	3.31	0.07
	A_D-	3.37	0.08
	A_U-	3.35	0.08

Note. N=20

Table 3-2

Experiment 2: Mean Response Times (in milliseconds)

Task	Condition	*m*	*sd*
Location	A_D+	2016.95	480.76
	A_U+	2257.80	470.14
	A_D-	2314.45	640.54
	A_U-	2332.50	664.50
Identity	A_D+	2357.50	483.76
	A_U+	2054.60	332.07
	A_D-	2390.45	465.28
	A_U-	2279.95	450.51

Note. N=20

3.3.1. Response times

Performance was dependent on task, even under slightly different conditions (a modification of the instructions given to participants). A hypothesis-driven 2 x (Task: Location, Identity) x 2 (False-Belief Condition: A_D+, A_U+) repeated measures ANOVA was undertaken in order to examine the conditions in which the agent had a false belief that the object was *present*. Crucially, an interaction was revealed, F (1, 19) = 22.51, $p < .001$, $\eta p^2 = .54$; participants were quicker to respond when they expected the agent to reach for a desired object in the Location task, but were quicker in the Identity task when the agent was not explicitly expected to reach for undesired object (see Figure 3-1a). Whilst explicitly accurate, participants' implicit mindreading was adversely affected by limits to efficient processing; in the Identity task it failed to account for the way in which the agent perceived the object.

A 2 (Task: Location, Identity) x 4 (False-Belief Condition: A_D+, A_U+, A_D-, A_U-) repeated measures ANOVA determined that there was an interaction between Task and False-Belief Condition, F (3, 57) = 8.68, $p < .001$, $\eta p^2 = .31$. A main effect of False-Belief Condition, F (3, 57) = 4.58, $p = .006$, $\eta p^2 = .19$ was also found. Subsequent analysis considered mean response times for each task in turn (see Figure 3-1b).

Location Task: A repeated measure ANOVA determined a main effect of False-Belief Condition, F (2.19, 41.53) = 4.18, $p = .02$, $\eta p^2 = .18$. Pairwise comparisons with Bonferroni corrections revealed that response times in the A_D+ condition were significantly faster than those in the A_U+ condition, t (19) = 2.94, $p = .046$. There were no other significant differences, though the pattern of response times does trend towards the findings of Experiment 1.

Identity Task: Analysis showed that mean response times differed between conditions, F (3, 57) = 13.93, $p < .001$, $\eta p^2 = .42$, with participants responding significantly faster in the A_U+ condition than in the A_D+ condition, t (19) = 5.70, $p < .001$, the A_D- condition, t (19) = 6.30, $p < .001$, or in the A_U- condition, t (19) = 4.30, $p < .005$. All other comparisons were non-significant. This replicates the findings in Experiment 1, in that participants' responses were significantly faster when the actor falsely believed the unwanted dog-robot was present.

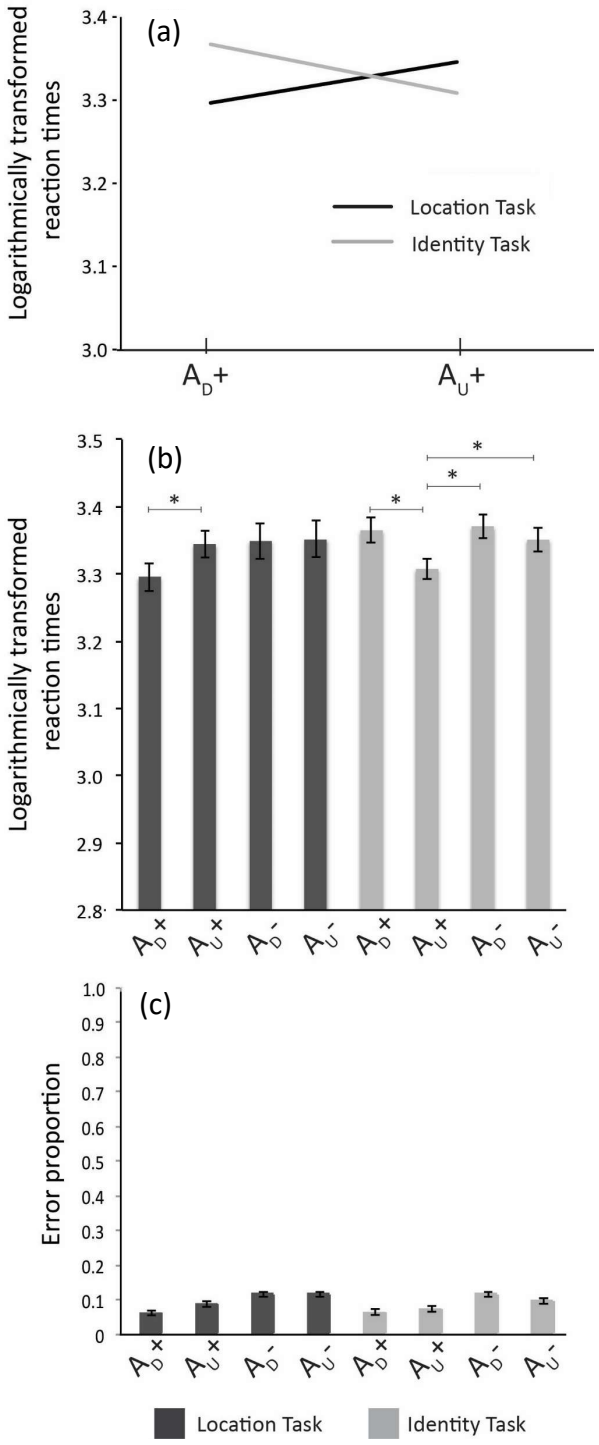

Figure 3-1 Experiment 2: Line chart and bar charts of reaction times and error proportions for the Location and Identity tasks.

*The Task x False-Belief Condition interaction (a) support the Location and Identity Hypotheses for Experiment 2. Bar charts show the logarithmically transformed response times (b) and mean error proportions (c) for the Location and Identity tasks. Error bars represent the standard error of the mean. Note: N=20; * significance level, $p < .01$.*

3.3.2. Errors

Overall, explicit responses in Experiment 2 revealed low error rates for the Location and Identity tasks (10% and 9% respectively; see mean error proportions in Figure 3-1c). Transformed and untransformed mean error proportions are presented in Tables 3-3 and 3-4.

Table 3-3

Experiment 2: Logarithmically Transformed Mean Error Proportions

Task	Condition	*m*	*sd*
Location	A_D+	.03	.03
	A_U+	.04	.04
	A_D-	.05	.06
	A_U-	.05	.04
Identity	A_D+	.03	.04
	A_U+	.03	.04
	A_D-	.04	.06
	A_U-	.04	.05

Note. N=20

Table 3-4

Experiment 2: Mean Error Proportions

Task	Condition	*m*	*sd*
Location	A_D+	.06	.07
	A_U+	.09	.10
	A_D-	.12	.16
	A_U-	.12	.11
Identity	A_D+	.07	.10
	A_U+	.08	.11
	A_D-	.12	.17
	A_U-	.10	.13

Note. N=20

There were no signs of a speed-accuracy tradeoff in the critical (A_D+/A_U+) conditions; faster response times in one condition over the other was not accompanied by significantly higher errors in that condition. A 2 (Task: Location, Identity) x 2 (False-Belief Condition: A_D+, A_U+) ANOVA revealed no difference in error rates, between tasks or conditions, when the agent falsely believed that an object was in the box, F (1, 19) = .21, p = .65. Following on from this a 2 (Task: Location, Identity) x 4 (False-Belief Condition: A_D+, A_U+, A_D-, A_U-) repeated measures ANOVA was undertaken which also revealed no interaction, F (2.84,

53.91) = .18, p = .90. Further analysis revealed no significant difference in error proportions across the Location task conditions, F (2.82, 53.65) = 1.59, p = .20 or Identity task conditions, F (2.19, 41.53) = .87, p = .44. Unlike in Experiment 1, no evidence of a possible approach bias in the Location task's A_D- condition was found.

3.4. *Summary*

There was no notable effect of the instruction manipulation on the pattern of performances. Despite the explicit instruction to focus on the object's location (rather than the agent's mental state or behaviour) the findings suggest that participants' efficient mindreading processes were implicated in tracking an agent' false beliefs. Crucially, Experiment 1's dissociation of behaviour between two different tasks was preserved in Experiment 2: reaction times were appropriately fast in a location scenario that involved tracking an agent's false belief that a desired object was present; but unduly fast in an identity scenario that involved tracking an agent's false belief that an undesired object was present.

3.5. *Discussion of Experiments 1 and 2*

Experiments 1 and 2 tested the extent to which participants' action predictions were affected by the specific content of an agent's false beliefs. Both experiments revealed a blind spot in the efficient mindreading system when participants performed in a false-belief task with a dual-aspect component. In each trial, participants were instructed to select the appropriate outcome of a sequence of events featuring an actor with a clear goal. In one condition (A_D+), the actor watched a desired object enter a box in front of her, but did not see it subsequently leave. Participants responded accurately if they predicted that the actor would reach to retrieve it. In the remaining three conditions, 'no reach' was the correct response as the actor falsely believed that the box was empty (A_D-, A_U-), or that it contained an undesired object (A_U+). Performances were assessed across two tasks; in the Location task all conditions involved either a blue or red single-aspect object, whereas in the Identity conditions there was only one, dual-aspect (red *and* blue) object. As expected, participants were highly accurate in their responses across all trials, but the focus of Experiments 1 and 2 was the *pattern* of reaction times resulting from the four differing conditions.

It was conjectured that, in the Location task, lower-level mindreading processes rapidly triggered motor cortex activity in the A_D+ condition, in anticipation of the actor's reach for a

desired blue ball. However, this appropriate response was not mirrored in the Identity task. Here, efficient mindreading processes failed to predict the actor's action in the A_D+ condition because it was subject to a signature limit over the processing of how the actor perceived the dual object. High-level flexible processing allowed participants to effortfully (and correctly) reason, "I see it as red, but she sees it as blue; she likes blue things therefore she will reach for the box", but the response time was not expedited. Efficient mindreading processes unduly triggered motor cortex activation in the A_U+ condition, incorrectly tracking the relation between the actor, object location and *desired* (i.e., blue) object. The blind spot was revealed by the fastest response times in this condition. To predict the actor's probable action, the participant must infer (from the circumstances in which the actor encounters the object) that she will register it as being blue all over. This is no problem for a flexible mindreading system: someone who only sees one side of a blue ball is likely to assume that it is blue all over because this sort of thing tends to have a uniform colour (or at least she would not expect it to be precisely the colour she dislikes on its reverse side). However, because this requires an appreciation of *how* an object is perceived from different viewpoints, it is suggested that this type of processing it is not within the scope of an efficient system.

Critical to the current study was an experimental design that allowed for the teasing apart of opposing mindreading accounts. Within the constraints of the current rationale, advocates of the early mindreading account would predict anticipatory motor cortex activity when the actor falsely believed a desired object was in the box, irrespective of the object's dual aspect; therefore, fastest reactions would occur in the A_D+ condition for *both* tasks. The present findings, then, question how a single-system framework can account for the inappropriately expedited performance in the identity task. Why do adults respond fastest in a scenario where they explicitly expect no response from the agent? The current data also qualifies a rich interpretation of Southgate and Vernetti's (2014) work. While their findings can be claimed as evidence that infants use mental state representations to predict agents' actions, such claims would require that infants were able to generate on-line representations of a person's perspective irrespective of the object's form. It is proposed that inclusion of a dual aspect component into their paradigm, as demonstrated here, would result in inappropriate action-predictions (as shown by anticipatory motor activation) by infants. Early mindreading advocates would put this failure down to task difficulty but the result of this study support the claim that whilst infants accurately predict the actor's action in the Location task, they do so by tracking her registrations.

Comparing the reaction times in the Location and Identity conditions, Experiments 1 and 2 bear on the speculation that a minimal model of the mind can modulate motor processes; and the fact that reaction times diverge from button selections in the Identity conditions suggests that the influence of efficient processing on motor processes does not involve flexible mindreading or practical deliberation. A full discussion of this point, as well as other areas for future research, will be taken up in the General Discussion.

Experiments 1 and 2 have uncovered something important about how mindreading systems with different processing constraints handle different tasks. Instead of considering two responses to a scenario involving false belief (e.g. anticipatory looking and verbal response), the present innovative method considers two aspects of a single response to a scenario involving false belief. It also shows that incompatible predictions can be manifest in a single response. In the Identity conditions, response times indicate one prediction about an observed action whereas button selections indicate a different, incompatible prediction. The findings also show that task instructions did not influence participant behaviour – an important observation that will be considered in the following chapter.

The findings from Experiments 1 and 2 converge with a range of visual perspective-taking studies that have documented limits on people's ability to track how others may experience the same object or scenery in a different way (e.g., Keysar et al., 2003; Masangkay et al., 1974; Moll, Meltzoff, Merzsch, & Tomasello, 2012; Surtees et al., 2012). Chapter 4 reflects on the overlapping concerns and interests of belief-reasoning and visual perspective-taking research, and highlights how the latter has informed another hotly debated mindreading puzzle.

CHAPTER 4. *Belief Reasoning and Visual Perspective Taking*

This chapter contains content written by Katheryn Edwards, from the following manuscript accepted for publication in Cognition:

Edwards, K., & Low, J. (2019). Leve1 2 perspective-taking distinguishes automatic and non-automatic belief-tracking. *Cognition, 193*. doi.org/10.1016/j.cognition.2019.104017

4.1. *Introduction*

Mature mindreading involves fairly incompatible cognitive demands - sometimes we have to track others' mental states without getting bogged down by detailed ascriptions over what is going on inside their minds, and yet sometimes we do have to make detailed deliberations about the content of others' minds. In Chapter 4 a second mindreading puzzle is considered: How can adult mindreading be both automatic *and* non-automatic? Conflicting findings from belief-reasoning and visual perspective-taking are presented, before the chapter concludes with the rationale behind the design of an object-detection task (Experiments 3, 4 and 5) which weaves together belief-ascription and visual perspective-taking to further test the proposal that the dual and contradictory demands of mature mindreading can be managed by having relatively distinct mindreading systems that impose a trade-off between cognitive efficiency and cognitive flexibility (Apperly & Butterfill, 2009).

4.2. *Belief reasoning and visual perspective-taking*

According to standard philosophical accounts (Davidson, 1980, 1990), beliefs have distinctive features that make inferences about mental states relatively demanding and made only when necessary. Beliefs carry propositional content (i.e., the referents of "that" clauses) and indicate the psychological relation between an individual and the world. Grasping that propositions can be evaluated in different ways by different people helps us appreciate that false beliefs are possible. Belief reasoning also has logical affinities with visual perspective-taking in the sense that both involve representing as well as integrating how the particular way an object, scene or state of affairs is experienced can give rise to different impressions, such as, "I see it as [the turtle standing on its feet], but he sees it as [the turtle lying on its back]." And analogously, "I know that [the chocolate is in the cupboard], but Maxi believes that [the chocolate is in the drawer]" (Apperly, 2010; Hamilton, Brindley, & Frith, 2009; Moll, Meltzoff, Merzsch, & Tomasello, 2012; Zeman, 2017). Appreciating beliefs and visual perspectives supports our inferences of others' actions, and yet the very characteristic that makes such processes cognitively flexible—simultaneously acknowledging contrasting models of a particular thing to different people—is the same characteristic that makes mindreading slow and effortful. On the other hand, it is also commonly supposed that mindreading must be cognitively efficient to play a role during fast-moving social interaction.

Given that these tensions tend not to co-occur in cognitive systems, a mindreading process is computationally efficient if there are signature limits on the kinds of input that can be automatically processed.

4.3. *Sometimes automatic and sometimes not automatic*

4.3.1. Belief-reasoning

It is puzzling that there are seemingly conflicting sets of findings regarding the automaticity of belief inferences. On the one hand, studies measuring response times to unpredictable probe questions in incidental false-belief tasks show that adult humans can work out what someone is thinking, but this is not something that is performed automatically (Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006; Back & Apperly, 2010). Adults take longer to respond to probes enquiring about an agent's belief of where an object is located than they take to respond to probes concerning the object's actual location (Apperly et al., 2006). Adults are only just as fast to respond to belief questions as they are to reality questions when explicitly instructed to keep track of an agent's belief of a target's whereabouts (Back & Apperly, 2010). There is also converging evidence suggesting that adults find it difficult to overcome egocentric biases when making judgements about others' beliefs (Birch & Bloom, 2007; Keysar, Barr, Balin, & Brauner, 2000; Keysar et al., 2003) and their reasoning is impeded by increased cognitive load or decreased executive functioning (Apperly et al., 2008; Bull, Phillips, & Conway, 2008; McKinnon & Moscovitch, 2007; Rowe, Bullock, Polkey, & Morris, 2001).

On the other hand, there is also evidence suggesting that belief inferences can be made automatically (Schneider, Bayliss, Becker, & Dux, 2012; Schneider et al., 2014; Schneider, Slaughter, & Dux, 2017). Schneider and colleages (2014) found that a character's false belief affected participants' behaviour even though it had no bearing on their task. They asked one group of participants to track a character's belief and another to track the location of a ball; both looked longer at an empty box in which the character falsely believed a ball to be, compared to a true belief condition in which the character's belief and ball location were consistent. Automatic belief ascription occurs to the extent that people's own action selections may be influenced by others' beliefs (van der Wel, Sebanz, & Knoblich, 2014), even when participants are explicitly instructed to prioritize their own beliefs (Meert, Wang, & Samson, 2017). Even in a simple object-detection task, where the goal is just to press a

button to detect the presence of a ball, adults' reaction times are speeded when only a bystander happens to believe the object is present, compared to when neither the participant nor the bystander believes the object is present (Bardi, Desmet, & Brass, 2019; Deschrijver, Bardi, Wiersema, & Brass, 2016; El Kaddouri, Bardi, De Bremaeker, Brass, & Wiersema, 2019; Kovács et al., 2010; Nijhof, Brass, Bardi, & Wiersema, 2016; Nijhof, Brass, & Wiersema, 2017). In Kovács and colleagues' object-detection task, adults watched animated movies in which a Smurf character observed a ball move around a table. In the outcome phase a barrier fell away and the participant had to respond if the ball was present. In ball-present trials the critical finding was that, compared to a baseline condition (termed P-A-) in which neither the participant nor Smurf expected the ball to be present, participants were faster to respond when only the Smurf expected the ball to be present (termed P-A+). This finding suggested to the authors that the Smurf's belief regarding the ball's location was automatically encoded. In a follow up set of VOE experiments, 7-month-olds' looking behaviour supported the conjecture that automatic belief ascription is also available to infants.

4.3.2. Visual perspective-taking

Research also shows that calculating others' visual perspectives is sometimes, but not always, automatic. In Samson and colleagues' dot-counting task (Samson et al., 2010) adults are instructed to indicate how many dots they themselves can see inside a room. Studies show that participants experience an altercentric interference effect whereby they respond more slowly and with more errors when an avatar in the room sees a different number of dots, compared to when he or she saw the same number as them (Furlanetto, Becchio, Samson, & Apperly, 2016; Qureshi, Apperly, & Samson, 2010; Samson et al., 2010). Such findings suggest that the mental content of the avatar's visual perspective can be automatically computed, which results in interference during on-line judgements about self-perspective (though the interpretation of such work has been challenged on the grounds that altercentric interference may also be the result of experimental artefacts such as attentional cueing (Cole, Atkinson, Le, & Smith, 2016; Conway, Lee, Ojaghi, Catmur, & Bird, 2017; Heyes, 2014c; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014).

There are, however, different forms of visuospatial perspective processing. Pursuant to Flavell's model (1978, 1992), a simple case, referred to as Level 1 perspective-taking (L1PT), involves calculating the content of what is seen when someone gazes, and can be

processed using line-of-sight information. A higher-level visuospatial perspective problem, termed Level 2 perspective-taking (L2PT), requires an understanding of *how* an entity is appreciated. The latter is regarded as the more representationally complex of the two, evidenced by later ontogenetic development and phylogenetic differences (Flavell, Everett, Croft, & Flavell, 1981; Karg, Schmelz, Call, & Tomasello, 2016; Masangkay et al., 1974; Moll & Meltzoff, 2011). The later-developing L2PT ability has been characterized as involving perspective-confrontation, which entails integrating in a single representation how two people looking at the self-same object from different viewpoints can arrive at different and contradictory descriptions (Moll et al., 2012; Perner, Stummer, Sprung, & Doherty, 2002). Confrontation of perspectives can come about not just when they are mutually exclusive (e.g., that the turtle is perceived as standing on its feet as opposed to lying on its back (Masangkay et al., 1974); or the object is believed to be in one location and not the other) but also arise when the alternatives are compatible. For example, a particular animal can be given two sortals (e.g., bunny, rabbit) allowing individuation of the self-same thing in distinct but synonymous ways (Doherty & Perner, 1998). Nonetheless, young children still treat alternative names as being somehow mutually exclusive. Overall, L2PT involves more than just tracking what someone else sees, but constructing and holding in mind a meta-relation that integrates alternative representations of one and the same thing held by two different people at the same time under a superordinate viewpoint.

Several studies show that humans do not automatically compute how an object might appear differently to people with different perspectives (Hamilton & Ramsey, 2013; Surtees, Butterfill, & Apperly, 2012). In Surtees and colleagues' digit-appearance task, for example, adults were instructed to indicate the numeral that was shown on a table (the stimulus was a rotationally asymmetrical digit such as a '6' or a '9', and there was an avatar positioned behind the table such that he or she saw the digit from the opposite point of view from participants). In contrast to findings from the dot-counting task, there was no evidence of altercentric interference on the self-trials of the digit-appearance task: adults were no slower to respond when the avatar's perspective of the digit was different from their own than when it was the same.

4.4. *The signature limit revisited: Is it all about numerical identity?*

A cornerstone prediction of the dual-process account is that signature limits on the efficient mindreading process arise from the fact that only objects and their relations to agents

can be automatically computed to predict others' behaviour, which in turn means that false belief involving identity in the numerical sense cannot be ascribed by representing registrations. Experiments 1 and 2 add to the supporting evidence showing that humans automatically compute people's false beliefs about an object's location but not its numerical identity (e.g., Fizke et al., 2017; Low et al., 2014; Low & Watts, 2013; Mozuraitis et al., 2015; Oktay-Gür et al., 2018). For example, Low and Watts found that adults' efficient mindreading, as indicated by certain eye movements, allowed participants to make accurate search anticipations when the agent had a false belief about an object's location but not when the agent's false belief about object identity led him to think that there were two objects present when, in fact, there was only one. Experiments 1 and 2 echo these findings, however, the dual-process account has yet to fully articulate the boundaries of the signature limit that distinguishes the automatic but rigid process of efficient mindreading.

Representing mistakes over how objects are represented in the numerical sense may not be the elemental or primary marker that distinguishes efficient from flexible mindreading processes. In Experiments 1 and 2, confronting the truth of the agent's belief certainly requires making attributions of the agent's belief about there being multiple objects versus the reality that there is only one object. However, the absence of an altercentric interference effect on adults' performance on the self-trials of the digit-appearance task (Surtees et al., 2012) is also treated as converging evidence of a signature limit on adults' efficient mindreading, and yet that task does not involve tracking mistakes over numerical identity per se (i.e., the participant and the avatar are both aware there is a single digit on the table and there really is a single digit on the table). Instead of object identity per se, the commonality between such tasks and their constellations is that they require a meta-representational understanding of perspective, evaluating how people's epistemic states are relativized to the specific perspective by which others regard the world. L2PT, involving perspective-confrontation, may be the core signature limit operating on the automaticity of the efficient mindreading process whilst L1PT (e.g., tracking relational attitudes in object-location false-belief tasks or visibility in the dot-counting task) is potentially stimulus-driven and goal independent.

4.5. *The current research*

Experiments 1 and 2 revealed the same reaction time profiles irrespective of task instruction (i.e., "focus on the agent's mental state" versus "focus on the agent's behaviour"

versus “focus on the ball’s location”). This finding is in line with Schneider and colleagues' (2014) change-of-location study in which anticipatory looking (to the empty box) implied that adults tracked an agent’s mental state despite receiving explicit instructions to attend to an object’s location (e.g., “Where do you think the ball is?”). According to the authors, this ‘implict’ belief tracking is consistent with the idea of an automatic theory of mind that is unintentional and unconscious. However, they also point out that implicit belief-tracking does not satisfy the classic classification of automaticity (Bargh, 1994; Shiffrin & Schneider, 1977) given that it is not entirely independent of working memory (Schneider, Lam, Bayliss, & Dux, 2012).

The dual-process account considers mindreading automatic if there is evidence of it occurring “to a significant degree independent of its relevance to the particulars of the subject’s motives and aims” (p.609, Butterfill & Apperly, 2013). Flexible and efficient mindreading are distinct in the sense that the circumstances which influence whether they occur - and which outputs they generate - do not completely overlap. For example, a situational factor like task instructions may strongly affect a flexible (or non-automatic) response but barely effect an efficient (or automatic) one. This is apparent in Schneider et al. (2014) where different explicit instructions, “Where do you think the girl will look for the ball?” or “Where do you think the ball is?” affected explicit responding, but not implicit belief-tracking behaviour. Task instructions also had impact on efficient mindreading in Experiments 1 and 2 (as indexed by reaction time profiles). However, because accurate responding to the primary task (selecting whether or not the agent would reach for the object) required the flexible system to attend to the agent’s mental state in each trial, Experiments 1 and 2 were not suited to exploring automaticity. Therefore, to tackle the automaticity puzzle, Experiments 3, 4 and 5 utilise a task in which the agent’s belief or perspective is never referred to, nor is the participant ever required to predict the agent’s behaviour based on the agent’s belief.

CHAPTER 5. *Experiment 3*

This chapter contains the methodology and results, written by Katheryn Edwards, from an experiment contained in the following manuscript accepted for publication in Cognition:

Edwards, K., & Low, J. (2019). Leve1 2 perspective-taking distinguishes automatic and non-automatic belief-tracking. *Cognition, 193*. doi.org/10.1016/j.cognition.2019.104017

5.1. *Introduction*

As adult humans, we recognize from our everyday experiences that we possess the capacity to make snap judgments about, and slowly cogitate over someone's behaviour. The challenge is to determine the cognitive components that underlie these distinct mindreading abilities. Chapter 5 reports converging findings of a new paradigm, which weaved together belief-attribution and perspectivization, to delineate the boundary of the signature limit operating on automatic mindreading. Specifically, the novel object-detection paradigm measured the extent to which adults were automatically influenced by the belief of a passive bystander in tasks that did and did not necessitate integrating contrasting perspectives. Using a within-subjects design, Experiment 3 profiled adults' reaction times in two closely-matched tasks. In the L1PT task, the participants and the bystander-agent observed a homogenous blue ball and a homogenous red ball moving around a table. At the end of each trial, one of the balls was hidden behind two screens so that neither the participant nor the agent could see it. In the L2PT task, the scene was identical except that a single heterogeneous object (a dog-robot) moved around the table, finishing its movements between the screens by the end of each trial. Both participant and agent were simultaneously shown that the object appeared blue from one viewing perspective and red from the opposite viewing perspective. Critically, the agent was irrelevant to both tasks; the participant was simply required to select the colour (blue or red) that was revealed to himself or herself when the screen rapidly dropped away. The agent either witnessed all events (and so had beliefs consistent with the participant) or was absent for some of the events (so that the agent and participant had inconsistent beliefs). Kovács et al.'s (2010) object-detection paradigm was adjusted as follows: First, the agent was positioned so that he faced the participant, viewing events from the opposite (rather than same) perspective. Second, the opposing viewpoints necessitated the use of two screens (rather than one) to simultaneously mask the objects from the participant and the agent. Third, participants made forced-choice rather than Go/NoGo responses. All trials featured video clips of a human agent in a real-life setting rather than an avatar in an artificial environment.

In the L1PT task, the agent may hold a false belief about the final location of each ball because he was absent when the red ball and blue ball switched places. For example, before the reveal, the agent believes that there is a red ball between the screens and the participant believes that there is a blue ball between the screens. In this task, the agent's belief but not his

visuospatial perspective is relevantly different, for when the screens drop both parties will see a blue ball. There is no confrontation of visuospatial perspective in the two-ball task because the two people looking at the object from different viewpoints will arrive at the same description. In the L2PT task, the agent may hold a false belief about the colour that will be revealed when the occluders drop because he was absent during the object's final rotation. In this case, however, there is also confrontation of visuospatial perspective because at the reveal the two people looking at the self-same object from opposite viewpoints will arrive at different and contradictory descriptions. While both tasks involve tracking another's perspective of an object or objects (the content of what is seen when someone gazes), only the dog-robot task has the additional requirement of confronting perspectives: in this case the participant is required to evaluate how the self-same object is construed from one location, when that construal simultaneously represents the alternative viewpoint that the agent is instead expecting to only perceive from his opposite location. The L1PT task can be differentiated from the L2PT task in that only the latter involves simultaneously confronting two different visuospatial perspectives on the self-same object, which may require embodied self-rotations to imagine assuming others' positions in the world so as to reason about how an object in their environment is experienced by them (Kessler & Rutherford, 2010; Surtees, Apperly, & Samson, 2013b).

For the L1PT task, it was predicted that a bystander's belief about the presence of a specific object would helpfully modulate adults' own reaction times when detecting the presence of that object. However, for the closely matched L2PT task it was expected that adults' reaction times would *not* be speeded when the bystander's belief about the presence of a specific object was dependent on his location in space. If, on the other hand, a facilitating influence of the bystander's belief extended to the L2PT task involving perspective-confrontation, then the dual-process account may be inaccurate and humans instead have a single mindreading process that is context sensitive.

5.2. *Method*

5.2.1. Participants

An a priori analysis using G*Power (Faul et al., 2007) (input parameters: $\alpha = .05$, power =.8) determined that a sample size of at least 33 participants was required to detect the standardised effect size. While not a direct replication, the standardised effect size (r =.45)

was calculated using the formula, $r^2 = t^2 / (t^2 + df)$, where t = reported t-test statistic of Kovács et al.'s (2010) critical effect = 2.42, and df = 23. A total of 54 adult participants, made available by the Victoria University of Wellington's Introduction to Psychology Research Programme (IPRP), signed up to take part in the study. Having a larger number of individuals safeguarded against participant dropout, and other factors affecting data collection such as experimenter error or computer malfunction. All participants signed informed consent forms prior to participation and were debriefed orally at the end of the session. One participant did not perform above chance level and was excluded. As a result, analysis was undertaken on the data of 53 participants. The ratio of females to males was 42/11 and the age mean was 18.36 years (Range 17 to 24). The study was approved by Victoria University of Wellington's Human Ethics Committee.

5.2.2. Materials

All stimuli and instructions to the participant were presented via E-Prime 2.0. Each individual watched a total of 80 videos in an object-detection paradigm. The on-screen video dimensions were 38cm x 21cm; all videos had a frame rate of 25 frames per second (fps) and a 720 x 576 resolution. There were 40 videos in the L1PT task and 40 videos in the L2PT task. Due to total experimental length considerations the duration of each video was reduced by speeding the footage by 120% using Adobe Premiere Pro. As a result, each L1PT video was 13.2 (from 15.8) seconds and each L2PT video was 17.8 (from 21.4) seconds in length. Sample videos used in the L1PT (S1 Movie and S2 Movie) and L2PT tasks (S3 Movie and S4 Movie) are available as supporting information via the following link: https://www.sciencedirect.com/science/article/abs/pii/S0010027719301908?via%3Dihub.

L1PT videos: The L1PT videos began with an agent seated at a table facing the participant. On the table, visible to both agent and participant, were two stationary homogenous balls (one red, one blue) and two wooden screens. In the first movement, the two balls simultaneously moved between the two screens so that they could not be seen by either the participant or agent. Following this movement, the events in the videos varied to create four belief-induction conditions. These conditions differed according to whether the *participant* expected a particular colour to be present (P+) or absent (P-) in the outcome phase and, further, whether the *agent* expected a particular colour to be present (A+) or absent (A-) in the outcome phase.

Expectations were induced by manipulating the movements of the balls and by varying the time that the agent left the scene. The agent's return to the scene signalled the onset of the final phase. There were two possible outcomes in the final phase: either a blue ball or a red ball was revealed when the screens rapidly fell away. As such, participants experienced 8 trial types, comprised of four belief-induction conditions paired with one of two possible outcomes (see Table 5-1a for an overview of conditions). For clarity and efficiency, the four conditions (P+A+, P-A-, P+A-, P-A+) are detailed, when paired with the blue outcome only (trials 1, 3, 5 and 7, as shaded in Table 5-1a).

Table 5-1

Experiment 3: Belief-induction Conditions in the L1PT and L2PT Tasks

(a) L1PT task			
Condition	**Trial**	**Outcome P ≡ Outcome A**	**Expectations based on belief-induction phase**
P+A+	1	Blue	Both P and A expect the outcome.
	2	Red	Both P and A expect the outcome.
P-A-	3	Blue	Neither P or A expect the outcome.
	4	Red	Neither P or A expect the outcome.
P+A-	5	Blue	P, but not A, expects the outcome.
	6	Red	P, but not A, expects the outcome.
P-A+	7	Blue	A, but not P, expects the outcome.
	8	Red	A, but not P, expects the outcome.

(b) L2PT task				
Condition	**Trial**	**Outcome P**	**Outcome A**	**Expectations based on belief-induction phase**
P+A+	1	Blue	Red	Both P and A expect the outcome.
	2	Red	Blue	Both P and A expect the outcome.
P-A-	3	Blue	Red	Neither P or A expect the outcome.
	4	Red	Blue	Neither P or A expect the outcome.
P+A-	5	Blue	Red	P, but not A, expects the outcome.
	6	Red	Blue	P, but not A, expects the outcome.
P-A+	7	Blue	Red	A, but not P, expects the outcome.
	8	Red	Blue	A, but not P, expects the outcome.

Each condition is described following the first movement (in which both balls moved between the screens). Let us first consider the P+A+ and P-A- conditions which resulted in expectations that were consistent between the participant and agent. As illustrated in Fig 1,

events in the P+A+ condition led both the participant and the agent to expect the presence of the blue ball in the outcome phase; in the final movement, both saw the red ball exit the scene, inducing a belief that the blue ball remained between the screens. Likewise, in the P-A- condition, both participant and agent witnessed the blue ball ultimately exit the scene, so that neither were led to believe that a blue ball would be revealed in the outcome phase (i.e., both were expecting the presence of the red ball). The P+A- and P-A+ conditions induced inconsistent expectations. In the P+A- condition, the participant and agent saw the blue ball leave the scene. However, the agent was absent when the red ball exited and the blue ball returned to rest between the screens. In this case, the participant was led to expect the outcome but agent was not. Finally, in the P-A+ condition the agent was present when the red ball left the scene but did not witness the red ball's return after the blue ball's exit. Again, the agent's and participant's expectations were inconsistent as the eventual outcome was not expected by the participant, but it was expected by the agent.

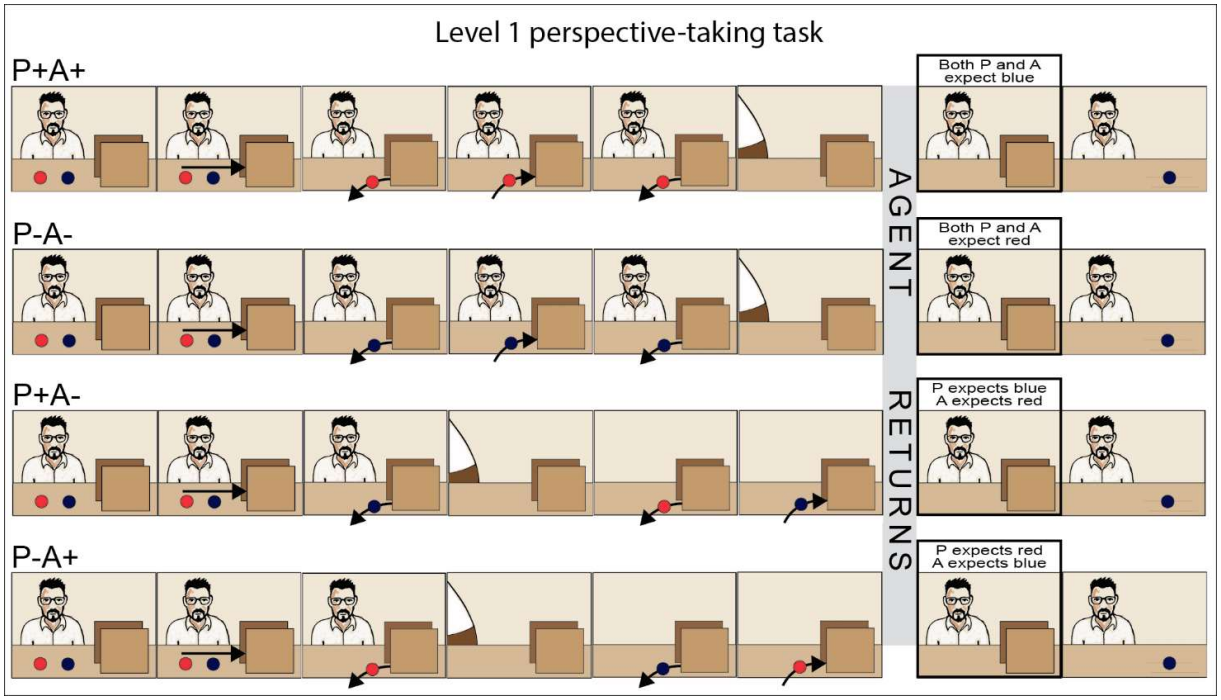

Figure 5-1 Experiment 3: Schematic storyboard showing the main belief-inducing events of the four conditions in the L1PT task movies.

The main belief-inducing events represent conditions where there is a blue outcome. In the P+A+ condition (consistent), both participant and agent expected blue; in the P-A- condition (consistent) neither participant nor agent expected blue. In the P+A- condition (inconsistent) only the participant expected blue, and in the P-A+ condition (inconsistent) only the agent expected the blue.

L2PT videos: The L2PT videos were designed to match the L1PT videos as closely as possible. Each video began with the same agent seated at a table facing the participant. The screens were present but instead of there being two balls on the table there was a single object (a dog-robot) that was blue on one side and red on the other (see Figure 2-1's Identity task object). The dual nature of this object was revealed to the participant and agent at the beginning of each video; it twice turned 180˚ (anticlockwise) before making its initial move behind the screens.

As in the L1PT task, the sequence of events leading up to the final phase varied (according to the object's movements and timing of agent's departure) to create four conditions culminating in one of two outcomes ('blue-facing-participant' or 'red-facing-participant'). This combination resulted in 8 trials types (see Table 5-1b for an overview of all conditions in the L2PT task). Here, the four conditions of blue-facing-participant outcomes are described (see trials 1, 3, 5 and 7, as shaded in Table 5-1b). Figure 5-2 illustrates the critical belief-inducing events following the initial spinning motion of the dog-robot and its first movement between the screens (common to all conditions). Due to the dual nature of the object, the participant's and agent's beliefs were consistent when they expected *different* colours in the outcome phase. For example, in the P+A+ condition, where both the participant and agent expect the eventual outcome (blue-facing-participant, red-facing-agent) the dog-robot's blue aspect was presented to the participant in its last movement inducing a belief in the participant that the blue aspect would be revealed in the outcome. From the agent's viewpoint, the red aspect was presented when the dog-robot made its last move behind the screens so the agent was induced to believe he would see a red aspect when the screens dropped. Similarly, expectations were consistent in the P-A- condition. Before its final move between the screens, the dog-robot spun to reveal its red aspect to the participant and its blue aspect to agent. As a result, neither the participant nor agent expected the eventual outcome. In the P+A- condition the agent was induced to believe that the blue aspect would be revealed as he saw the object's blue aspect enter the screens before he left the scene. In the agent's absence, the participant then saw the dog-robot re-emerge and spin to reveal its blue aspect to the participant before returning behind the screens (with its red aspect facing the agent). In this case, both the participant and the agent last saw the object's blue aspect, but the outcome only met the participant's expectation. Finally, in the P-A+ condition the agent expected the eventual outcome (blue-facing-participant) because he last saw the dog-robot's red aspect enter the screens. The participant, however, saw (in the agent's

absence) that the dog re-emerged, turned to present its red aspect to the participant, then retreated behind the screens. The events of the P-A+ condition induced the agent, but not the participant, to expect the outcome.

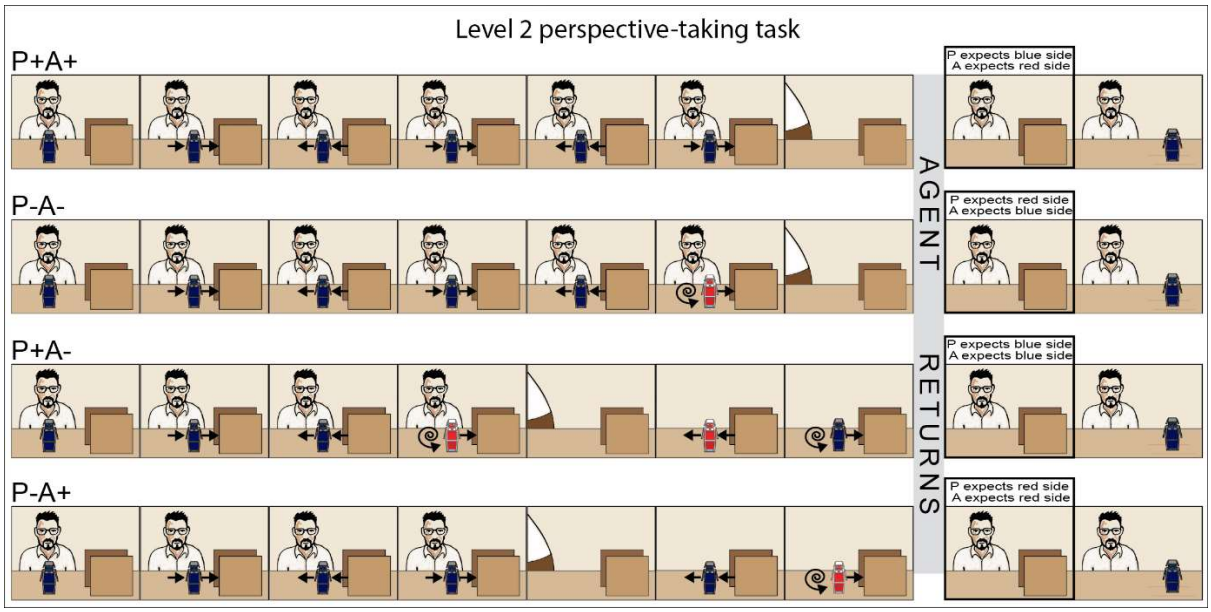

Figure 5-2 Experiment 3: Schematic storyboard showing the main belief-inducing events of the four conditions in the L2PT task movies.

The main belief-inducing events represent conditions in blue-facing-participant outcomes. In the P+A+ condition (consistent), the participant expects blue and the agent expects red; in the P-A- condition (consistent) the participant expects red and the agent expects blue; in the P+A- condition (inconsistent) both participant and agent expect blue; and in the P-A+ condition (inconsistent) both participant and agent expect red.

5.2.3. Procedure

Participants were tested in a room in which there were two stand-alone workstation cubicles, so that one or two adults could separately and privately undertake the experiment in a single session. Each person sat at a Dell Optiplex 9020 desktop with a 23" screen (16:9 aspect ratio). Participants were guided through each task via on-screen directions which explained the format of the test trials and provided the correct procedure for responding. The initial screen stated, "This is an object-detection task. Your job is to press a key as quickly as you can when you see something appear behind a wall". Task order was counterbalanced; the L1PT task instructions were as follows (the L2PT task instructions were identical except for the information in brackets): "In the first half of the experiment you will see 40 videos, lasting a total of about 10 (15) minutes. They will look like this (relevant frame of video

provided). In each video, the person will leave the scene, then return. Press the ‘Q’ key with your left hand as soon as the person has completely left the scene. When the walls disappear do one of the following with your right hand: Press the ‘N’ key if BLUE is revealed; Press the ‘M’ key if RED is revealed”. The outcome response buttons, both depressed by fingers of the right hand, were not counterbalanced.

Each trial consisted of an initial fixation cross (1000ms), then a short video. During each video, the participant had to make two responses: an attention check (pressing a key within 2000ms of the agent leaving the scene), and a colour detection (selecting blue or red when an object was revealed). The timings of each trial’s events differed by task and condition (see Figure 5-3 for timings of critical events in the L1PT and L2PT tasks). For each task, 40 test trials were presented in a pseudorandom order in two blocks. The first block contained 24 trials comprising three cycles of four different conditions with a red or a blue outcome. After a student-led break the participants experienced another block of 16 trials (two cycles of four different conditions with either a red or a blue outcome). Thus, over the two tasks, participants experienced 80 trials in total. A training phase exposed participants to four practice trials with feedback. These were undertaken before the experimental trials of each task. No performance feedback was given during the test phase to minimize trial time and distraction. The entire experiment took approximately 30 minutes in total. On completion of the experiment participants were asked to complete a form purportedly surveying their experience of how easy it was to sign up for laboratory experiments in exchange for partial course credits (e.g., “Have you found it easy to find suitable timeslots?”.“ The final question, “What was the experimenter testing?” sought to determine whether the participants were primed to consider the bystander’s belief. Although not a funneled debriefing protocol there was reason to be confident that mental state attribution was not deemed to be the target of the current research; all survey answers referred to the measuring of attention and/or reaction times in the pursuit of object detection.

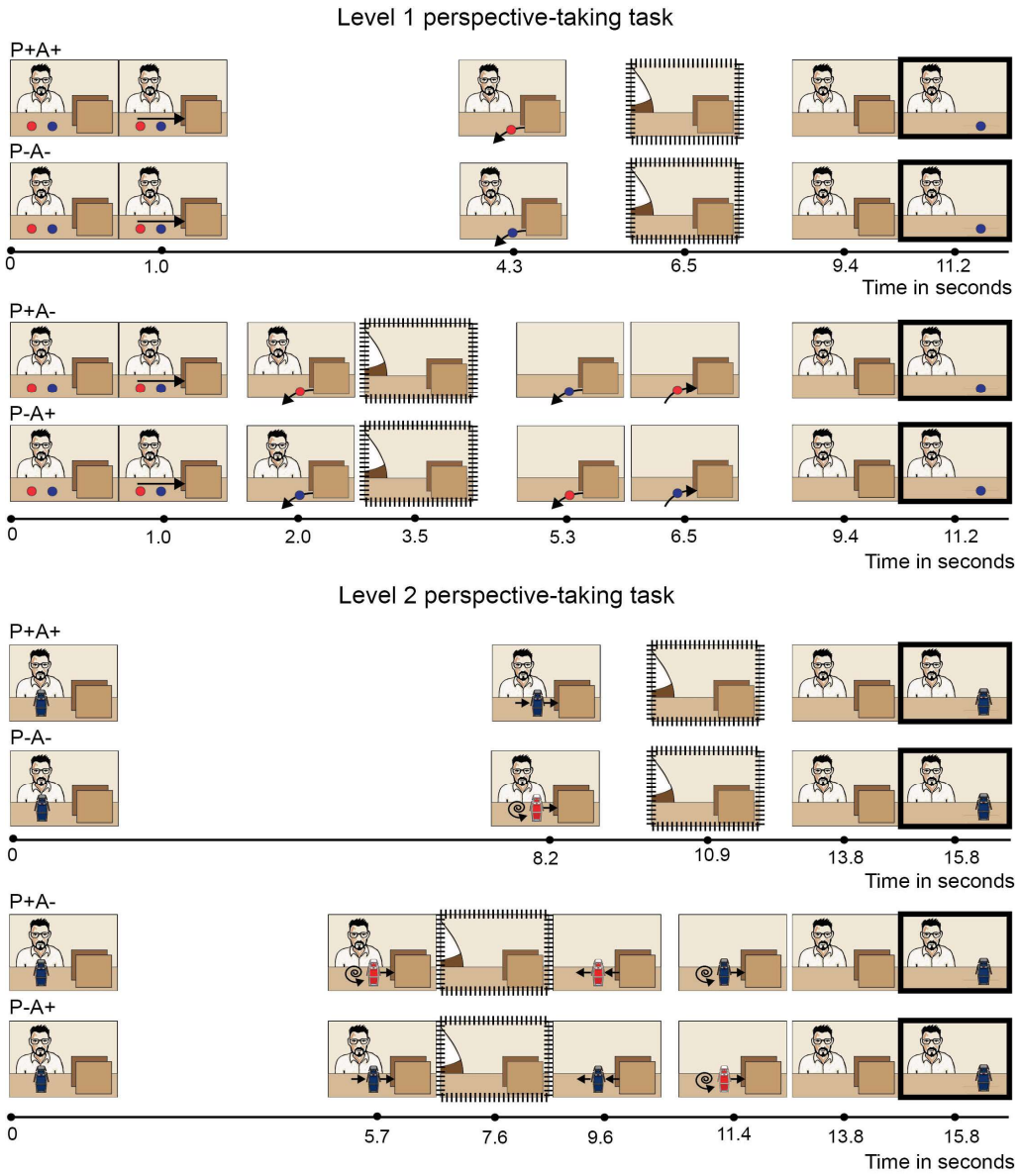

Figure 5-3 Experiment 3: Timings of the main events of the four conditions in the Level 1 perspective-taking and Level 2 perspective-taking tasks.

Two main events are highlighted: (1) the attention check triggered by the exit of the agent (hatched border); (2) removal of the screens (bold border).

5.3. *Results and discussion*

All statistical analyses were conducted with IBM SPSS Statistics 23 (SPSS Inc., Chicago, IL, USA). Analysis was undertaken on correct responses, defined as those in which the participant detected a colour that matched the revealed object. All statistical tests were two-tailed. Reaction times for trials in which participants failed to respond to an attention check were excluded (1.5% of trials). Following an outlier analysis, all data points greater

than 3 standard deviations above or below the participant's overall mean in each task was removed. As a result, 39 individual RTs were omitted (0.6% of individual responses in the L1PT task and 1.2% of individual responses in the L2PT task). Tests for normality revealed a positive skew in reaction times and error rates. A logarithmic transformation of the reaction time data was performed to fit the assumptions of an ANOVA before proceeding with further statistical analyses. As such, all means and standard deviations reported in the main text describe logarithmically transformed data. Mean response times for each condition are presented in Table 5-2 (transformed) and Table 5-3 (untransformed). The extent of the positive skew for the error data necessitated non-parametric testing (see Section 5.1.6). Greenhouse Geisser corrections were used whenever the assumption of sphericity was violated.

Table 5-2

Experiment 3: Logarithmically Transformed Mean Response Times

Task	Condition	*M*	*SD*
Level 1 perspective-taking	P+A+	2.50	.13
	P+A-	2.59	.90
	P-A+	2.64	.10
	P-A-	2.72	.08
Level 2 perspective-taking	P+A+	2.68	.10
	P+A-	2.68	.09
	P-A+	2.74	.08
	P-A-	2.74	.11

Note. N=53

Table 5-3

Experiment 3: Mean Response Times (in milliseconds)

Task	Condition	*M*	*SD*
Level 1 perspective-taking	P+A+	331.51	116.40
	P+A-	399.29	86.89
	P-A+	452.29	113.73
	P-A-	530.30	107.66
Level 2 perspective-taking	P+A+	489.20	123.32
	P+A-	490.30	109.74
	P-A+	560.27	121.57
	P-A-	572.32	165.52

Note. N=53

5.3.1. Response times

In keeping with Kovács et al.'s (2010) analyses, responding between conditions was initially compared. There was no theoretical basis to suggest that the colour of the target in the outcome phase (blue or red) would influence responding, so a 2 (Task: L1PT, L2PT) x 2 (Order: L1PT first, L2PT first) x 4 (Condition: P+A+, P+A-, P-A+, P-A-) mixed model ANOVA was performed. A main effect of Task, F (1, 51) = 215.00, $p < .001$, $\eta p^2 = .81$ was discovered; reaction times in the L1PT task (m = 2.61, sd = 0.13) were significantly faster than those in the L2PT task (m = 2.71, sd = 0.10). Planned comparisons between the corresponding conditions in each task (see Table 5-4 for an overview of analysis) revealed that reaction times were consistently slower in the L2PT task. A main effect of Condition, F (1, 51) = 149.17, $p < .001$, $\eta p^2 = .75$ was uncovered, but no main effect of Order (p =.423. There was no 3-way interaction (p = .482), but a two-way Task x Order interaction was found, F (1, 52) = 20.19, $p < .001$, $\eta p^2 = .28$. Post hoc independent samples t-tests found a single significant difference when comparing how participants performed in conditions depending on what order they completed the tasks; participants were faster in the P+A- condition if they completed the L1PT task first, t (52) = 2.17, $p < .036$, though this did not survive a Bonferroni correction. Finally, a two-way Task x Condition interaction was found, F (1, 52) = 50.06, $p < .001$, $\eta p^2 = .50$, which was explored further by task.

Table 5-4

Experiment 3: Planned Comparisons of Reaction Times between Tasks

	Paired Differences				
L1 - L2 Comparison	*m*	*sd*	*t*	*df*	*p*
P+A+	-.18	.09	-14.74	52	<.001*
P+A-	-.09	.08	-7.87	52	<.001*
P-A+	-.10	.06	-10.79	52	<.001*
P-A-	-.03	.08	-2.49	52	.016*

Notes. L1 = Level 1 perspective-taking task; L2 = Level 2 perspective-taking task; analyses undertaken on logarithmically transformed data; N=53; * indicates a significant effect.

L1PT task: A one-way ANOVA revealed that response times differed significantly between conditions, F (2.54, 131.86) = 173.93, $p < .001$, $\eta p^2 = .78$. This was explored by performing Bonferroni-corrected pairwise comparisons. The critical prediction was supported: response times were significantly faster in the P-A+ condition than in the P-A- condition, t (52) = 11.60, $p < .001$. Response times for the other conditions were then

compared (see Table 5-5 for an overview of pairwise comparisons). The pattern of responding is shown in Figure 5-4A: participants were fastest to respond in the P+A+ condition and slowest to respond in the P-A- condition; in addition, their reaction times in the P+A- condition were significantly faster than in the P-A+ condition. These findings suggest that, in the L1PT task, speed of response was modulated by both the participants' and the bystander's beliefs.

Table 5-5

Experiment 3: Overview of Pairwise Comparisons of Conditions

		Paired Differences			
Task	Comparison	*m*	*sd*	*t*	*p*
Level 1 perspective-taking	P-A- versus P-A+	.07	.05	11.60	<.001*
	P+A+ versus P+A-	.09	.08	8.58	<.001*
	P+A+ versus P-A-	.22	.07	21.29	<.001*
	P+A+ versus P-A+	.14	.07	14.27	<.001*
	P+A- versus P-A-	.12	.07	12.53	<.001*
	P+A- versus P-A+	.05	.08	4.82	<.001*
Level 2 perspective-taking	P-A- versus P-A+	.003	.07	.40	.689
	P+A+ versus P+A-	.003	.06	.44	.661
	P+A+ versus P-A-	.07	.07	7.03	<.001*
	P+A+ versus P-A+	.06	.05	8.31	<.001*
	P+A- versus P-A-	.06	.09	5.26	<.001*
	P+A- versus P-A+	.06	.06	7.02	<.001*

Notes. Planned comparisons in bold; * indicates a significant effect after Bonferroni correction; N=53.

L2PT task: Participants' reaction times differed per condition, as revealed by a one-way ANOVA, F (2.35, 122.19) = 31.32, $p < .001$, $\eta p^2 = .38$. Bonferroni-corrected pairwise comparisons showed that there was no difference between response times in the P-A+ and P-A- conditions ($p = .689$) supporting the primary hypothesis for this task. As illustrated in Figure 5-4B, the pattern of responding diverged from the L1PT task. In the L2PT task there was no difference between the P+A+ and P+A- conditions, and no difference between the P-A- and P-A+ conditions, suggesting that participants were not influenced by the bystander's belief. A statistical overview of the pairwise comparisons for each condition is provided in Table 5-5.

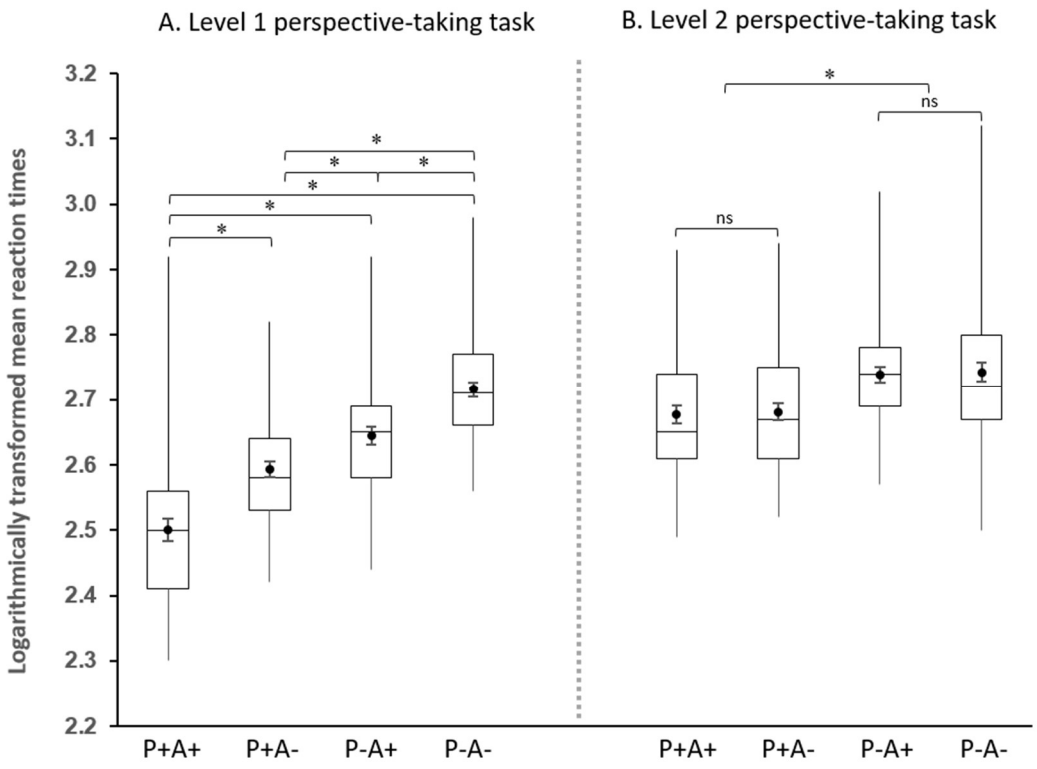

Figure 5-4 Experiment 3: Logarithmically transformed mean response times.

*Panel A shows box plots and logarithmically transformed mean response times for the four conditions in the L1PT task. Panel B shows box plots and logarithmically transformed response times in the L2PT task. Means are represented by dot markers; associated error bars represent the standard error of the mean. Note: * $p < .01$, two-tailed tests; N=53.*

An orthogonal analysis was also undertaken to explore the influence of the participant's belief and agent's belief. A 2 (Task: L1PT, L2PT) x 2 (Belief holder: P, A) x 2 (Belief: +, -) repeated measures ANOVA was performed. To do this the data was first organised to create four scenarios, P+, P-, A+ and A-. In P+ scenarios ([P+A+] + [P+A-])/2), participants were led to expect the outcome, whereas in P- scenarios ([P-A+] + [P-A-])/2), events were designed so that the participant did not expect the outcome. In A+ scenarios ([P+A+] + [P-A+])/2) the agent is led to expect the outcome, whereas in A- scenarios ([P+A-] + [P-A-])/2), the outcome is unexpected by the agent. Main effects of Task, F (1, 52) = 151.49, $p < .001$, $\eta p^2 = .74$, Belief holder, F (1, 52) = 23.15, $p < .001$, $\eta p^2 = .31$, and Belief, F (1, 52) = 366.29, $p < .001$, $\eta p^2 = .88$ were revealed. There was no three-way interaction ($p = .634$) but a Task x Belief-holder interaction was discovered, F (1, 52) = 6.41, $p = .014$, $\eta p^2 = .11$, and a Task x Belief interaction, F (1, 52) = 125.34, $p < .001$, $\eta p^2 = .71$, which were further investigated by task.

L1PT task: A 2 (Belief-holder: P, A) x 2 (Belief: +, -) repeated measures ANOVA revealed a main effect of Belief holder, F (1, 52) = 19.67, $p < .001$, $\eta p^2 = .28$, and a main effect of Belief, F (1, 52) = 477.35, $p < .001$, $\eta p^2 = .90$. However, these main effects were qualified by an interaction, F (1, 52) = 22.44, $p < .001$, $\eta p^2 = .30$. Overall, individuals were quicker to respond when outcomes were expected, compared to when they were not, but the effect of belief depended on the Belief holder. As depicted in Figure 5-5A, participants were faster to respond when the agent expected the outcome (A+; $m = 2.58$, $sd = .11$) compared to when the agent did not expect the outcome (A-; $m = 2.66$, $sd = .08$), and they were faster when they themselves expected the outcome (P+; $m = 2.55$, $sd = .10$) compared to when they did not (P-; $m = 2.68$, $sd = .09$), but the difference between expecting outcomes and not expecting outcomes was greater for the participant-held beliefs.

L2PT task: A 2 (Belief: P, A) x 2 (Belief holder: +, -) repeated measures ANOVA also found main effects of Belief holder, F (1, 52) = 9.60, $p = .003$, $\eta p^2 = .16$, and Belief, F (1, 52) = 47.91, $p < .001$, $\eta p^2 = .48$. Again, these main effects were qualified by an interaction, F (1, 52) = 53.18, $p < .001$, $\eta p^2 = .501$ (see Figure 5-5B). In this case, whilst individuals were faster to respond when they expected the outcome (P+; $m = 2.68$, $sd = .09$), compared to when they did not (P-; $m = 2.74$, $sd = .09$) scenarios), there was no significant difference in responding between scenarios in which the agent expected the outcome (A+; $m = 2.71$, $sd = .09$)) and those in which agent did not (A-; $m = 2.72$, $sd = .09$).

The finding that participants were faster in P+ compared to P- scenarios, often referred to as the reality bias (e.g., Bardi et al., 2018; Bardi, Six, & Brass, 2017; Deschrijver et al., 2016), suggests that participants were attending to each trial's events and using them to predict outcomes, rather than just waiting for the screens to drop to make their colour selection. Moreover, the reality bias was observed in both tasks. Comparing performances in the A+ and A- scenarios, it appears that there was only a facilitating influence of the agent's belief-like state in the L1PT task.

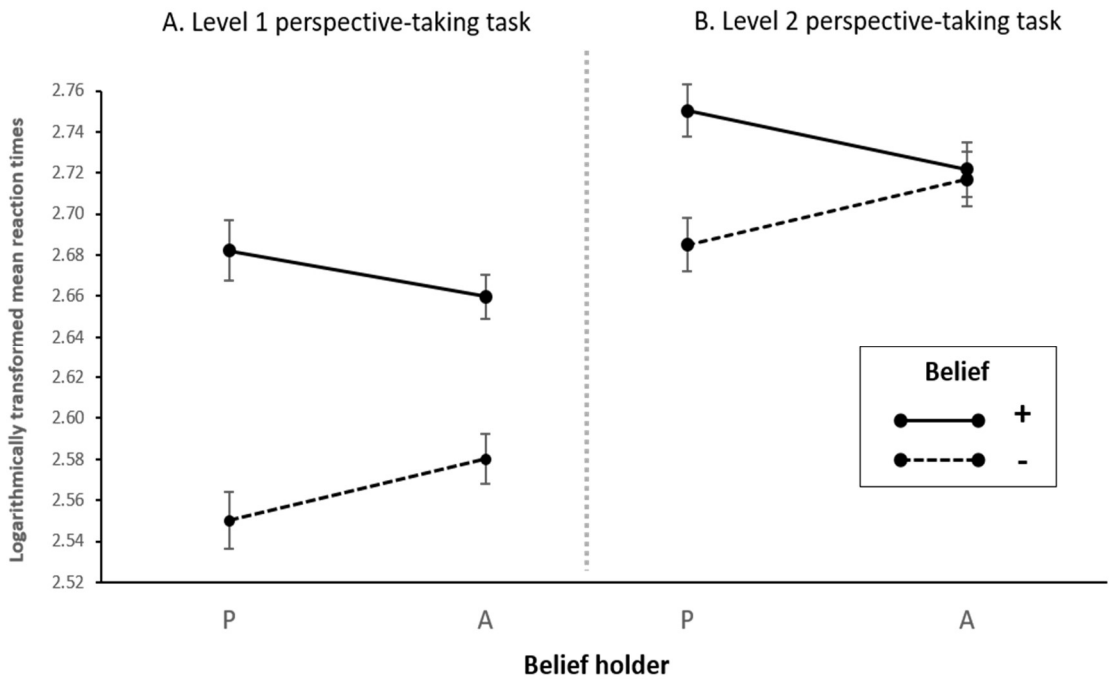

Figure 5-5 Experiment 3: Orthogonal analyses.

Panels A and B show the interactions between Belief-holder and Belief for the L1PT and L2PT tasks, respectively. Means are represented by dot markers; associated error bars represent the standard error of the mean. Note: N=53; 'P' = Participant; 'A' = Agent; '+' = Expected outcome; '-' = Unexpected outcome.

5.3.2. Errors

Overall, participants displayed high accuracy levels; the median error proportion was zero for each of the 16 trial types. The mean error proportions in the L1PT and L2PT tasks were .05 and .04 respectively (see Table 5-6 for mean error proportions and standard deviations for each condition and trial type). Mean error rates were analyzed using non-parametric tests as tests for normality revealed a large positive skew. After collapsing the colour of the outcome variable, a Friedman test revealed no difference in error proportions across the 8 conditions (the 4 conditions in each task), $\chi^2(7) = 7.87$, $p = .344$.

Table 5-6

Experiment 3: Mean Error Proportions

Task	Condition	*m*	*sd*
Level 1 perspective-taking	P+A+	.04	.07
	P+A-	.05	.07
	P-A+	.07	.09
	P-A-	.06	.06
Level 2 perspective-taking	P+A+	.04	.07
	P+A-	.05	.07
	P-A+	.06	.09
	P-A-	.07	.11

Note. N=53

5.4. *Summary*

To summarise, in keeping with Kovács et al.'s (2010) original study, not only were participants in the L1PT task faster to detect the outcome when *they* expected the outcome, they were also faster to detect the outcome when only the agent expected the outcome (P-A+, compared to P-A- condition). By contrast, in the L2PT task there was no facilitating influence of the agent's belief, indicating that his belief relativized to his visuospatial perspective about the outcome was not automatically processed. However, before drawing any strong conclusions regarding the implications of the data, it was necessary to address a potential limitation of the current study: perhaps the findings were an artifact of the particular methodology used (stimuli, materials, procedure) rather than a conceptual replication of the original object-detection study. To address this concern, we undertook another study in which we attempted to replicate Kovács and colleagues' findings with a single ball.

CHAPTER 6. *Experiment 4*

This chapter contains the methodology and results, written by Katheryn Edwards, from an experiment contained in the following manuscript accepted for publication in Cognition:

Edwards, K., & Low, J. (2019). Leve1 2 Perspective-taking distinguishes automatic and non-automatic belief-tracking. *Cognition, 193*. doi.org/10.1016/j.cognition.2019.104017

6.1. *Introduction*

Experiment 3's findings elaborate upon the dual-process account of human mindreading by suggesting that registration of perspective differences is likely to be eschewed by the efficient mindreading process. However, to be confident that the findings (that adults automatically track an agent's belief about which of two objects he is expecting to see) are a conceptual extension of classical findings from the original object-detection paradigm (and not a completely different phenomenon), Experiment 4 was undertaken to explore whether the findings could be replicated when participants had to detect the presence or absence of a single object.

6.2. *Methods*

6.2.1. Participants

Participants in Experiment 4 were 60 right-handed adults, 39 of which were students who participated in partial fulfilment of course requirements, and 21 who were adult volunteers who responded to an advert placed in a community playcentre. There were 38 females and 22 males, with an age mean of 21.88 years (Range 18 to 36). The study was approved by Victoria University of Wellington's Human Ethics Committee. The sample size of 60 participants was greater than the minimum number of participants required to detect Kovács, Téglás and Endress' (2010) critical effect, providing safeguards against potential procedural errors and/or absenteeism.

6.2.2. Materials

Stimuli and instructions were presented using E-Prime 2.0 using the same display parameters as Experiment 3. Each individual watched 40 short videos as part of an object-detection task. Each video was 10 seconds in length (after speeding the original footage by 120% in Adobe Premiere Pro). As in Experiment 3, the videos began with an agent seated at a table (on which were two screens) facing the participant. In contrast to the videos shown in Experiment 3, the to-be-detected object was now a single black ball. In the first movement, the ball moved between the two screens so that it could not be seen by either the participant or agent. Following this movement, the events in the videos varied to create four belief-induction conditions. These conditions differed according to whether the participant expected the ball to be present (P+) or absent (P-) in the outcome phase, and whether the agent

expected the ball to be present (A+) or absent (A-) in the outcome phase.

Expectations were induced by manipulating the movements of the ball and by varying the time that the agent left the scene (before or after critical events). The agent's return to the scene signalled the onset of the final phase. There were two possible outcomes in the final phase: the ball was either present or absent when the screens rapidly fell away. As such, participants experienced 8 trial types, comprised of four belief-induction conditions paired with one of two possible outcomes.

Events in the P+A+ condition led both the participant and the agent to expect the presence of the ball in the outcome phase. In the P-A- condition, both participant and agent were led to believe that the ball had left the scene. The P+A- and P-A+ conditions induced inconsistent expectations. In the P+A- condition, the participant and agent saw the ball leave the scene. However, the agent was absent when the ball returned to rest between the screens. In this case, the participant was led to expect the presence of the ball but agent was led to expect its absence. Finally, in the P-A+ condition both participant and agent witnessed the ball moving between the screens but only the participant saw the ball leave the scene. In the outcome phase, the agent's and participant's expectations were inconsistent; the participant expected the ball to be absent while the agent expected it to be present (Figure 6-1 for a schematic showing the main belief-inducing events of the four conditions).

6.2.3. Procedure

Participants were tested in the same room (with an identical arrangement) as in Experiment 3. Guidance regarding test format and response requirements was provided via on-screen prompts. Participants were instructed to detect the presence or absence of a single black ball. The initial screen stated, "This is an object-detection task. Your job is to press a key as quickly as you can when you see something appear behind a wall". Further instructions explained, "You will see 40 videos lasting a total of about 10 minutes. They will look like this (relevant frame of video provided). In each video, the person will leave the scene, then return. Press the 'Q' key with your left hand as soon as the person has completely left the scene. When the walls disappear do one of the following with your right hand: Press the 'N' key if the ball is present; Press the 'M' key if the ball is absent". The outcome response buttons were not counterbalanced.

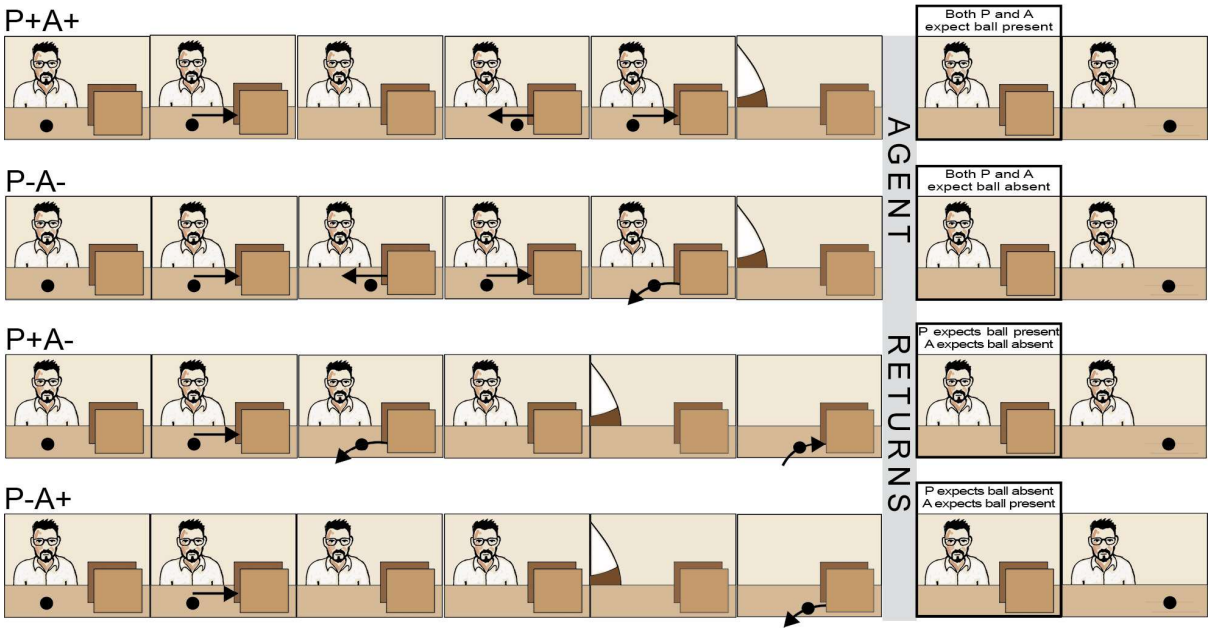

Figure 6-1 Experiment 4: Schematic showing the main belief-inducing events of the four conditions (ball-present outcome).

In the P+A+ condition, both participant and agent expected the ball to be present; in the P-A- condition neither participant or agent expected the ball to be present; in the P+A- condition only the participant expected the ball to be present; and in the P-A+ condition only the agent expected the ball to be present

As in Experiment 3 the timings of each trial's events differed by condition (see Figure 6-2 for timings of critical events). The 40 test trials were presented in a pseudorandom order in two blocks. The first block contained 24 trials comprising three cycles of 8 trials (4 conditions x 2 outcomes) and the second block contained 16 trials (two cycles of 8 trials). A training phase exposed participants to 4 practice trials with feedback. These were undertaken before the experimental trials. No performance feedback was given during the test phase to minimize trial time and distraction. The entire experiment took approximately 15 minutes in total. On completion of the experiment participants were asked to fill out a survey asking them about their experience taking part in the University's research programme. As in Experiment 3, all survey answers pertaining to the nature of the current task referred to attention and speed of object detection. Finally, participants were debriefed and their data collected.

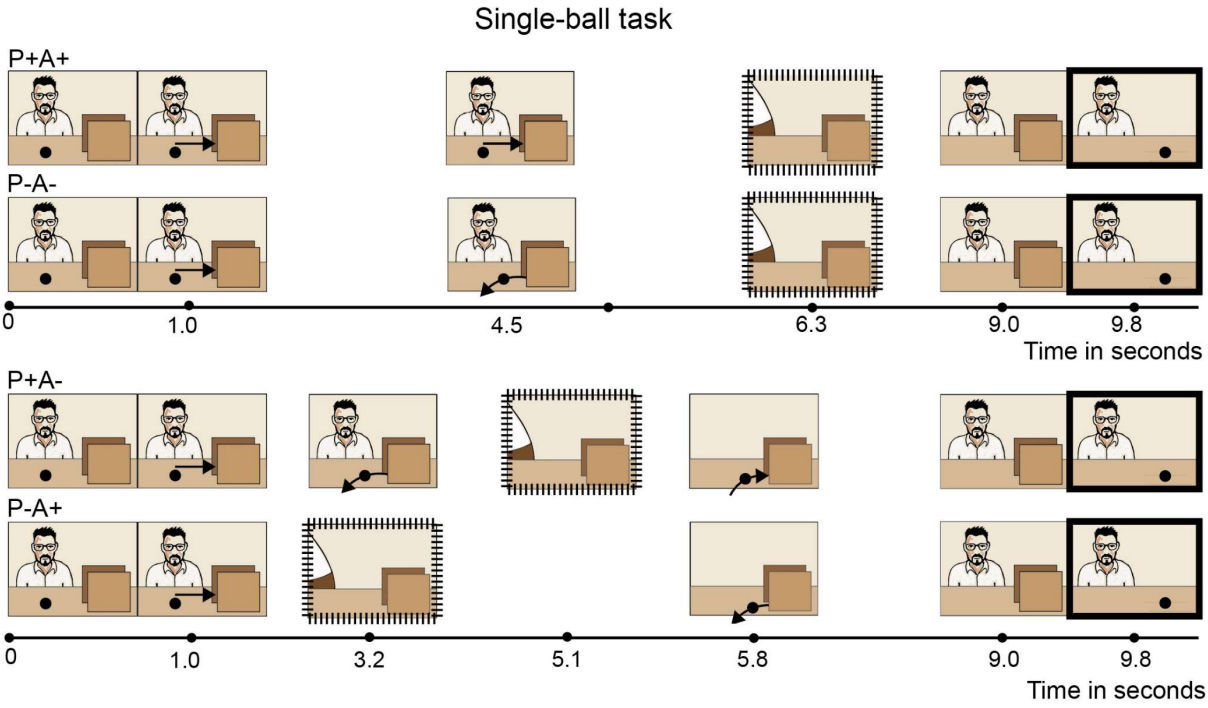

Figure 6-2 Experiment 4: Timings of the main events of the four conditions

The schematic highlights three main events: firstly, the attention check triggered by the exit of the agent (hatched border); secondly, removal of the screens, prompting colour selection (bold border); and thirdly, the last object movement witnessed by the participant and agent. For the P+A+ and P-A- conditions the participant and agent simultaneously witness the last movement so that their expectations are the same. In the P+A- and P-A+ conditions the last observed movement differs as the agent leaves the scene before the objects reach their final destination.

6.3. *Results and discussion*

All statistical analyses were conducted with IBM SPSS Statistics 23 (SPSS Inc., Chicago, IL, USA). Analysis was undertaken on correct responses, defined as those in which the participant accurately detected the presence or absence of the ball. All statistical tests were two-tailed. Error rates are reported separately below. Reaction times for trials in which participants failed to respond to the attention check were excluded (4.42% of trials). Following an outlier analysis, all data points greater than 3 standard deviations above or below the participant's overall mean in each task were also removed. As a result, 11 individual reaction times were omitted (0.45% of individual responses). Tests for normality revealed a positive skew in reaction times and error rates. A logarithmic transformation of reaction time data to fit the assumptions of an ANOVA was performed before proceeding with further statistical analysis. Transformed and untransformed means for response times are

presented in Tables 6-1 and 6-2, respectively. Due to the nature of the error data, analysis was conducted via non-parametric tests (see Section 6.3.2). Greenhouse Geisser corrections were used whenever the assumption of sphericity was violated.

Table 6-1

Experiment 4: Logarithmically Transformed Mean Response Times (N=60)

Outcome	Condition	m	sd
Ball-present	P+A+	2.60	.20
	P+A-	2.63	.14
	P-A+	2.72	.10
	P-A-	2.81	.10
Ball-absent	P+A+	2.80	.11
	P+A-	2.79	.11
	P-A+	2.73	.12
	P-A-	2.75	.11

Note. N=60

Table 6-2

Experiment 4: Mean response times (in milliseconds) (N=60)

Outcome	Condition	m	sd
Ball-present	P+A+	436.60	193.37
	P+A-	457.76	163.91
	P-A+	536.54	131.34
	P-A-	666.73	155.64
Ball-absent	P+A+	646.23	159.58
	P+A-	631.15	165.32
	P-A+	555.03	183.47
	P-A-	582.86	161.62

Note. N=60

6.3.1. Response times

A 2 (Outcome: ball-present, ball-absent) x 4 (Condition: P+A+, P+A-, P-A+, P-A-) repeated measures ANOVA was undertaken. Main effects of Outcome (F (1, 59) = 35.62, p < .001, ηp^2 = .63) and Condition, F (2.72, 160.03) = 27.58, p < .001, ηp^2 = .32, were revealed, and a significant Outcome x Condition interaction was confirmed, F (1.54, 91.21) = 35.62, p < .001, ηp^2 = .38. To interpret the interaction, a repeated measures ANOVA was performed

for each outcome. For the ball-present conditions the repeated measures ANOVA revealed a main effect of Condition, F (1.56, 91.94) = 47.32, $p < .001$, $\eta p^2 = .45$. Post hoc tests showed that the critical prediction was supported: response times were significantly faster when just the agent expected the ball to be present (P-A+), compared to when neither agent nor participant expected it to be present (P-A-), t (59) = 7.83, $p < .001$.

A statistical overview of the pairwise comparisons for all conditions in the ball-present and ball-absent trials is presented in Table 6-3. Participants were fastest to detect the presence of the ball when both the participant and agent expected it to be present (P+A+ condition), and slowest to detect the ball when neither the participant nor agent expected it to be present (P-A-). Lastly, participants were quicker to detect the ball when they, but not the agent believed it was present compared to when the agent, but not the participant expected it to be present (see Figure 6-3). These findings support the hypothesis that participants' reaction times are automatically influenced by the mere presence of others.

Table 6-3

Experiment 4: Overview of Pairwise Comparisons

		Paired Differences				
Outcome	Comparison	*m*	*sd*	*t*	*df*	*p*
Ball-present	P-A- *versus* P-A+	.10	.09	-7.83	59	<.001*
	P+A+ *versus* P+A-	-.04	.11	-2.83	59	.006*
	P+A+ *versus* P-A-	-.22	.20	-8.24	59	<.001*
	P+A+ *versus* P-A+	-.12	.20	-4.66	59	<.001*
	P+A- *versus* P-A-	-.18	.14	-10.00	59	<.001*
	P+A- *versus* P-A+	-.08	.13	-4.71	59	<.001*
Ball-absent	P-A- *versus* P-A+	.02	.11	1.81	59	.076
	P+A+ *versus* P+A-	.01	.10	.85	59	.398
	P+A+ *versus* P-A-	.05	.11	3.28	59	.002*
	P+A+ *versus* P-A+	.07	.14	3.95	59	<.001*
	P+A- *versus* P-A-	.04	.13	2.06	59	.044
	P+A- *versus* P-A+	.06	.12	4.00	59	<.001*

Note: P=Participant, A=Agent; planned comparisons are in bold; * indicates a significant effect after Bonferroni correction; N=60.

A repeated measures ANOVA also revealed a main effect of Condition for ball-absent trials, F (2.44, 144.01) = 9.13, $p < .001$, $\eta p^2 = .13$. Pairwise comparisons revealed no

difference between the baseline condition (P-A-), in which neither the participant nor agent was expecting the absence of the ball, and the condition in which only the agent expected there to be no ball present (P+A-). P+A+ and P+A- responding was significantly slower than P-A- and P-A+ responding, though the P+A- versus P-A- comparison did not survive the Bonferroni correction. There was also no difference between response times in the P-A- and P-A+ conditions (see Figure 6-3). As suggested by Kovács et al. (Supporting Information; 2014) this mixed response pattern is not surprising, as detecting the presence of an object is likely easier than detecting its absence.

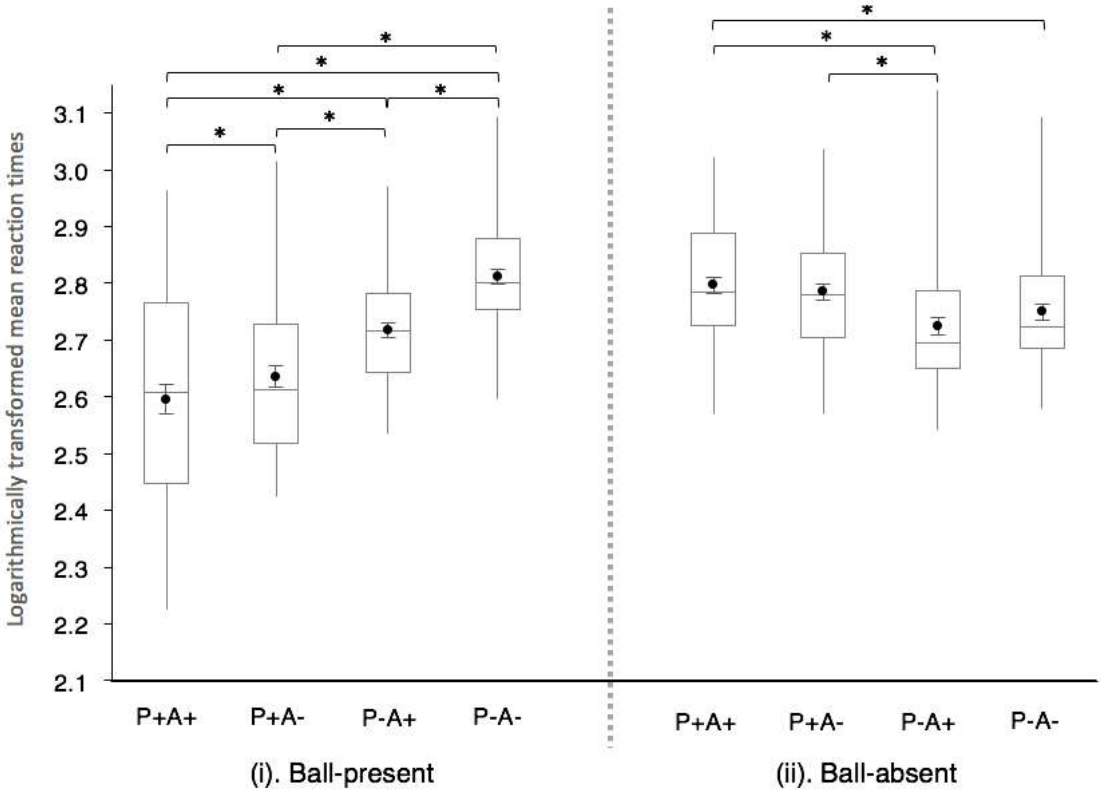

Figure 6-3 Experiment 4: Logarithmically transformed mean response times

*Box plots and logarithmically transformed mean response times based on ball-present (i) and ball-absent (ii) outcomes in the single ball task. Error bars represent the standard error of the mean. Means are represented by dot markers; associated error bars represent the standard error of the mean. Note: N= 60; * p < .01, two-tailed tests.*

Finally, an orthogonal analysis of the ball-present data was undertaken, with a 2 (Belief holder: P, A) x 2 (Belief; +, -) repeated measures ANOVA. A main effect for Belief holder was found, F (1, 59) = 22.35, $p < .001$, $\eta p^2 = .28$, and a main effect of Belief, F (1, 59) =

71.35, $p < .001$, $\eta p^2 = .55$. However, the main effects were qualified by a Belief holder x Belief interaction, F (1, 59) = 22.97, $p < .001$, $\eta p^2 = .28$. This was explained by an observation that the effect of belief was stronger for P scenarios compared to A scenarios, that is, the difference between P+ (m = 2.62, sd = .16) and P- (m = 2.77, sd = .08) responding was larger than that between A+ (m = 2.67, sd = .11) and A- (m = 2.74, sd = .10).

6.3.2. Errors

Participants showed a high level of accuracy, revealed by low mean error proportions in both the ball-present and ball-absent conditions (.06 and .05 respectively; see Table 6-4 for error proportions in each condition). Tests for normality revealed that the error data was positively skewed. A Friedman test revealed no statistically significant differences in mean error proportions across the 8 trial types, $\chi^2(7) = 1.86$, $p = .967$.

Table 6-4

Experiment 4: Mean Error Proportions

Outcome	Condition	*m*	*sd*
Ball-present	P+A+	.060	.11
	P+A-	.050	.11
	P-A+	.070	.14
	P-A-	.060	.12
Ball-absent	P+A+	.053	.11
	P+A-	.060	.11
	P-A+	.057	.11
	P-A-	.047	.11

Note. N=60

6.4. *Summary*

To conclude, when expecting the ball to be present, responding is fastest when both the participants' and agents' beliefs match the outcome, and slowest when neither are induced to expect the outcome. In keeping with the theoretical basis for the study, not only are participants faster than the baseline condition (P-A-) to detect the ball when they, but not the agent, expect the outcome, they are also speeded when only the agent expects the ball to be present. Nonetheless it is possible that the critical effect was an artifact of the attention check used in the classic object-detection task. This possibility is addressed in Chapter 7 (Experiment 5).

CHAPTER 7. *Experiment 5*

This chapter contains the methodology and results, written by Katheryn Edwards, from an experiment contained in the following manuscript accepted for publication in Cognition:

Edwards, K., & Low, J. (2019). Leve1 2 Perspective-taking distinguishes automatic and non-automatic belief-tracking. Cognition 193. doi.org/10.1016/j.cognition.2019.104017

7.1. *Introduction*

In Chapter 6 the current study's novel object-detection paradigm was seen to provide a conceptual replication of Kovács et al.'s (2010) findings. However, while Kovács and colleagues' task has been considered a useful tool for investigating the automaticity of mindreading, its validity has been questioned. Phillips et al. (2015) argue that the critical effect is driven by timing variations in the attention check (often referred to as the attention-check hypothesis). If so, could the automatic belief-tracking suggested in Experiments 3 and 4 merely reflect such methodological inconsistencies? Perhaps, in the L1PT task, adults are significantly slower to detect the correct colour in the P-A- than in the P-A+ condition because there is a shorter duration between the attention check (which requires the participant to press a button when the agent leaves the scene) in the P-A- condition than in the P-A+ condition. In other words, a shorter stimulus onset asynchrony (SOA) in the P-A- condition than in the P-A+ condition leads to more protracted response times in the former. There are a number of reasons to challenge the attention-check hypothesis as an explanation of the current findings. First, it has been contested (e.g., Nijhof et al., 2016, 2017) on the grounds that the influence of a short SOA on the reaction time to a second stimulus (known as psychological refractory period) is a short-term effect, and only observable at SOAs up to several hundred milliseconds. The shortest SOAs found in the typical object-detection paradigm tend to be over 2,000 milliseconds, and the shortest time between the attention check and detection response in the current paper (> 4000ms) is substantially longer than refractory periods discussed in past literature. Second, in Experiment 3's L1PT task, there was consistently faster responding in the P+A+ condition than in the P+A- condition, even though the former condition had a shorter SOA. Third, in Experiment 3 adults were not faster to respond in the P-A+ condition than in the P-A- condition of the L2PT task, which would *not* be predicted if the key difference between those conditions was merely the result of a shorter SOA. Nonetheless, it may be argued that some factor associated with tracking a rotating object may have interfered with a potential attention-check effect. To fully mitigate concerns over differences in refractory periods across trial types, a second replication of Experiment 3 was attempted, removing the attention checks from each condition.

7.2. *Method*

7.2.1. Participants

A total of 108 right-handed psychology students volunteered in partial fulfilment of course requirements. There were 82 females and 26 males, with an age mean of 18.92 years (Range 17 to 34 years). The study was approved by Victoria University of Wellington's Human Ethics Committee. A greater number of individuals in this study was recruited due to an increase in the availability of students in Victoria University of Wellington's IPRP and because there was reason to be concerned that the removal of the attention check could result in a greater number of participants failing to meet the accuracy threshold of 75%.

7.2.2. Materials and procedure

The materials and procedure were identical to Experiment 3, except that there was no requirement for the participants to respond (by pressing the Q key) when the agent left the scene.

7.3. *Results and discussion*

All statistical analyses were conducted with IBM SPSS Statistics 23 (SPSS Inc., Chicago, IL, USA). Analysis was undertaken on correct responses, defined as those in which the participant detected a colour that matched the revealed object. Five participants were excluded from analysis as their performances were below the 75% accuracy threshold across all trials. Of the 103 remaining participants there were 79 females and 24 males, with a mean age of 18.8 years (range 17 to 34). All individual data points greater than 3 standard deviations above or below the participant's overall mean in each task were removed. As a result, 103 individual reaction times were omitted (0.6% of individual responses in the L1PT task and 0.6% of individual responses in the L2PT task). All statistical tests were two-tailed. Tests for normality revealed a positive skew in reaction times and error rates. In order to proceed with further statistical analysis a logarithmic transformation of the reaction time data was undertaken to fit the assumptions of an ANOVA. Mean response times are presented in Table 7-1 (transformed) and Table 7-2 (untransformed). Error rates were compared across conditions using non-parametric tests (see Section 7.3.2). Greenhouse Geisser corrections were used whenever the assumption of sphericity was violated.

Table 7-1

Experiment 5: Logarithmically Transformed Mean Response Times

Task	Condition	*m*	*sd*
Level 1 perspective-taking	P+A+	2.56	.06
	P+A-	2.59	.07
	P-A+	2.60	.06
	P-A-	2.64	.07
Level 2 perspective-taking	P+A+	2.59	.06
	P+A-	2.60	.06
	P-A+	2.65	.06
	P-A-	2.66	.07

Note. N=103

Table 7-2

Experiment 5: Mean Response Times (in milliseconds) (N=103)

Task	Condition	*m*	*sd*
Level 1 perspective-taking	P+A+	364.47	47.20
	P+A-	389.89	59.18
	P-A+	402.56	56.44
	P-A-	447.75	65.23
Level 2 perspective-taking	P+A+	389.34	55.01
	P+A-	399.60	58.80
	P-A+	450.05	68.18
	P-A-	458.03	70.50

Note. N=103

7.3.1. Response times

Informed by previous research, a 2 (Task: L1PT, L2PT) x 2 (Order: L1PT first, L2PT first) x 4 (Condition: P+A+, P+A-, P-A+, P-A-) mixed model ANOVA was performed. There was no three-way interaction ($p = .597$), Task x Order interaction ($p = .311$), Condition x Order interaction ($p = .876$), or main effect of Order ($p = .556$). However, there was main effect of Task, F (1, 101) = 30.68, $p < .001$, $\eta p^2 = .23$; the mean reaction time in the L1PT task ($m = 2.60$, $sd = .07$) was smaller than that of the L2PT task ($m = 2.62$, $sd = .07$). There was also a main effect of Condition, F (2.67, 269.75) = 144.45, $p < .001$, $\eta p^2 = .57$, and a two-way Task x Condition interaction, F (2.71, 273.58) = 10.78, $p < .001$, $\eta p^2 = .10$, which was explored further after separating the data by task.

L1PT task: A repeated measures ANOVA showed that performance significantly differed across conditions, F (2.91, 296.78) = 85.26, $p < .001$, $\eta p^2 = .45$. Supporting the critical prediction, it was determined that response times were significantly faster in the P-A+ condition than in the P-A- condition, t (102) = 8.05, $p < .001$. Bonferroni-corrected pairwise comparisons between the other L1PT conditions (see Table 7-3 for an overview) provided a pattern of findings that is illustrated in Figure 7-1A. Fastest responding was found in the P+A+ condition and slowest responding in the P-A- condition, but there was no significant difference between the P+A- and P-A+ conditions. These findings indicate that speed of response was modulated by both the participants' and the bystander's beliefs.

Table 7-3 Exp 5: Overview

Experiment 5: Overview of Pairwise Comparisons

		Paired Differences			
Task	Comparison	*m*	*sd*	*t*	*p*
Level 1 perspective-taking	P-A- *versus* P-A+	.05	.06	8.05	<.001*
	P+A+ *versus* P+A-	.03	.05	5.31	<.001*
	P+A+ *versus* P-A-	.09	.06	15.96	<.001*
	P+A+ *versus* P-A+	.04	.06	7.33	<.001*
	P+A- *versus* P-A-	.06	.05	11.22	<.001*
	P+A- *versus* P-A+	.01	.06	2.33	.132
Level 2 perspective-taking	P-A- *versus* P-A+	.008	.05	1.54	.756
	P+A+ *versus* P+A-	.01	.04	2.52	.078
	P+A+ *versus* P-A-	.07	.07	10.73	<.001*
	P+A+ *versus* P-A+	.06	.06	10.06	<.001*
	P+A- *versus* P-A-	.06	.06	10.53	<.000*
	P+A- *versus* P-A+	.05	.06	8.81	<.001*

Notes: Critical comparisons in bold; * indicates a significant effect after Bonferroni correction; *df* = 102, N=103.

L2PT task: Reaction times differed between conditions, as revealed by a repeated measures ANOVA, F (2.43, 247.59) = 79.71, $p < .001$, $\eta p^2 = .44$. Focusing on the critical conditions, support was found for the primary L2PT hypothesis: there was no difference between response times in the P-A+ and P-A- conditions ($p = .75$). As depicted in Figure 7-1B, the pattern of responding diverged from the L1PT task. In the L2PT task there was no difference between the P+A+ and P+A- conditions, and no difference between the P-A- and P-A+ conditions, indicating that participants were not influenced by the bystander's belief. A statistical overview of the pairwise comparisons for each condition is provided in Table 7-3.

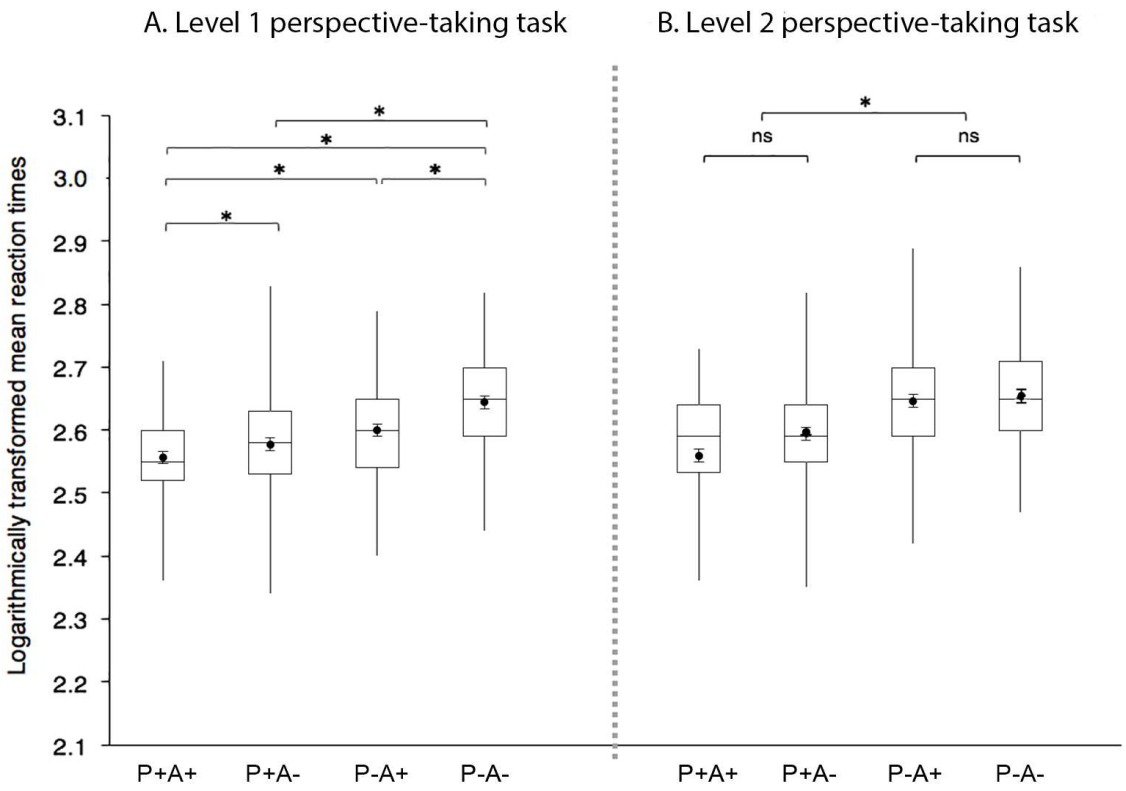

Figure 7-1 Experiment 5: Logarithmically transformed mean response times.

*Panel A shows box plots and logarithmically transformed mean response times for the four conditions in the L1PT task. Panel B shows box plots and logarithmically transformed response times in the L2PT task. Means are represented by dot markers; associated error bars represent the standard error of the mean. Note: N=103; * p < .01, two-tailed tests.*

As in Experiments 3 and 4, orthogonal analyses were conducted to examine the influence of participants' and agent's beliefs. A 2 (Task: L1PT, L2PT) x 2 (Belief holder: P, A) x 2 (Belief: +, -) repeated measures ANOVA revealed main effects of Task, F (1, 102) = 35.66, $p < .001$, $\eta p^2 = .26$, Belief holder, F (1, 102) = 7.09, $p = .009$, $\eta p^2 = .07$, and Belief, F (1, 102) = 259.79, $p < .001$, $\eta p^2 = .72$. There was no Task x Belief interaction ($p = .066$), but a three-way interaction was found, F (1, 102) = 32.80, $p < .001$, $\eta p^2 = .24$, and two-way interactions between Task and Belief holder, F (1, 102) = 5.52, $p = .021$, $\eta p^2 = .05$ and between Belief holder and Belief, F (1, 102) = 62.70, $p < .001$, $\eta p^2 = .38$. These were investigated further by task.

L1PT task: A 2 (Belief holder: P, A) x 2 (Belief: +, -) repeated measures ANOVA uncovered a main effect of Belief holder, F (1, 102) = 5.77, $p = .020$, $\eta p^2 = .05$, and a main effect of Belief, F (1, 102) = 261.84, $p < .001$, $\eta p^2 = .72$. However, these findings were qualified by an interaction, F (1, 102) = 5.69, $p = .019$, $\eta p^2 = .05$. Replicating Experiment 3's

findings, it was determined that, overall, individuals were quicker to respond when beliefs contained an expectation of the outcome (+), compared to when they did not (-), however the effect of Belief depended on the Belief holder. As illustrated in Figure 7-2A, the response differential between P+ scenarios (m = 2.57, sd = .06) and P- scenarios (m = 2.63, sd = .06), was greater than the response differential between A+ (m = 2.58, sd = .05) and A- scenarios (m = 2.62, sd = .06).

L2PT task: A 2 (Belief holder: P, A) x 2 (Belief: +, -) repeated measures ANOVA revealed a main effect of Belief holder, F (1, 102) = 6.37, $p < .013$, $\eta p^2 = .06$, and a main effect of Belief, F (1, 102) = 107.12, $p < .001$, $\eta p^2 = .05$. Again, these main effects were qualified by an interaction, F (1, 102) = 106.10, $p < .001$, $\eta p^2 = .05$, which is depicted in Figure 7-2B. It was observed that individuals were faster to respond in P+ scenarios (m = 2.59, sd = .06) compared to P- scenarios (m = 2.65, sd = .06), but there was no significant difference in responding for A+ (m = 2.62, sd = .05) versus A- (m = 2.63, sd = .06) scenarios.

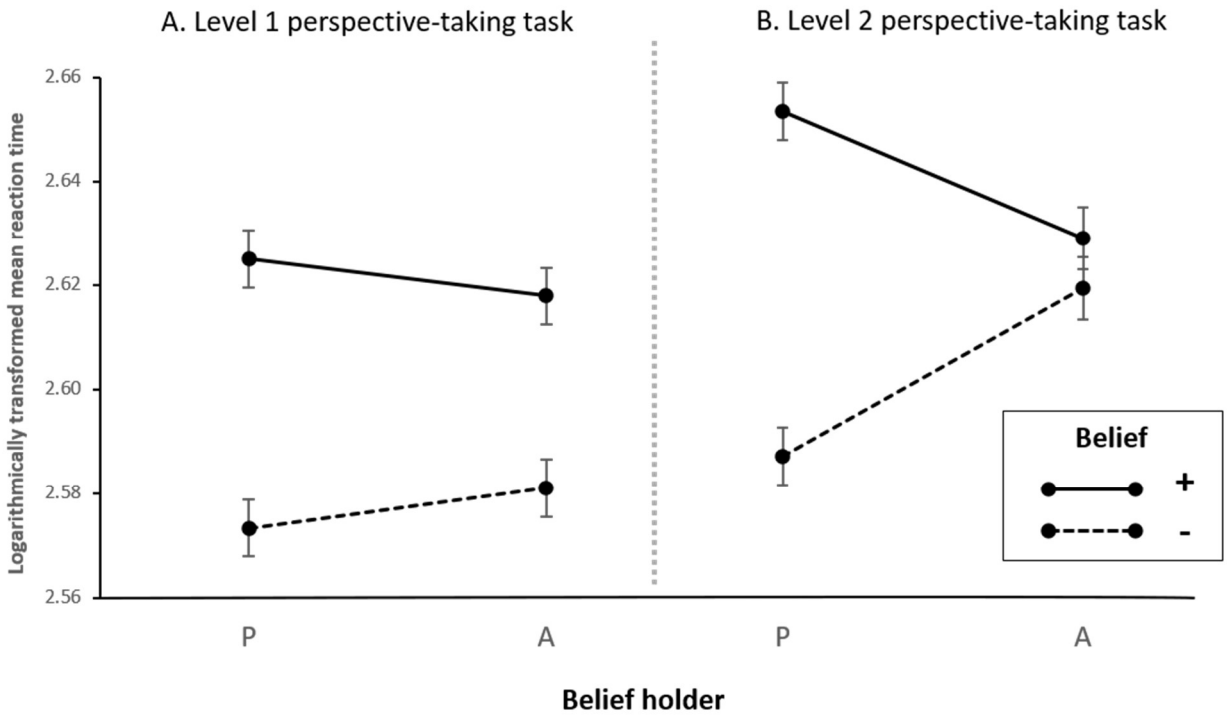

Figure 7-2 Experiment 5: Orthogonal analyses.

Panels A and B show the interactions between Belief-holder and Belief for the L1PT and L2PT tasks, respectively. Means are represented by dot markers; associated error bars represent the standard error of the mean. Note: 'P' = Participant; 'A' = Agent; '+' = Expected outcome; '-' = Unexpected outcome. Note: N=103.

Replicating Experiments 3 and 4, these findings suggest the presence of a reality bias (Bardi et al., 2019, 2017; Deschrijver et al., 2016) in both tasks, inferring that participants do use their own beliefs about the position and/or orientation of the object/s when detecting the colour outcome. However, it seems that the agent's beliefs are only taken into account in the L1PT task which does not involve contrasting perspectives.

7.3.2. Errors

Overall, participants displayed high accuracy levels; the median error proportion was zero for each of the 16 trial types. The mean error proportions in the L1PT and L2PT tasks were .05 and .04 respectively (see Table 7-4 for mean error proportions and standard deviations for each condition). Mean error proportions were analyzed using non-parametric tests as tests for normality revealed a large positive skew. A Friedman test revealed no significant difference in mean error proportions across the 8 conditions (4 conditions in each task), $\chi^2(7) = 13.32$, $p = .065$.

Table 7-4

Experiment 5: Mean Error Proportions

Task	Condition	*m*	*sd*
Level 1 perspective-taking	P+A+	.04	.06
	P+A-	.05	.07
	P-A+	.06	.07
	P-A-	.03	.05
Level 2 perspective-taking	P+A+	.04	.07
	P+A-	.04	.06
	P-A+	.04	.06
	P-A-	.05	.08

Note. N=103.

7.4. *Summary*

The pattern of reaction-times in Experiment 5 emulates that of Experiment 3, even in the absence of the attention check. In the L1PT task not only are participants faster to detect the outcome when they *expect* the outcome, they are also faster to detect the outcome when only the agent expects the outcome. By contrast, in the L2PT task there is no facilitating influence of the agent, indicating that his belief about the outcome is not automatically

processed in this instance. A post hoc power analysis determined that the study had 99.99% power to calculate the critical effect with the current sample size. Removing the attention check did not have any impact on participants' accuracy compared to Experiment 3, implying that this procedural change did not adversely affect engagement with either task.

To conclude, when adults' efficient processing was examined in a Level 1 perspective-taking context (where an agent's belief, but not his visuospatial perspective, was relevantly different) and in a Level 2 perspective-taking context (where both the agent's belief and visuospatial perspective are relevantly different), evidence was found to suggest that automatic mindreading draws upon a distinctively minimalist model of the mental that underspecifies representation of differences in perspective relative to an agent's position in space. The following chapter considers the implication of these findings a for a dual-process account of human mindreading.

CHAPTER 8. *General Discussion*

8.1. *Introduction*

The overarching aim of this thesis was to test the dual-process account of human mindreading. The chosen approach was to devise two empirical techniques to inform whether the dual-process account offers fitting explanations for two prominent puzzles in the theory of mind field. The first puzzle concerns developmental evidence: conflicting findings from direct and indirect false-belief tasks suggests that there is a major conceptual breakthrough in belief reasoning sometime around children's 4th birthday, but also that infants should be credited with abstract belief understanding. The second puzzle concerns mature mindreading characteristics: there is little consensus as to why adults are sometime fast and sometimes slow to reason about others' mental states.

According to the dual-process account the two puzzles can be explained by drawing upon two relatively distinct processes. Efficient mindreading allows for speedy mental state processing and guides rapid. non-verbal responding in infants, children and adults. The later developing flexible mindreading permits effortful mental state calculations and supports success in verbal false-belief tests. The current research exploits the theory that efficient mindreading exhibits 'signature limits' to the type of information that can be processed. Informed by previous research, two empirical chapters (Chapters 2 and 3) report findings of a novel procedure designed to test for signature limits regarding numerical identity. Three further empirical chapters (Chapters 5, 6 and 7) document an attempt to advance the current understanding of signature limits by designing a second novel paradigm in which numerical identity cannot be used to explain response patterns. This present chapter reviews the main empirical findings, discusses the implications for a dual-process account of mindreading and considers alternative explanations and limitations of the current approach. Finally, some suggestions for future research are posited, before the chapter closes with the author's conclusory remarks.

8.2. *Summary of findings*

Using an action-prediction paradigm, two experiments tested the conjecture that representations underpinning efficient belief-tracking relate agents to objects, leading to the prediction that efficient processing cannot handle false-beliefs involving identity. More specifically, Experiments 1 and 2 tested the hypotheses that efficient, but inflexible, mindreading may give rise to appropriate reaction time facilitation in a standard unexpected-

transfer task, but not in a task involving an identity component. Marrying ideas from a looking time study (Low & Watts, 2013) with an electroencephalogram study (Southgate & Vernetti, 2014), a new behavioural paradigm was created in which adults had to quickly select whether an actor would reach, or not reach, for an object based on an actor's false belief about an object's location. In the object-location scenario, the object was either a red or a blue ball that looked the same from all viewpoints, whereas in the object-identity scenario the object was a toy that appeared red from one perspective and blue from another. While adults' button selections were accurate and reflected a high level of flexible processing across both tasks, reaction times showed distinct profiles in rapid efficient processing. In the object-location scenario, participants' response times were appropriately facilitated when they accurately predicted that an agent would reach for a desired ball based on her false belief that it was in the box. In stark contrast, although participants in the object-identity task were very accurate in predicting that the agent would reach for a toy (based on her false belief that a preferred toy was in the box) their response times were not suitably speeded. The (relatively) automatic operation of the efficient mindreading system seen in the object-location task was not evident in the task incorporating an identity component. Experiments 1 and 2 show for the first time, that there are indeed different profiles of reaction times for object-location scenarios and for object-identity scenarios. In sum, while the overall accuracy reflected a high level of flexible belief reasoning across both tasks, the pattern of response times across conditions revealed a limit in the processing scope of efficient processing, that is, efficient mindreading eschews an understanding of others' false beliefs about numerical identity.

In a second set of experiments a novel object-detection paradigm was used to test the extent to which participants automatically tracked the beliefs of a passive bystander in two closely-matched but conceptually distinct tasks. Experiments 3 and 5 report that, in an L1PT object-detection task involving homogenous objects, adults' reaction times were involuntarily influenced by the presence of a passive bystander. Participants were faster to detect the colour of an object when the agent, but not the participant (P-A+), expected the outcome, compared to a baseline condition in which neither expected the outcome (P-A-). By contrast, in an L2PT task, the presence of the agent did not influence adults' response times when the to-be-detected object could be differently perceived depending on where the agent was located in relation to that object. In this scenario, reaction times for the pairs of conditions in which a participant expected a certain colour to be revealed (P+A+, P+A-) were significantly faster than the pair of conditions (P-A-, P-A+) in which the participant did not expect a

certain colour to be revealed. The pattern of responding in the L2PT task indicated that reaction times were contingent on participants' expectations only. In Experiment 4, the critical effect of automatic belief-tracking was replicated when only one homogenous object was used and the agent's perspective was not relevantly different. Experiment 5 sought to rule out the possibility that response times in the object-detection paradigm may be influenced by differences in the timings of the attention checks (the requirement for the participant to respond when the agent left the scene) across conditions. Using the same procedure and materials as in Experiment 3 it was found that the overall pattern of responding was not affected when the attention check was removed. These findings were also supported by an orthogonal analysis investigating the influence of participants' own beliefs (P+, P-) and the belief of the agent (A+, A-). Overall, the conclusion is that adults automatically track others' beliefs concerning where an object is located but not their beliefs regarding how an object is perceived from a certain perspective.

8.3. *Alternative explanations and limitations*

Heretofore it has been proposed that the findings from Experiments 1 to 5 support a dual-process account of mindreading. However, the current thesis should also consider how the findings sit with the possibility that human beings only have a single mindreading system that is sufficiently sophisticated to also enable speedy calculations of wide-ranging mental contents. Carruthers (2015, 2016) describes a high-level, one-system account in which infants start out with core mental-state concepts (e.g., thinks, likes) which allows them to attribute meta-representational states to others (e.g., "Daddy thinks there is a toy in my box") without an explicit understanding of what is true and false. Mastery of these concepts occurs in time, but there is no change to the fundamental architecture of the representational mechanism from infant to adult. According to such an account, observations that adults are sometime fast and sometimes slow to reason about others' mental states are explained by appealing to context. Thus, tasks that do not tax executive function are processed speedily by adults, using the same mindreading processes as used by young children. More demanding belief-reasoning tasks require adults to draw upon further cognitive resources resulting in more protracted reasoning.

Applying Carruther's (2015, 2016) reasoning to the current findings, a single-system proponent may claim that the differing response patterns in Experiments 1, 2, 3 and 5 are explained by extraneous task variables. Specifically, the Identity tasks (Experiments 1 & 2)

and Level 2 tasks (Experiments 3, 4 & 5) may have imposed unnecessary demands on rotation skills that mask the expression of sophisticated mindreading that is both flexible and efficient. Testing signature limits often relies on documenting that efficient mindreading supports tracking others' mental ascriptions of where an object is located or what is or is not perceptible to someone, but not how someone identifies an object from a different perspective or viewing angle (Low et al., 2014; Low & Watts, 2013; Mozuraitis et al., 2015; Surtees et al., 2012; Surtees, Samson, & Apperly, 2016). Carruthers (2015, 2016) suggests that there is a confound because different performances between object-location/L1PT and object-identity/L2PT tasks might be instead due to non-mental content; identity/L2PT tasks often involve spatial rotation to represent another's perspective.

In the current body of work, great efforts were made to match the cognitive structure of the false-belief (Location versus Identity) tasks in the action-anticipation paradigm, and visual perspective-taking (Level 1 versus Level 2) tasks in the object-detection paradigm. Participants in the Experiments 1 and 2 did make more errors in the Identity trials, but overall their accuracy was very high in both tasks. The claim from the single-system viewpoint, that tasks involving a dual-aspect object place great demands on executive functioning, do not explain the dissociation of performances across the two tasks discovered in this research. Regarding the object-detection paradigm, a single-system advocate may suggest that the passive bystander of Experiments 3 and 5 might not have influenced adults' own responses in the L2PT task due to extraneous demands associated with the chosen stimuli. In response, it must be noted that the LPT1 and LPT2 tasks were well-matched in terms of overall difficulty: one task involved tracking a single object with two distinct sides, and the other involved tracking two distinct objects. Moreover, it was revealed that adults were influenced automatically by the agent's beliefs in the L1PT two-ball task and in the L1PT one-ball task, even when tracking beliefs about the path of two distinct objects is more cognitively demanding than tracking beliefs about the paths of one distinct object (Horowitz & Cohen, 2010). One possible method for reducing mental rotation demands is outlined in Section 8.5, but it is important to note here that signature limits on efficient mindreading have been documented with identity tasks that do not involve rotation (e.g., Fizke et al., 2013). Furthermore, performance on independent measures of mental rotation ability are not correlated with the appreciation of how an object that is simultaneously visible to the self and other can give rise to different representations of identity (Hamilton et al., 2009).

Another important factor to consider is visual saliency. It is difficult to quantify the extent to which the current findings reflect the saliency of the self-perspective in the object-identity (action-prediction paradigm) and L2PT (object-detection paradigm) tasks. Perhaps reducing the visual saliency of the self-perspective would produce differing patterns of results. That said, consider Keysar et al.'s (2003) referential communication task where adults had to move objects about a vertical grid according to the directions of a confederate. Participants were aware that only some of the objects were mutually visible, but when asked to "move the small candle" they often mistakenly attended to the smallest candle from their privileged perspective, rather than to the smallest candle experienced from the confederate's perspective. Keysar and colleagues went to great lengths to reduce the saliency of self-perspective by inviting the participant to set up the array of objects on the grid from the confederate's point of view, making it clear which objects could be seen and which could not. Despite these steps, participants' errors persisted when interpreting the confederate's referential communication, which suggests that the early moments of mental state processing may be relatively less affected by manipulations of low-level visual factors.

At this point we must also consider the deflationary explanations of the current findings. Let us first consider the action-prediction paradigm. In evaluating Southgate and Vernetti's (2014) work, it may be argued that participants did not need to track others' mental states to predict their actions. Infants and adults could have learned, over successive trials, to associate outcomes (reaching or no reaching) with prior events occurring in the video sequences. However, Southgate and Vernetti reasoned that if the infants were learning over multiple trials, their anticipatory motor cortex activation (for reaching trials) would have gradually strengthened. Instead, they found that activation was greatest over the first few trials. The procedure in the action-prediction task of Experiments 1 and 2 differed from Southgate and Vernetti's in that the participants never saw the agent reaching (or not reaching) in the practice or test trials preventing a link being formed between trial events and final outcomes. Furthermore, although such a deflationary account would predict faster responding over time, statistical analysis revealed no significant difference between first half and second half reaction times for either the Location or Identity task. In sum, a learned-association deflationary account does not sufficiently explain Experiment 1 and 2's data.

Turning to the object-detection paradigm, one deflationary explanation of the critical effect revolves around the attention check. To clarify, while there is a growing number of

studies utilizing the object-detection paradigm for measuring whether and to what extent certain mindreading inferences can be automatic, the conclusions drawn have been contentious given criticisms that the critical effects are just artefacts of the timings in the attention checks used by the researchers to ensure participants' task compliance (Phillips et al., 2015). However, a recent object-detection study found that Kovács and colleagues' (2010) critical P-A+<P-A- effect was maintained despite ensuring that the attention check occurred at exactly the same time across all trials (El Kaddouri et al., 2019). Another study, involving a group of adults with high functioning autism (Deschrijver et al., 2016), found a negative correlation between the size of the critical effect and the severity of autism spectrum disorder symptoms. Assuming that attention check performances were consistent across the group, this finding does not support the idea that attention check timings alone drive the difference between P-A+ and P-A- responding. In addition, Bardi et al. (2018) showed that whilst a critical effect was uncovered in a ball-detection task involving a human-like bystander, it was not revealed in a ball-detection task involving a dog bystander, despite the attention check timings being the same for the two tasks. In alignment with these studies, Experiment 5's findings suggest that the critical P-A+<P-A- effect is stable and maintained even when attention checks are removed completely from the current task context. However, we cannot rule out the possibility that the P-A+<P-A- effect may be the result of other methodological factors. Furthermore, as per Phillip's et al.'s (2015) findings, it is too soon to make any firm conclusions about the confounding role of attention checks given that the present paradigm involves different materials and set up (e.g., forced choice instead of go-no-go response, real-life as opposed to animated agent, two occluders rather than one).

Another low-level interpretation of the object-detection paradigm is offered by Heyes (2014a) in her review of Kovács et al.'s (2010) infant data. According to Heyes, the difference in looking times between the critical conditions of the test trials (P-A+ < P-A-) is explained by retroactive interference, a phenomenon in which the encoding of the first of two events (that occur one after the other) may be disrupted by the encoding of the second event (Dewar, Cowan, & Della Sala, 2007). Let us consider this in more detail. In Kovács et al.'s experiment 5, 7-month-old infants were presented with two identical familiarisation movies in which an agent (a Smurf) placed a ball on a table which then moved behind an occluder. In the end phase of the familiarisation movies the occluder dropped to reveal the ball. The infants then watched two different test-trial movies. These had the same start phase as the familiarisation movies but then diverged as follows: in the P-A- movie, the agent was present

when the ball exited the scene in its final movement; the agent then left the scene but returned before the occluder dropped to reveal no ball; in the P-A+ movie, the agent left the scene *before* the ball exited in its final movement but, as in the P-A- movie, the agent returned to the scene before the occluder dropped to reveal no ball. Kovács et al. found that infants looked longer at the outcome of the P-A+ movie than the P-A- movie, and interpreted this as evidence that the agent's belief was automatically encoded, that is, infants calculated the agent's belief that the ball was behind the occluder and were surprised when this was not confirmed when the occluder fell away. However, Heyes suggests that the disparity in infants' looking behaviour occurs because retroactive interference is present in the P-A+ movie but not in the P-A- movie, and as a result, the outcome of former is more novel than the outcome of the latter, when comparing both to the preceding familiarisation movies. More specifically, the main event of the P-A+ trial that distinguished it from the familiarisation movie (the ball leaving the scene) may not have been encoded by the infant due to disruption from the following event, that is, the agent's return to the scene. As a result, the infants may have only encoded events that *matched* the familiarisation movies, like the ball moving behind the occluder. If so, when comparing the test movies to the preceding familiarisation movies, the events of the P-A+ version resemble the events of the familiarisation movies whereas the events of the P-A- version do not. As a result, the no-ball outcome in the P-A+ movie was more surprising. Thus, Heyes concluded that the infants' reaction to the ball-absent outcome in the P-A+ trial reveals more about novelty than about automatic belief ascription.

Heyes' (2014a) retroactive interference explanation of Kovács et al.'s (2010) infant data does not apply to the adult data of the current object-detection paradigm because Experiments 3 to 5 feature multiple test trials and no familiarisation phase. Yet it is fitting to briefly reconsider Heyes' low-level interpretation of infant data, given that the dual-process account construes Kovács et al.'s (experiment 5) findings as evidence that infants are tracking belief-like states rather than reacting to perceptual novelty. In reviewing Heyes' comments, one point of contention is that, when describing the movement of the ball in the test trials (which is the same in both P-A- and P-A+ conditions), Heyes states that "the memory for the latter part of this sequence – the part that made it different from the familiarisation sequence – is likely to have been impaired in the Novel Absent [P-A+] condition by the reappearance of the agent at the end of the sequence" (p. 651). That is, the ball leaves the scene at 14 seconds (into the movie), but the encoding of this information is disrupted by the entrance of

the agent two seconds later (at 16 seconds). Heyes suggests that the return of the agent does not disrupt the encoding of the ball's exit in the P-A- condition because there is a longer duration between events - the ball exits at 12 seconds and the agent returns 4 seconds later. However, it is important to question the assumption that the agent's return is more salient than the agent's exit. After all, the familiarisation movies do not entail any movement from the agent (except for placing the ball on the table in the start phase) so the agent's return and exit are arguably equally as attention-grabbing. If so, in accordance with Heyes' retroactive interference argument, the agent's exit should disrupt the encoding of the ball's exit in the P-A- movie *and* the P-A+ movie (as the agent exits 2 seconds after the ball exits in both movies). Thus, in the P-A- condition (as in the P-A+ condition) the infants would only encode the movements that matched the familiarisation movie (ball moving behind the occluder). It follows that equivalent looking times should be predicted between the two conditions as they were equally novel in comparison to the familiarisation movies. As this was not the case, it is reasonable to be cautious regarding Heyes' retroactive interference interpretation.

Ruffman et al. (2012) concede that the object-detection paradigm does go some way to avoiding the typical confounds of AL and VOE tasks, as the infants looking behaviour is not interpreted as a prediction of, or a reaction to, agents' search behaviours. However, perceptual access discrepancies do allow for a non-mentalistic explanation of infants' looking behaviour. For instance, in the P-A- scenario both the agent and infant saw the ball leave the scene in its final movement. When the occluder dropped and the ball was absent the infant had no reason to consider the agent's perceptual experience as the perceptual access of the infant and agent was the same. However, in the P-A+ condition both the agent and infant saw the ball move behind the occluder, but only the infant saw it subsequently leave the scene. Recognising this disparity in perceptual access, longer looking times may reflect infants' prior learning that people tend to search for things that have apparently disappeared. Alternatively, infants may look longer because they are simply puzzled by the relevance of the inconsistent perceptual access. It is important to acknowledge these possible explanations here, but it is difficult to tease them apart from those offered by the dual-process or early mindreading account; after all a minimal or full-blown mindreader may still expect searching behaviour, but the expectation would come about via the ascription of registrations or beliefs (respectively), rather than behaviour rules. Ruffman et al. appreciate that adults (in a multi-trial scenario) show similar patterns of responding, however they suggest that the patterns

may converge "for very different reasons (i.e., adults might register the false belief, infants might be confused)" (p.99). The current thesis suggests that the response patterns converge across age groups for the *same* reason, that is, both infants and adults have the capacity to efficiently register false belief-*like* states. But such a task-irrelevant, relatively automatic undertaking is limited to the processing of limited information, so is not available in situations when infants or adults must handle false beliefs about numerical identity (Experiment 1 and 2) or perspective confrontation (Experiments 3 and 5).

In Experiments 3 and 5, adults were slower to react in the L2PT task than in the L1PT task (error rates were very low in both tasks). A potential concern might be that the critical P-A+ < P-A- effect was present in the L2PT task but hidden by the longer detection responses. For example, participants may have acknowledged the difference in perspective between self and agent and slowed down accordingly, masking the effect of the automatic processing. However, for this claim to be substantiated there would have been greater reaction times in the L2PT task than in the L1PT task only when there was a difference of belief between the participant and agent (i.e., the inconsistent conditions: P+A- and P-A+). On comparing reaction times in each condition, it was found that this was not the case. One explanation of the condition-wide slowing down of L2PT reaction times may be that the participant, made aware of the perspective-relevant nature of the object for the self and other, is motivated to engage in flexible off-line mindreading by using an embodied representation of the self that is then rotated to the current bodily position of the agent's position in space (Surtees et al., 2013b).

A different explanation for Experiments 3 and 5, which still preserves a dual-process account of human mindreading is that the content of the agent's registration that is efficiently tracked differs between tasks. For example, in the L1PT task, the participant tracks the agent's registration that a blue ball left the scene in the P-A- condition. When the blue ball is revealed, the encoded registration interferes with the colour detection response, prolonging the reaction time in comparison with the P-A+ condition in which there is no such interference. By contrast, in the P-A- condition of the L2PT task, the participant may simply compute the agent's registration that the 'dog-robot' moved behind the occluders, so when the object is revealed in the outcome phase there is no such interference in comparison with the P-A+ condition. The nature of the task provokes the idea that neither participant nor agent tracked the dog-robot's colour as it moved through the scene: participants may have paid no

attention to the movements of the heterogeneous object during the trial and relied only upon the final revelation to make a colour selection. Experiment 1's reality bias (P+<P-) was reduced for the L2PT task compared to the L1PT task, which in some part supports this conjecture. However, there was no replication of this finding with Experiment 3's larger sample, with the P+ versus P- differential being greater in the L2PT task then the L1PT task.

Finally, it is acknowledged that the reaction times measured across Experiments 1 to 5 are potentially influenced by three factors: the accuracy of participants' own beliefs, the accuracy of the agent's belief, and the content of the agent's belief (which may or may not be accurate). We should consider the possibility that there is a confound between the latter two factors, so that when we refer to the tracking of an agent's beliefs we are not clear whether it is the accuracy of the belief that is influencing the participant's behaviour, or the content of the belief, or both. That said, the current paradigm is designed to de-confound the first two factors, as is standard in false-belief testing; the experimental conditions in the present set of experiments exist precisely to separate the participant's own beliefs and expectations from the agent's beliefs or expectations. The distinction between belief content and belief accuracy is an important one but an experiment to de-confound them would need to be the subject of a future project.

8.4. *Implications*

8.4.1. Linking minimal mindreading and motoric processes

The findings of Experiments 1 and 2 bear on the speculation that a minimal model of the mind can modulate motor processes. To explore this idea further we must first address the issue of goal representation. According to the early mindreading account, Southgate and Vernetti's study indicates that infants demonstrated a sophisticated mindreading capacity because they were able to represent both the agent's goal (e.g., she desires a ball from the box) and the agent's belief about this goal (she believes it's in the box). The current thesis has discussed the efficient processing of other's beliefs, but it has not yet considered the way in which an efficient processing of desire can lead to goal identification. According to Butterfill and Apperly (2013) a minimal mindreader can represent goals without having to represent intentions or propositional attitudes like desires or beliefs. In their view goals are inferred from bodily movements. However, Michael and Christensen (2016) argue that Butterfill and Apperly underspecify *how* context-dependent goals are represented without a sophisticated

appreciation of others' mental states. By offering evidence that infants demonstrate situational goal attribution (e.g., Kim & Song, 2015; Luo & Baillargeon, 2007; Woodward, 1998) and the ability to link goals to specific agents rather than actions (Buresh & Woodward, 2007), Michael and Christensen suggest that goal attribution is likely a flexible operation, requiring psychological-state representation. In response, Butterfill and Apperly propose that early-developing flexible goal attribution is not necessarily incompatible with a dual-process account; after all, flexible reasoning about certain mental states like desire precedes the ability to reason about beliefs (e.g., Rakoczy, Warneken, & Tomasello, 2007). However, they also acknowledge than any account of minimal mindreading ought to explain how goal ascription *could* be cognitively efficient.

To address this challenge, Butterfill and Apperly (2016) offer an explanation of efficient goal tracking that draws upon an understanding of preferences. A minimal mindreader may track others' preferences as a proxy of desire, where a preference describes a relationship between an agent, two outcomes (A and B) and a probability. For example, when the participants in Experiments 1 and 2 repeatedly witness an agent's action resulting in A (reaches for blue things), and never B (reaches for red things), the probability of A increases. Within such a model, tracking desires allows a minimal mindreader to efficiently expect that when an agent can act either with the goal of reaching for blue things or with the goal of reaching for red things, she will act with the former goal. In Experiments 1 and 2 then, tracking registration as a proxy of belief enables the participant to expect that when the agent acts on the goal involving the desired blue object, her action will accord with what she registers concerning the location of the blue object. Experiments 1 and 2 found that participants were able to track, within limits, whether someone with a false belief would or would not reach for a box to retrieve a desired (e.g., blue) or undesired (e.g., red) object. Supporting Butterfill and Apperly, these findings suggest that it is possible for the efficient tracking of belief and the efficient tracking of desire to jointly inform expectations of others' future actions. In this way, a minimal model of mind that involves principles linking preferences, registrations and goals can help us to efficiently anticipate others' future actions.

It must be stressed here that the precise role played by motor processes in the understanding others' actions is under debate, and outside the remit of the current work. However, it is apposite at this juncture to draw upon a growing body of research that has implicated motor processes in the attribution of others' goals. As reported in Chapter 2

(Section 2.1.1), there is evidence that the motor cortex is recruited, not only when one is observing another's action, but also when one is generating a prediction of that action. More specifically, research in mirror neurons and motor simulation suggests that the observer motorically represents others' bodily actions, such as reaching and grasping, and that these motor representations of action outcomes can generate expectations concerning another agent's behaviour (Cross et al., 2013; Flanagan & Johansson, 2003; Kilner et al., 2004; Rizzolatti & Sinigaglia, 2010; Sinigaglia & Rizzolatti, 2011; Southgate, Johnson, et al., 2010). The observer's motor representation of the outcome of an agent's action is essentially the goal of the action, which suggests that motor processes are 'planning-like' (Butterfill & Apperly, 2016). In support of this, research has shown that observers are quicker and more accurate in predicting the target object, or goal, of an agent when they can exploit specific motor cues (Ambrosini, Costantini, & Sinigaglia, 2011). In their experimental trials Ambrosini and colleagues presented adults with multiple movie clips in which an individual reached towards one of two objects with a 'no-shaped' hand (a closed fist) or a pre-shaped hand (which was either a whole-hand grip, appropriate for grasping the larger object, or a pincer-type grip, appropriate for grasping the smaller object). The authors found that participants were quicker and more accurate in detecting the appropriate target in the pre-shape, compared to the no-shape trials. If motor cues can facilitate anticipatory looking, then it may be argued that one's own motor representations can drive one's own eye movements in anticipation of another's goal-directed action. This in turn suggests that motor representations *are* involved in understanding and predicting other's actions in an efficient manner – in a way that is consistent with minimal mindreading.

However, the goal-tracking process cannot 'ignore' the efficient belief-tracking process, or else it would lead to errors regarding the goals of others in false-belief scenarios. One consideration, that emerges from the findings of Experiments 1 and 2, is that efficient belief-tracking influences the process by which the behavioural expectation is generated – in other words, one's representation of an agent's registration can influence the environment as 'seen' by one's motor system. This would suggest that belief-tracking processes interact with and motor processes in some way – and perhaps that they even share a common representational format. I return to this notion in Section 8.5.

8.4.2. More than identity

Little is known about whether human beings' fast-paced mindreading is

computationally restricted to processing a limited kind of content, and what exactly the nature of that signature limit might be. The findings from Experiments 3 and 5 raise a fundamental point that proponents of the dual-process account of human mindreading have not addressed in the literature. To remind the reader, Experiments 3 and 5 report that Kovács et al.'s (2010) (P-A+<P-A-) critical effect is maintained in an object-detection task involving Level 1 perspective-taking but not in the same task incorporating a Level 2 (perspective confrontation) component. The presence of the critical effect in the L1PT task is readily explained in terms of a minimal model of the mind: humans efficiently model other people's minds in terms of registrations (relationships to objects), even when the encoding of others' belief-like states is completely irrelevant to the task being performed. However, the obliteration of the critical effect in the L2PT task cannot be explained by a breakdown in the ability to efficiently process object identity per se. Explorations of signature limits on efficient processing often rely on belief-reasoning tasks that are designed to exploit the subtle understanding that attributions of identity can generate mistakes in the numerical sense. To clarify, there are two kinds of numerical identity mistakes: compression, in which there are in fact two entities but someone falsely believes there is one, and expansion, in which there is in fact one entity but someone falsely believes there are two. The rotation of the dog-robot toy was revealed to the agent so there is nothing to suggest that the agent is necessarily going to make mistakes about identity in the numerical sense, that is, to think that there are two dog-robots when there is really one.

The current thesis offers a new conjecture: that representations underlying automatic belief-tracking either do not specify agents' locations or do not specify objects' orientations, or perhaps neither. This conjecture generates the prediction that automatic belief-tracking alone will not yield expectations about agents' perspectives, which would explain the elimination of the critical effect in the L2PT task. If the participant has not encoded where the agent was when she last encountered the object (the agent could have been on either side of the table), she cannot make a prediction about what the agent expects to see. If the participant has encoded the agent's location but only encoded the object as a bare object (that is, its orientation is not part of the registration), then the participant has the object, the registration, and the agent's current location, but he or she cannot go back and work out what the agent is expecting to see.

Prior to this thesis, it was an open question as to whether registration, being a

relationship to an object and its location, might include detailed information about the agent. Data revealed in the L1PT task suggest that the P-A+<P-A- effect can be explained by registration alone (where the object was at time of registration) without the need to assume that the registered location amounts to a belief state. The elimination of the P-A+<P-A- effect in the L2PT task suggests that registration as a belief-like state is further impoverished in not taking into account the agent's position in space in relation to the object. In belief-tracking, representing the agent's location and orientation would be relevant to understanding how someone perceives and expects the world to be, but perhaps there is a distinction between representing the agent merely as an individual when assigning the representation, and representing the agent's position in space as *part* of the registration. Thus, one possibility is that the registration comprises the spatial location of the agent and all entities in the agent's field. Another possibility is that the agent's presence may trigger the generation of a registration containing only [Objects seen by agent] (see Surtees, Samson, & Apperly, 2016). In other words, the agent's visual as well as spatial perspective can be important for what the agent registers, but the efficient mindreading process may not necessarily encode and/or store those parameters within the registration itself. If we are asking a question, as applied to the L2PT task, about what the agent expects to see or happen when the screens drop, we can answer that question using a flexible mindreading process based both on what the agent believed he last perceived from that spatial position and imagining ourselves in the agent's current position. The findings of Experiments 3 and 5 suggest that efficient mindreading is not set to handle different beliefs in combination with perspectives, as it seems that tracking registration encodes where the dog-robot-object's is placed in the scene but perhaps not how the agent is located with respect to the dog-robot, or how the dog-robot is represented from that location. The findings showing adults' resistance to the influence of an agent's perspective and belief in the L2PT task reveals important information about the specific parameters of the signature limit that constrains the efficient and relatively automatic mindreading process. If the encoding of someone's belief, vis-à-vis how the person's location in space restricts the aspects of the object in focus, is naturally eschewed by an efficient mindreading process, it would explain why studies show that adults are immune to altercentric interference over how others experience the meaning of rotationally asymmetrical digits (e.g., a number that looks like a 6 to the participant and a 9 to the agent) (Surtees et al., 2012).

8.5. *Future research*

8.5.1. Representational underpinnings

The current work highlights the need to investigate the nature of the representations underpinning efficient belief-tracking. In other cognitive domains, such as object cognition, hypotheses about the nature of the processes have generated many testable predictions (see Chapter 3, Carey, 2009). One potential avenue of exploration is the idea that efficient belief-tracking processes can influence and/or be influenced by motor processes. As discussed in Section 8.4.1, there is theoretical motivation for deliberating an interaction between motor and belief-tracking processes, and Experiments 1 and 2 already hint at this possibility.

In Section 8.4.1, the notion of efficient goal-tracking was outlined, with reference to a number of studies that have reported a link between motor representations and action understanding. However, comparatively little research has been undertaken incorporating false-belief scenarios. Afterall, as Experiments 1 and 2 indicate, accurate action predictions require belief-tracking processes to inform goal ascription processes. That is, if participants had ignored the agent's false beliefs in the action-prediction paradigm they would have performed poorly in each trial. Butterfill and Apperly (2016) suggest that efficient belief-tracking processes may influence behavioural expectations in the same way that perceptual processes do. To clarify, planning-like motor processes incorporate various 'actual' environmental factors in determining how others' actions will unfold. If, as Jeannerod (2006) infers, these planning-like processes also take into account non-actual environmental factors, then it may be proffered that the agent's registration (as a non-actual environmental factor) could modulate them resulting in effective and efficient belief-tracking.

One possible way to investigate the interaction between motoric and efficient belief tracking, would be to draw upon Ambrosini, Sinigaglia, and Costantini's (2012) study which sought to determine how behaviour expectations would be modified by restricting the observer's own bodily movements. Based on a prior procedure (in which eye movements were recorded as participants watched an agent reach for one of two objects with a pre-shaped or no-shaped hands; Ambrosini et al., 2011), they found that action prediction was modulated when the participants' own movements were appropriately restrained. Specifically, for pre-shaped trials, anticipatory gaze behaviour was significantly impaired in participants whose own hands were tied behind their backs. In contrast, there was no

difference between hands-free and hands-tied eye movements in the no-shape trials.

How would restrictions on motor representations affect action-prediction in scenarios requiring the tracking of false beliefs? The present action-prediction paradigm would not be suitable for such an investigation given that the participant's 'reach-and-press' response was specifically designed to match the agent's reaching behaviour, thereby allowing for primed reaction times in cases where the participant anticipated the agent to act in accordance with her false belief (in AD+ trials). A promising avenue of future research, however, could be to build upon Ambrosini and colleagues (2011, 2012) work. For example, in a new study, participants could observe the reaching action of an individual towards one of two objects, with their hands tied or with their hands free. By manipulating motor cues (the agent's hand shape) and belief content (the agent's true or false belief about the objects' locations) it may be possible to learn more about the possible interaction between motor representations and minimal mindreading. One possibility, conjectured above, is that they even share a common representational format. If so, one would expect the restriction to bodily movements to impact the anticipatory looking such that their accuracy (as measured by anticipatory looking) would be impaired, in the hand-tied conditions only. Certainly, further work is required to provide a more fine-grained picture of the conceivable link between belief-tracking and the motor processes underpinning goal-tracking, for example, garnering information regarding the timings of these processes.

8.5.2. Rotational confound

The present study acknowledges the rotational confound integral to many false belief tasks involving aspectuality. One way to address the concerns of single-system protagonists regarding the additional demands of mental rotation tasks would be to modify the object-detection paradigm of Experiments 3 and 5. In an alternative version of the L2PT task the participant would see the dog-robot zip behind the screens. Demands on rotation would be attenuated by having the agent move around to the participant's position in space when he returns (so that he shares the participant's view of the final outcome). Now, if adults reason with a single mindreading system that is context sensitive, it is possible to predict that participants' reaction times will be modulated by what the agent believes he is expecting to see from his new position in space. This prediction involves participants successfully tracking both the nature of the object and the agent's position in space. On the other hand, if adults have an efficient mindreading process where the agent's location is just not encoded or stored

as part of the registration itself (as the findings of Experiments 3 and 5 would suggest), then there should be no evidence of adults being automatically influenced by the agent's belief relative to how his expectations change as he moves in relation to the object. If the latter turns out to be the case, it is less about differences in demands between tasks that mask expression of mindreading competency but more about embodied mental rotation being conceptually and mechanistically closer to flexible rather than efficient mindreading processes. I will briefly return to this in closing the following section.

8.6. *Reflections*

Since the completion of Experiments 1 to 5, Scott, Roby and Baillargeon (in press) have presented their latest views on the debate between one-system and two-system accounts. In advocating the assumption that infants' and adults' psychological reasoning is qualitatively comparable, they position themselves at odds with the idea that there are signature limits to infant belief-reasoning. In particular they question the claims that infants cannot process false beliefs in scenarios that incorporate identity and deal with multiple interlocking mental states that causally interact. In support of their argument the authors continue to draw heavily upon the findings of two studies - Scott and Baillargeon's (2009) penguin task and Scott et al.'s (2015) rattle task – without dealing head on with the challenges to a rich interpretation outlined in Section 1.3.3. Bearing in mind that these critiques (e.g., Butterfill & Apperly, 2013; Heyes, 2014a; Low et al., 2016; Ruffman, 2014; Wellman, 2014) are directed at the penguin task, it would be prudent here to address the rattle task by examining Scott et al.'s findings under a dual-process lens. The rattle task data continues to be presented as evidence that infants' psychological reasoning system is conceptually rich and abstract, but the following sections offer a minimal mindreading explanation of the findings.

Scott et al. (2015) investigated infants' understanding of how an agent sought to implant a false belief about an object's identity in another agent. Infants were presented with a scenario in which a thief (T) tried to steal a desirable (rattling) toy while its owner (O) was absent, by exchanging it with a less desirable (non-rattling) version. First, the infants watched six familiarisation trials featuring different coloured rattling and non-rattling toys (see Figure 8-1). In the rattling-toy familiarisation trials, O emerged through a curtain after knocking twice and placed a toy on the table (on a tray). She shook the toy, causing it to rattle, and placed it back on the tray. Then she left the scene saying, "I'll be back!". In her absence T picked up the toy, shook it; when she heard a "knock, knock" sound she placed it back on the

tray. O returned, picked up the toy without shaking it, and placed her rattling toy in the box next to the curtain. Both T and O then paused until the trial ended. The silent-toy trials were the same except that: the toy did not rattle when O shook it; T did not pick the toy up when O left the scene; and when O returned, she threw the non-rattling toy into the bin. Under a dual-process account, by the end of the familiarisation trials both T and O have registered that there are three non-rattling (red, orange and purple) toys in the bin and three rattling toys (green, yellow and blue) in the box (see Figure 8-1).

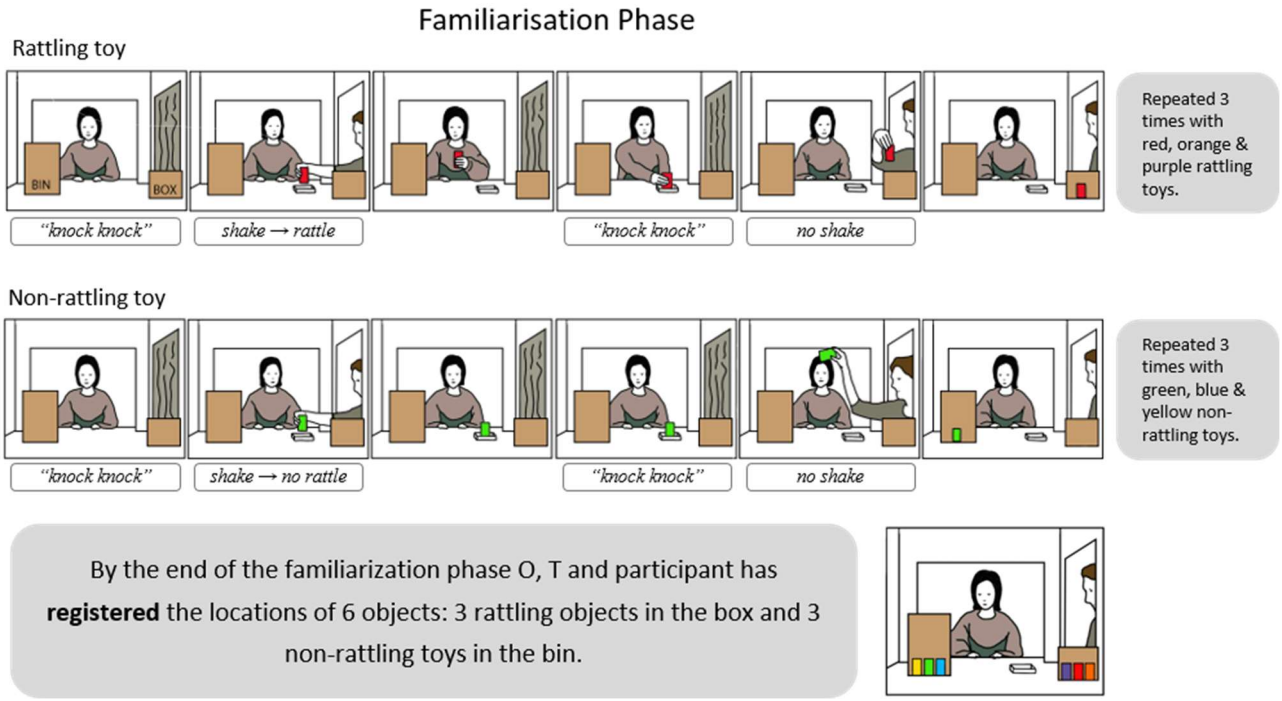

Figure 8-1 Events in the familiarisation trials of Scott et al. (2015)

To assist in the explanation of events, the location of the object/s is depicted at all times - in the actual trials the objects were not visible after being placed in the bin or box. The colour scheme shown in this figure relates to this summary only (Scott et al. applied different combinations of colours and patterns for their rattling and non-rattling toys). The order of the trials was: rattling, non-rattling, non-rattling; rattling, non-rattling, rattling. T = thief; O = owner.

In the test phase of their 'Deception Condition', the infants either viewed a matching or a non-matching trial (see Figure 8-2). In the matching trial O returns with a green toy on a tray - it is identical in appearance to the one she previously threw into the bin. However, this particular green toy rattles when she shakes it. O places the rattling green toy back on the tray and leaves saying, "I'll be back!" In her absence, T picks it up with her left hand, and reaches into the bin for the identical looking, non-rattling green toy with her right hand. She places

the non-rattling green toy on the tray and hides the rattling green toy in her large jumper pocket. T returns her hands to the table and the participants watch the paused scene until the trial ends. The non-matching trial is the same except that T replaces the green rattling toy with a *yellow* non-rattling toy from the bin.

Scott et al. (2015) found that infants tended to look longer at non-matching compared to matching trials. According to the authors, this disparity reflects early, sophisticated belief-reasoning capabilities: the infants reasoned that T wished to steal the preferred rattling toy, but could only do so if O falsely believed (on her return) that the toy in the tray was the rattling green toy. Hence, it was reasonable for T to substitute the desired toy with one that matched its appearance, but not reasonable to substitute it with one that did not. Furthermore, to rule out a possible low-level explanation - that the infants looked longer in the non-matching trials because of a not-seen-before switch of toy colour-switch – the authors added a 'Silent Condition'. In this condition O brought a non-rattling toy to the matching and non-matching test trials, and the data revealed similar looking times between matching and non-matching trials. The authors claimed that these results reflected infants' failure to generate a causally coherent explanation for T's actions given her preference for rattling toys.

However, these results can also be interpreted in keeping with a dual-process account. In the non-matching trial, when O returns and re-encounters the non-rattling yellow toy, there would be significant disruption to her registrations: O would have to revise her registrations of having placed the green rattling toy on the tray, and thrown the yellow, non-rattling toy in the bin. In contrast, when O eventually re-encounters the green, non-rattling toy on the tray there would be minimal disruption to her registrations: O would still hold the registration that she placed the green rattling toy on the tray. Infants' abilities to track O's registrations of locations of objects over time can just as well predict the outcome that infants showed longer looking at the non-matching trial as compared to the matching trial.

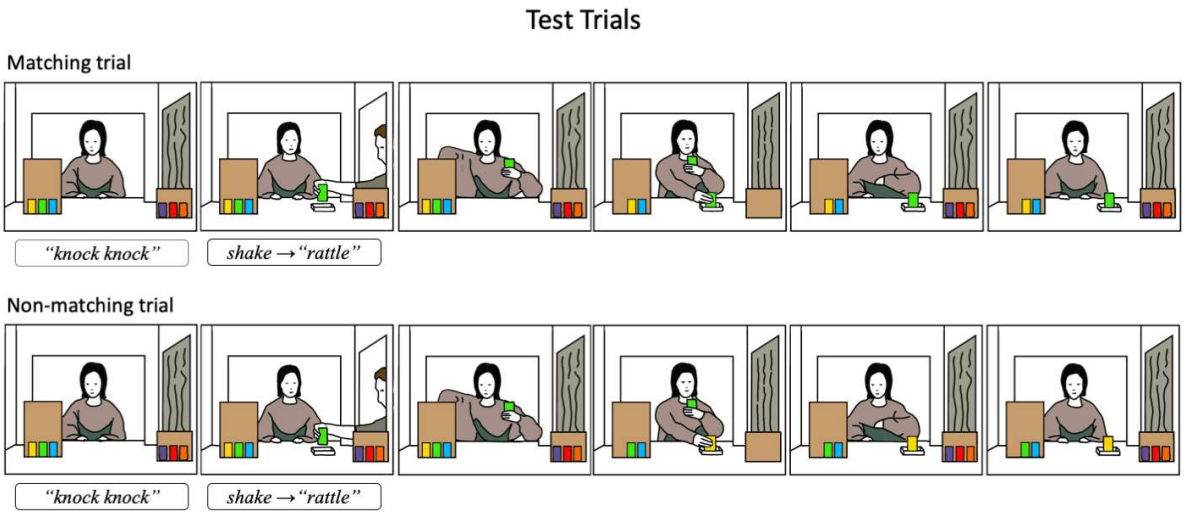

Figure 8-2 Events in the test trials of Scott et al. (2015)

Notes. As in Figure 8-1, the locations of the object are depicted at all times - in the actual trials the objects were not visible when in the bin or box.

However, these results can also be interpreted in keeping with a dual-process account. In the non-matching trial, when O returns and re-encounters the non-rattling yellow toy, there would be significant disruption to her registrations: O would have to revise her registrations of having placed the green rattling toy on the tray, and thrown the yellow, non-rattling toy in the bin. In contrast, when O eventually re-encounters the green, non-rattling toy on the tray there would be minimal disruption to her registrations: O would still hold the registration that she placed the green rattling toy on the tray. Infants' abilities to track O's registrations of locations of objects over time can just as well predict the outcome that infants showed longer looking at the non-matching trial as compared to the matching trial.

But what of Scott et al.'s (2015) silent condition? Adopting a minimal mindreading explanation, similar looking times across the two trials is perhaps due to the likely disruption to O's registration being less salient in the Silent Condition (where the T's scheme involves a single cue; colour only, with a green non-rattling toy being swapped for a yellow non-rattling toy) than in the Deception Condition (where T's scheme covers two cues; colour and sound, with a green rattling toy being swapped for a yellow non-rattling toy).

In their second experiment, Scott et al. (2015) modified each of the six familiarisation trials so that O, on returning to the scene, shook the toy again before placing it in either the box (rattling toys) or the bin (non-rattling toys). The test trials were identical to that of their

first experiment. Their "Shake Twice" Condition revealed similar looking times for matching and non-matching scenarios, which the authors interpreted as evidence for infants understanding that substituting a rattling toy with an identically looking non-rattling toy would now be as ineffective as substituting a different coloured toy. However, under a minimal mindreading account there would be disruption to O's registrations in both trials: in the non-matching trial, when O re-encounters the yellow non-rattling toy back on the tray she would have to revise her registrations of having placed a green rattling toy on the tray, and of already binning the yellow non-rattling toy; in the matching trial, when O shakes the green non-rattling toy, she would have to revise her registrations of having placed the green rattling toy on the tray and of having placed the green non-rattling toy in the bin.

Finally, in their third experiment Scott and colleagues (2015) sought to further explore infants' sensitivity to the circumstances that could facilitate effective deception. They used the familiarisation trials from their first experiment but modified the test trials to create a Deceived Condition and an Alerted Condition. In the former, the test trials were initially the same as the matching trial from Figure 8-2, except that at the end of the trial the infants either saw O discard the matching-in-appearance (non-rattling) toy (discard trial) or place it in her box (store trial). As expected, they found that infants looked longer at the discard, compared to the store trial, reflecting infants' surprise that O would throw the toy in the bin when she had every reason to (mistakenly) believe that matching-in-appearance toy was the same (rattling) toy she had just placed in the tray. Scott et al. argued that a minimal mindreading account would make the opposite prediction: when O returned and saw the toy she would register it as the green non-rattling toy "for what it really was" (p.48) and discard it, so that infants would be surprised (look longer) if O chose to store a non-rattling store trials. In response to this claim, an alternative minimal mindreading account is offered. In this explanation O registered the locations of the two green toys in the world each having a different attribute – the green non-rattling toy in the bin and the green rattling toy on the tray. Given that, on her return, her encountering of the object on the tray matches her registration of the object at said location, she should store the toy on the tray. Hence, in line with Scott et al.'s findings, the minimal mindreading account predicts that infants should look longer at discard compared to store trials.

The Alerted Condition was the same as the Deception Condition except that O returned earlier to find T with a visually similar type toy in each hand. T then placed the non-rattling

version on the tray and the rattling version in her pocket. As predicted by the authors there was no difference in looking time between discard and store trials - given that O did not know which toy was which, the infants could form no expectations about her future actions. In contrast, the authors suggested that a minimal mindreading account would have predicted longer looking in store trials, "as infants expect O to register the toy on the tray as the silent toy". This latter claim, however, misconstrues the predictions of a minimal mindreading account. Instead, consider the following: O registered the locations of two greens in the world, each having a different attribute – the green non-rattling toy is registered in the bin and the green rattling toy was registered on the tray. In returning to the scene and encountering T holding two green toys, side by side, O would be unable to make clear cut registrations of which toy came from which location. O's previous registrations would not be able to assist her either. Thus, there would be no clear indication of what O would do, and subsequently no prediction that infants would look at one (store or discard) trial longer than the other.

In conclusion, claims that 18-month-olds can reason about one person's intention to implant in another person a false belief about object identity should be treated with caution. Furthermore, given the replication issues raised in Section 1.2.3, a continued effort must be made to resolve the debate regarding the reliability and robustness of non-verbal theory of mind measures. Lack of a methodological consensus, and the general difficulties of working with infants and toddlers, suggest that questions relating to the development and nature of mindreading are best addressed via a collaborative multi-lab approach (e.g., *The ManyBabies* projects).

Another, more recent publication warrants attention here: Ward, Ganis, and Bach (2019) report evidence that L2PT may be automatic. While previous work has shown that adults and older children can spontaneously (i.e., independent of task instructions) represent how others view an object or spatial layout, these studies have all involved participant-agent interaction (such as active engagement in a joint task) or explicit judgements of the other's perspective (Elekes, Varga, & Király, 2016, 2017; Freundlieb, Kovács, & Sebanz, 2016; Surtees, Apperly, & Samson, 2016). In Ward, Ganis, and Bach, however, the agent played an entirely passive role, as participants judged whether an alphanumeric character, presented in differing orientations, was in its canonical or mirror-flipped form (e.g., 'R' or 'Я'). In keeping with previous findings (Shepard & Metzler, 1971) people were slower to respond the

more they had to mentally rotate the stimulus to its typical orientation. However, the presence of the passive agent facilitated judgements. For example, if the alphanumeric character was rotated away from the participant, correct responding was faster if the character appeared upright to the passive agent. The authors concluded that L2PT is not necessarily subject to effortful processing, nor does it require motivating factors like joint action.

At face value these findings directly challenge the conjecture that L2PT distinguishes automatic and non-automatic belief tracking. However, the present thesis (Experiments 3 and 5) emphasizes that it is L2PT *involving perspective-confrontation* (Moll et al., 2012; Perner et al., 2002) that stymies the automatic processing of others' viewpoints. As Apperly (2019) points out, in Ward et al.'s (2019) task there was never any conflict between the participant and the passive 'other' regarding the correct response. By comparison, in the L2PT tasks of Experiments 3 and 5, the agent and participant would produce conflicting responses if both were required to detect the colour revealed when the occluders fell away. While Ward and colleagues suggest they have found evidence of automatic L2PT, it is difficult to conclude that the participants effortlessly took on board the passive agent's perspective. Notably, there was a mental rotation effect for both the participant *and* the passive agent. That is, character recognition was slowed by an increase in the alphanumeric character's angular disparity from the participant's viewpoint, *and* an increase in angular disparity from the passive agent's viewpoint. It is argued here that a perspective-taking process that is affected by angular disparity cannot be described as automatic. By contrast, judging whether or not someone can see an object or not (L1PT) *is* unaffected by angular disparity (e.g., Kessler & Rutherford, 2010; Michelon & Zacks, 2006) and *can* be processed effortlessly. Ward and colleagues' data may speak to rapid L2PT, but their findings are not in direct dispute with the current work as their angular disparity data seem to reflect the workings of a sophisticated and cognitively demanding process.

The detection of an angular disparity effect in a number of L2PT tasks (Kessler & Rutherford, 2010; Kessler & Thomson, 2010; Michelon & Zacks, 2006) has led some researchers to suggest that mental rotation is "one source of difficulty" in calculating how others may experience the same entity from a different perspective (p. 9, Surtees et al., 2013b; Surtees, Apperly, & Samson, 2013a). Despite Carruthers' (2016) claims, this supposition does not necessarily support a one-system account of mindreading, in which infants and young children have the conceptual capacity to succeed in L2PT tasks but lack

the domain-general mental rotation skills to demonstrate their understanding. Rather, in this closing reflection I support the view that mental rotation skills (along with a range of other cognitive abilities) may be a *pre-requisite* for the conceptual (and flexible) understanding that others' perspectives can give rise to experiences and impressions that are radically different to our own.

8.7. *Conclusions*

This thesis has investigated the dual-process theory of human mindreading. Experiments 1 and 2 presented a new reaction time paradigm that successfully differentiated the cognitive mechanisms that underlie those distinct mindreading abilities. Both studies exposed dissociations in the profile of adults' speeded responses over certain phenomena, whereby appropriate response facilitation occurred when tracking false beliefs about location, whilst inappropriate facilitation occurred when tracking false beliefs about identity. Experiment 3 uncovered that adults' reaction times in the L1PT task were helpfully speeded by a bystander's irrelevant belief when tracking two homogenous objects but not in the L2PT task when tracking a single heterogeneous object. The limitation is especially striking given that the heterogeneous nature of the single object was fully revealed to participants as well as to the bystander. Together, the current behavioural data provides new and converging evidence for Apperly and Butterfill's (2009) dual-process account that adult humans draw upon multiple systems and models of mind for making action predictions.

The current body of work represents a move away from debating whether a mindreading process uses a minimal-theory-of-mind model, to assuming that it does and then working out what exactly the signature limit of the process might be. Furthermore, it advances the field by showcasing two innovative and promising tools for assessing the competing theories that seek to explain the cognitive architecture underlying humans' automatic and non-automatic mindreading abilities. Finally, this thesis provides evidence for a new signature limit on automatic belief-tracking, which both informs us about the nature of the representations involved and also motivates the field to search for further signature limits.

A legitimate question is why a relatively separate, and restricted, automatic mindreading process - which persists beyond infancy and childhood - would have evolved in humans: how adaptive is a mental-state calculator that, under certain circumstances, breaks down? One possibility is that fast, but limited processing in adulthood may be an adaptive

reaction to the demands of complex environments (Payne, Bettman, & Johnson, 1993). As social animals it is imperative that we have the ability to quickly predict the motives and actions of others (especially dangerous ones). And whilst, as fully matured humans we routinely come across hurried instances in which we erroneously infer others' intentions, desires and beliefs, our experiences also inform us that even the most studious deliberation of others' minds is far from fool proof. Limited processing may lead to erroneous judgements, but it is important to grasp that cognitive limitations are not exclusively linked to negative outcomes (Hertwig & Todd, 2003). Even in a simple object-detection task involving a homogenous object, we have shown that performance is enhanced by the automatic belief ascription of other agents.

The dual-process account of human mindreading continues to stand out as an influential and articulated theory, offering testable explanations of the 'developmental gap' and of fast/slow mature mindreading. While I hope that the current thesis has provided some answers regarding the nature of the mature mindreading, I acknowledge that a number of new issues have been raised. The processes underlying efficient mindreading are yet to be fully described, and questions remain as the scope of, and boundaries to, fast-paced belief-tracking in adults, children and infants. The challenge for future research is to build upon the current findings to determine the cognitive components that underlie what may be relatively distinct mindreading abilities. This body of work could be a launching point for future research looking at the interface between the efficient processes involved in belief tracking and the motor processes involved in goal-tracking. More generally, future investigations require a broadening of research horizons alongside comprehensive investigations of the multiple cognitive processes and cognitive systems that shape children's developing abilities to track and ascribe others' beliefs. This thesis shows that testing the boundary conditions of the different mindreading systems is a promising avenue to inform our knowledge of the cognitive mechanisms that make mindreading possible.

References

Ambrosini, E., Costantini, M., & Sinigaglia, C. (2011). Grasping with the eyes. *Journal of Neurophysiology*, *106*(3), 1437–1442. https://doi.org/10.1152/jn.00118.2011

Ambrosini, E., Sinigaglia, C., & Costantini, M. (2012). Tie my hands, Tie my eyes. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 263–266. https://doi.org/10.1037/a0026570

Apperly, I. A. (2010). *Mindreaders: The cognitive basis of "Theory of Mind."* Hove: Psychology Press/Taylor & Francis Group.

Apperly, I. A. (2013). Can theory of mind grow up? Mindreading in adults, and its implications for the development and neuroscience of mindreading. In S. Baron-Cohen, H. Tager-Flusberg, & M. V. Lombardo (Eds.), *Understanding other minds: Perspectives from developmental social neuroscience* (pp. 72–92).

Apperly, I. A. (2019). The benefit of seeing in company. *Trends in Cognitive Sciences*, *23*, 451–453. https://doi.org/10.1016/j.tics.2019.03.005

Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential theory of mind task. *Cognition*, *106*(3), 1093–1108. https://doi.org/10.1016/j.cognition.2007.05.005

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970. https://doi.org/10.1037/a0016923

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., & Moffitt, G. (2010). Why are there limits on theory of mind use ? Evidence from adults ' ability to follow instructions from an ignorant speaker. *The Quarterly Journal of Experimental Psychology*, *63*(6), 1201–1217. https://doi.org/10.1080/17470210903281582

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, *17*(10), 841–844. https://doi.org/http://dx.doi.org/10.1111/j.1467-9280.2006.01791.x

Avis, J., & Harris, P. L. (1991). Belief-Desire Reasoning among Baka Children: Evidence for a Universal Conception of Mind. *Child Development*, *62*(3), 460–467. https://doi.org/10.1111/j.1467-8624.1991.tb01544.x

Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition*, *115*(1), 54–70. https://doi.org/10.1016/j.cognition.2009.11.008

Baillargeon, R., Buttelmann, D., & Southgate, V. (2018). Invited Commentary: Interpreting failed replications of early false-belief findings: Methodological and theoretical considerations ☆. *Cognitive Development*, *46*(June), 112–124. https://doi.org/10.1016/j.cogdev.2018.06.001

Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review*

of Psychology., *67*, 159–186. https://doi.org/10.1146/annurev-psych-010213-115033

Baillargeon, R., Scott, R. M., & He, Z. (2010). False belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–117.

Bardi, L., Desmet, C., & Brass, M. (2019). Spontaneous Theory of Mind is reduced for nonhuman-like agents as compared to human-like agents. *Psychological Research*, *83*(7), 1571–1580. https://doi.org/10.1007/s00426-018-1000-0

Bardi, L., Six, P., & Brass, M. (2017). Repetitive TMS of the temporo-parietal junction disrupts participant's expectations in a spontaneous Theory of Mind task. *Social Cognitive and Affective Neuroscience*, *12*(11), 1775–1782. https://doi.org/10.1093/scan/nsx109

Bargh, J. A. (1994). The Four Horsemen of Automaticity: Awareness, Intention, Efficiency and Control in Social Cognition. In J. A. Bargh, J. Wyer, R. S., & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 1–40). Erlbaum: Hillsdale.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Boston: MIT Press.

Bidet-Caulet, A., Barbe, P.-G., Roux, S., Viswanath, H., Bartelemy, C., Bruneau, N., … Bonnet-Brilhault, F. (2012). NIH Public Access. *European Journal of Neuroscience*, *36*(7), 2996–3004. https://doi.org/10.1111/j.1460-9568.2012.08223.x.

Birch, S. A., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, *18*(5), 382–386. https://doi.org/10.1111/j.1467-9280.2007.01909.x

Bull, R., Phillips, L. H., & Conway, C. A. (2008). The role of control functions in mentalizing: Dual-task studies of Theory of Mind and executive function. *Cognition*, *107*(2), 663–672. https://doi.org/10.1016/j.cognition.2007.07.015

Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition*, *104*(2), 287–314. https://doi.org/10.1016/j.cognition.2006.07.001

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

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*(2), 337–342. https://doi.org/10.1016/j.cognition.2009.05.006

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. https://doi.org/10.1016/j.jecp.2013.10.002

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

Butterfill, S. A., & Apperly, I. A. (2016). Is goal ascription possible in minimal mindreading ? *Psychological Review*, *123*(2), 228–233. https://doi.org/10.1037/rev0000022

Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and

great apes. *Child Development*, *70*(2), 381–395. https://doi.org/10.1111/1467-8624.00028

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*, 187–192. https://doi.org/10.1016/j.tics.2008.02.010

Carey, S. (2009). *The Origin of Concepts*. New York: Oxford University Press.

Carlson, S. M., Claxton, L. J., & Moses, L. J. (2013). The Relation Between Executive Function and Theory of Mind is More Than Skin Deep. *Journal of Cognition and Development*, *16*(1), 186–197. https://doi.org/10.1080/15248372.2013.824883

Carlson, S. M., & Moses, L. J. (2001). Individual Differences in Inhibitory Control and Children's Theory of Mind. *Child Development*, *72*(4), 1032–1053. https://doi.org/doi.org/10.1016/j.jecp.2004.01.002

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, *87*(4), 299–319. https://doi.org/10.1016/j.jecp.2004.01.002

Carruthers, P. (2013). Mindreading in infancy. *Mind and Language*, *28*(2), 141–172. https://doi.org/10.1111/mila.12014

Carruthers, P. (2015). Mindreading in adults: Evaluating two-systems views. *Synthese*. https://doi.org/10.1007/s11229-015-0792-3

Carruthers, P. (2016). Two Systems for Mindreading? *Review of Philosophy and Psychology*, *7*(1), 141–162. https://doi.org/10.1007/s13164-015-0259-y

Carruthers, P. (2018). Young children flexibly attribute mental states to others. *Proceedings of the National Academy of Sciences*, *115*, 11351–11353. https://doi.org/10.1073/pnas.1816255115

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*, 377–395. https://doi.org/10.1016/0885-2014(94)90012-4

Cole, G. G., Atkinson, M., Le, A. T. D., & Smith, D. T. (2016). Do humans spontaneously take the perspective of others? *Acta Psychologica*, *164*, 165–168. https://doi.org/10.1016/j.actpsy.2016.01.007

Confais, J., Kilavik, B. E., Ponce-Alvarez, A., & Riehle, A. (2012). On the anticipatory precue activity in motor cortex. *Journal of Neuroscience*, *32*(44), 15359–15368. https://doi.org/10.1523/JNEUROSCI.1768-12.2012

Conway, J. R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 454–465. https://doi.org/10.1037/xhp0000319

Crivello, C., & Poulin-Dubois, D. (2018). Infants' false belief understanding: A non-replication of the helping task. *Cognitive Development*, *46*, 51–57. https://doi.org/10.1016/j.cogdev.2017.10.003

Cross, E. S., Stadler, W., Parkinson, J., Schütz-Bosbach, S., & Prinz, W. (2013). The influence of visual training on predicting complex action sequences. *Human Brain Mapping*, *34*(2), 467–486. https://doi.org/10.1002/hbm.21450

Csibra, G. (1998). Seeing is not believing. *Behavioral and Brain Sciences*, *21*(01), 117–118. https://doi.org/10.1017/S0140525X98250709

Davidson, D. (1980). Towards a unified theory of meaning and action. *Grazer Philosophische Studien*, *11*, 1–12. https://doi.org/doi.org/10.5840/gps19801120

Davidson, D. (1990). *The Structure and Content of Truth*. *87*(6), 279–328.

Dennett, D. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, *1*(4), 568–570. https://doi.org/https://doi.org/10.1017/S0140525X00076664

Deschrijver, E., Bardi, L., Wiersema, J. R., & Brass, M. (2016). Behavioral measures of implicit theory of mind in adults with high functioning autism. *Cognitive Neuroscience*, *7*(1–4), 192–202. https://doi.org/10.1080/17588928.2015.1085375

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

Dewar, M. T., Cowan, N., & Della Sala, S. (2007). Forgetting due to retroactive interference: A fusion of Müller and Pilzecker's (1900) early insights into everyday forgetting and recent research on anterograde amnesia. *Cortex*, *43*, 616–634. https://doi.org/10.1016/s0010-9452(08)70492-1

Dixson, H. G. W., Komugabe-Dixson, A. F., Dixson, B. J., & Low, J. (2018). Scaling Theory of Mind in a Small-Scale Society: A Case Study From Vanuatu. *Child Development*, *89*(6), 2157–2175. https://doi.org/10.1111/cdev.12919

Doherty, M., & Perner, J. (1998). Metalinguistic awareness and theory of mind: Just two words for the same thing? *Cognitive Development*, *305*(1998), 279–305. https://doi.org/10.1016/S0885-2014(98)90012-0

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

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

El Kaddouri, R., Bardi, L., De Bremaeker, D., Brass, M., & Wiersema, J. R. (2019). Measuring spontaneous mentalizing with a ball detection task: Putting the attention-check hypothesis by Phillips and colleagues (2015) to the test. *Psychological Research*, 1–9. https://doi.org/10.1007/s00426-019-01181-7

Elekes, F., Varga, M., & Király, I. (2016). Evidence for spontaneous level-2 perspective taking in adults. *Consciousness and Cognition*, *41*, 93–103. https://doi.org/10.1016/j.concog.2016.02.010

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

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

Fizke, E., Butterfill, S. A., & Rakoczy, H. (2013). Toddlers' understanding of false belief about an object's identity. *Poster Presented at the Biennial Meeting of the Society for Research in Child Development*. Seattle, WA.

Fizke, E., Butterfill, S. A., 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. https://doi.org/10.1016/j.jecp.2017.05.005

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*, 769–771. https://doi.org/10.1038/nature01861

Flavell, J. H. (1978). The development of knowledge about visualperception. In C.B.Keasey (Ed.), *Nebraska Symposium on Motivation (Vol. 25)*. Lincoln: University of Nebraska Press.

Flavell, J. H. (1992). Perspectives on perspective taking. In H. Beilin & P. B. Pufall (Eds.), *Piaget's theory: Prospects and possibilities.* (pp. 107–139). Hillsdale, NJ: Erlbaum.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1-Level 2 distinction. *Developmental Psychology*, *17*(1), 99–103. https://doi.org/10.1037/0012-1649.17.1.99

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

Frankish, K. (2010). Dual-process and dual-system theories of reasoning. *Philosophy Compass*, *5*(10), 914–926. https://doi.org/10.1111/j.1747-9991.2010.00330.x

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2016). When do humans spontaneously adopt another's visuospatial perspective? *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 401–412.

Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. A. (2016). Altercentric interference in Level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(2), 158–163. https://doi.org/http://dx.doi.org/10.1037/xhp0000138

German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in "theory of mind": Evidence from compromised belief-desire reasoning in old age. *Cognition*, *101*(1), 129–152. https://doi.org/10.1016/j.cognition.2005.05.007

Grosse Wiesmann, C., Friederici, A. D., Disla, D., Steinbeis, N., & Singer, T. (2018). Longitudinal evidence for 4-year-olds' but not 2- and 3-year-olds' false belief-related action anticipation. *Cognitive Development*, *46*, 58–68. https://doi.org/10.1016/j.cogdev.2017.08.007

Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior and Development*, *21*(2), 167–179. https://doi.org/10.1016/S0163-6383(98)90001-7

Hamilton, A. F., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition*, *113*(1), 37–44. https://doi.org/10.1016/j.cognition.2009.07.007

Hamilton, A. F., & Ramsey, R. (2013). How are the actions of triangles and people processed in the human brain? In D. Rutherford, M. & M. Kuhlmeir, V. (Eds.), *Social Perception: Detection and interpretation of animacy, agency, and intention* (pp. 231–257). https://doi.org/10.7551/mitpress/9780262019279.003.0010

Happé, F. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129–154. https://doi.org/10.1007/BF02172093

He, Z., Bolz, M., & Baillargeon, R. (2011). False-belief understanding in 2.5-year-olds: Evidence from violation-of-expectation change-of-location and unexpected-contents tasks. *Developmental Science*, *14*(2), 292–305. https://doi.org/10.1111/j.1467-7687.2010.00980.x

Hertwig, R., & Todd, P. M. (2003). More is not always better: The benefits of cognitive limits. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making* (pp. 213–231). John Wiley & Sons, Ltd.

Heyes, C. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, *21*(1), 101–148. https://doi.org/10.1017/S0140525X98000703

Heyes, C. (2014a). False belief in infancy: A fresh look. *Developmental Science*, 1–13. https://doi.org/10.1111/desc.12148

Heyes, C. (2014b). *Submentalizing: I Am Not Really Reading Your Mind*. https://doi.org/10.1177/1745691613518076

Heyes, C. (2014c). Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on Psychological Science*, *9*, 131–143. https://doi.org/10.1177/1745691613518076

Hinten, A. E., Labuschagne, L. G., Boden, H., & Scarf, D. (2018). Preschool children and young adults' preferences and expectations for helpers and hinderers. *Infant and Child Development*, *27*(4), e2093. https://doi.org/10.1002/icd.2093

Horowitz, T. S., & Cohen, M. A. (2010). Direction information in multiple object tracking is limited by a grade resource. *Attention, Perception & Psychophysics*, *72*(7), 1765–17. https://doi.org/10.3758/APP.72.7.1765

Jeannerod, M. (2006). *Motor cognition: What actions tell the self*. Oxford: Oxford University Press.

Kano, F., Krupenye, C., Hirata, S., Tomonaga, M., & Call, J. (2019). Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proceedings of the National Academy of Sciences of the United States of America*, 1–6. https://doi.org/10.1073/pnas.1910095116

Karg, K., Schmelz, M., Call, J., & Tomasello, M. (2016). Differing views: Can chimpanzees do Level 2 perspective-taking? *Animal Cognition*, *19*(3), 555–564. https://doi.org/10.1007/s10071-016-0956-7

Kessler, K., & Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology*, *1*, 1–12. https://doi.org/10.3389/fpsyg.2010.00213

Kessler, K., & Thomson, L. A. (2010). The embodied nature of spatial perspective taking: Embodied transformation versus sensorimotor interference. *Cognition*, *114*(1), 72–88. https://doi.org/10.1016/j.cognition.2009.08.015

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38. https://doi.org/10.1111/1467-9280.00211

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*, 25–41. https://doi.org/10.1016/S0010-0277(03)00064-7

Kilner, J. M., Vargas, C., Duval, S., Blakemore, S., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience*, *7*(12), 1299–1301. https://doi.org/10.1038/nn1355

Kim, E. Y., & Song, H. J. (2015). Six-month-olds actively predict others ' goal-directed actions. *Cognitive Development*, *33*, 1–13. https://doi.org/10.1016/j.cogdev.2014.09.003

Knudsen, B., & Liszkowski, U. (2012). Eighteen- and 24-month-old infants correct others in anticipation of action mistakes. *Developmental Science*, *15*, 113–122. https://doi.org/10.1111/j.1467-7687.2011.01098.x

Kovács, Á. M., Kühn, S., Gergely, G., Csibra, G., & Brass, M. (2014). Are all beliefs equal? Implicit belief attributions recruiting core brain regions of theory of mind. *PLoS ONE*, *9*(9). https://doi.org/10.1371/journal.pone.0106558

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–1834. https://doi.org/10.1126/science.1190792

Kristen, S., Thoermer, C., Hofer, T., Aschersleben, G., & Sodian, B. (2006). Scaling of theory of mind tasks. *Zeitschrift Fur Entwicklungspsychologic Und Padagogische Psychologie*, *38*(4), 186–195. https://doi.org/10.1026/0049-8637.38.4.186

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*(6308), 110–114. https://doi.org/10.1126/science.aaf8110

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

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

Leslie, A. M. (1987). Pretense and representation: The origins of "Theory of Mind." *Psychological*

Review, *94*(4), 412–426. https://doi.org/10.1037/0033-295X.94.4.412

Leslie, A. M. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. Hirschfield & S. Gelman (Eds.), *Domain Specificity in Culture and Cognition* (pp. 119–148). Cambridge: Cambridge University Press.

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. https://doi.org/10.1111/cdep.12183

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. https://doi.org/10.1111/cdev.12224

Low, J., & Edwards, K. (2018). The curious case of adults' interpretations of violation-of-expectation false belief scenarios. *Cognitive Development*, *46*, 86–96. https://doi.org/10.1016/j.cogdev.2017.07.004

Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology*, *30*(1), 1–13. https://doi.org/10.1111/j.2044-835X.2011.02074.x

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. https://doi.org/10.1177/0956797612451469

Luo, Y. (2011). Three-month-old infants attribute goals to a non-human agent. *Developmental Science*, *14*(2), 453–460. https://doi.org/10.1111/j.1467-7687.2010.00995.x

Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, *105*, 489–512. https://doi.org/10.1016/j.cognition.2006.10.007

Lurz, R. (2011). *Mindreading animals: The debate over what animals know about other minds.* Cambridge, MA: MIT Press.

Lurz, R., & Krachun, C. (2019). Experience-projection methods in theory-of-mind research: Their limits and strengths. *Current Directions in Psychological Science*, *28*(5), 456–462. https://doi.org/10.1177/0963721419850156

Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., & Flavell, J. H. (1974). The early development of inferences about the visual percepts of others. *Child Development*, *45*(2), 357–366. https://doi.org/10.1111/1467-8624.ep12154629

Mayer, A., & Träuble, B. E. (2013). Synchrony in the onset of mental state understanding across cultures? A study among children in Samoa. *International Journal of Behavioral Development*, *37*(1), 21–28. https://doi.org/10.1177/0165025412454030

McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, *102*(2), 179–218. https://doi.org/10.1016/j.cognition.2005.12.011

Meert, G., Wang, J., & Samson, D. (2017). Efficient belief tracking in adults: The role of task instruction, low-level associative processes and dispositional social functioning. *Cognition*, *168*, 91–98. https://doi.org/10.1016/j.cognition.2017.06.012

Michael, J., & Christensen, W. (2016). Flexible goal attribution in early mindreading. *Psychological Review*, *123*(2), 219–227. https://doi.org/10.1016/j.newideapsych.2015.01.003

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & Psychophysics*, *68*(2), 327–337. https://doi.org/10.3758/BF03193680

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. https://doi.org/10.1111/cdev.12581

Moll, H., & Meltzoff, A. N. (2011). How does it look? Level 2 perspective taking at 36 months of age. *Child Development*, *82*(2), 661–673. https://doi.org/10.1111/j.1467-8624.2010.01571.x

Moll, H., Meltzoff, A. N., Merzsch, K., & Tomasello, M. (2012). Taking versus confronting visual perspectives in preschool children. *Developmental Psychology*, *49*(4), 646–654. https://doi.org/10.1037/a0028633

Mozuraitis, M., Chambers, C. G., & Daneman, M. (2015). Privileged versus shared knowledge about object identity in real-time referential processing. *Cognition*, *142*, 148–165. https://doi.org/10.1016/j.cognition.2015.05.001

Nijhof, A. D., Brass, M., Bardi, L., & Wiersema, J. R. (2016). Measuring mentalizing ability: A within-subject comparison between an explicit and implicit version of a ball detection task. *PLoS ONE*, *11*(10), e0164373. https://doi.org/10.1371/journal.pone.0164373

Nijhof, A. D., Brass, M., & Wiersema, J. R. (2017). Spontaneous mentalizing in neurotypicals scoring high versus low on symptomatology of autism spectrum disorder. *Psychiatry Research*, *258*, 15–20. https://doi.org/10.1016/j.psychres.2017.09.060

Oberle, E. (2009). The development of theory of mind reasoning in Micronesian children. *Journal of Cognition and Culture*, *9*(1), 39–56. https://doi.org/10.1163/156853709X414629

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

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

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision Maker*. New York: Cambridge University Press.

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

Perner, J. (2010). Who took the cog out of cognitive science? *International Perspectives on Psychological Science*, *1*, 241–262. https://doi.org/10.4324/9780203845820-25

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in*

Cognitive Sciences, *3*(9), 337–344. https://doi.org/10.1016/S1364-6613(99)01362-5

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. https://doi.org/10.1111/j.2044-835X.1987.tb01048.x

Perner, J., Stummer, S., Sprung, M., & Doherty, M. (2002). Theory of mind finds its Piagetian perspective: Why alternative naming comes with understanding belief. *Cognitive Development*, *17*(3–4), 1451–1472. https://doi.org/10.1016/S0885-2014(02)00127-2

Peterson, C., Wellman, H. M., & Liu, D. (2005). Steps in theory-of-mind development for children with deafness or autism. *Child Development*, *76*(2), 502–517. https://doi.org/10.1111/j.1467-8624.2005.00859.x

Phillips, J., Ong, D. C., Surtees, A. D. R., Xin, Y., Williams, S., Saxe, R., & Frank, M. C. (2015). A second look at automatic theory of mind: Reconsidering Kovács, Téglás, and Endress (2010). *Psychological Science*, *26*(9), 1353–1367. https://doi.org/10.1177/0956797614558717

Poulin-Dubois, D., Polonia, A., & Yott, J. (2013). Is false belief skin-deep? The agent's eye status influences infants' reasoning in belief-inducing situations. *Journal of Cognition and Development*, *14*(1), 87–99. https://doi.org/10.1080/15248372.2011.608198

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

Poulin-Dubois, D., & Yott, J. (2017). Probing the depth of infants' theory of mind: Disunity in performance across paradigms. *Developmental Science*, *21*(4), e12600. https://doi.org/10.1111/desc.12600

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

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral & Brain Sciences*, *1*, 515–526. https://doi.org/10.1017/S0140525X00076512

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

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230–236. https://doi.org/10.1016/j.cognition.2010.08.003

Rakoczy, H., Bergfeld, D., Schwarz, I., & Fizke, E. (2015). Explicit theory of mind Is even more unified than previously assumed: Belief ascription and understanding aspectuality emerge together in development. *Child Development*, *86*(2), 486–502.

https://doi.org/10.1111/cdev.12311

Rakoczy, H., Warneken, F., & Tomasello, M. (2007). “This way!”, “No! That way!”-3-year olds know that two people can have mutually incompatible desires. *Cognitive Development*, *22*(1), 47–68. https://doi.org/10.1016/j.cogdev.2006.08.002

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology*, *33*(1), 12–21. https://doi.org/10.1037/0012-1649.33.1.12

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience*, *11*(4), 264–274. https://doi.org/10.1038/nrn2805

Rowe, A. D., Bullock, P. R., Polkey, C. E., & Morris, R. G. (2001). “Theory of mind” impairments and their relationship to executive functioning following frontal lobe excisions. *Brain*, *124*(3), 600-616. https://doi.org/10.1093/brain/124.3.600

Rubio-Fernández, P. (2018). What do failed (and successful) replications with the Duplo task show? *Cognitive Development*, *48*, 316–320. https://doi.org/10.1016/j.cogdev.2018.07.004

Ruffman, T. (2014). To belief or not belief: Children’s theory of mind. *Developmental Review*, *34*(3), 265–293. https://doi.org/10.1016/j.dr.2014.04.001

Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology*, *80*, 201–224. https://doi.org/10.1006/jecp.2001.2633

Ruffman, T., Puri, A., Galloway, O., Su, J., & Taumoepeau, M. (2018). Variety in parental use of “want” relates to subsequent growth in children’s theory of mind. *Developmental Psychology*, *54*(4), 677–688. https://doi.org/10.1037/dev0000459

Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children’s and mothers’ mental state language and theory-of-mind understanding. *Child Development*, *73*(3), 734–751. https://doi.org/10.1111/1467-8624.00435

Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology*, *30*(1), 87–104. https://doi.org/10.1111/j.2044-835X.2011.02045.x

Sabbagh, M. A., & Bowman, L. C. (2018). Theory of mind. In J. Wixted (Ed.), *Stevens’ handbook of experimental psychology and cognitive neuroscience* (pp. 249–288). https://doi.org/doi.org/10.1002/9781119170174.epcn408.

Sabbagh, M. A., & Paulus, M. (2018). Replication studies of implicit false belief with infants and toddlers. *Cognitive Development*, *46*, 1–3. https://doi.org/10.1016/j.cogdev.2018.07.003

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Scott, S. E. B. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception Performance*, *36*(5), 1255-1266.

https://doi.org/10.1037/a0018729

Santiesteban, I., Catmur, C., Hopkins, C. S., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing ? *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 929–937. https://doi.org/10.1037/a0035175

Scarf, D., & Ruffman, T. (2017). *Great Apes' Insight into the Mind: How Great? eLetter on C. Krupenye, F. Kano, S. Hirata, J. Call, M. Tomasello, Great apes anticipate that other individuals will act according to false beliefs. Science 354, 110 (2016).* Retrieved from http://science.sciencemag.org/content/354/6308/110/tab-e-letters

Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others' mental states. *Journal of Experimental Psychology: General*, *141*, 433–438. https://doi.org/10.1037/a0025458

Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts implicit theory-of-mind processing. *Psychological Science*, *23*, 842–847. https://doi.org/10.1177/0956797612439070

Schneider, D., & Low, J. (2016). Efficient versus flexible mentalizing in complex social settings: Exploring signature limits. *British Journal of Psychology*, *107*(1), 26–29. https://doi.org/10.1111/bjop.12165

Schneider, D., Nott, Z. E., & Dux, P. E. (2014). Task instructions and implicit theory of mind. *Cognition*, *133*(1), 43–47. https://doi.org/10.1016/j.cognition.2014.05.016

Schneider, D., Slaughter, V. P., & Dux, P. E. (2017). Current evidence for automatic Theory of Mind processing in adults. *Cognition*, *162*, 27–31. https://doi.org/10.1016/j.cognition.2017.01.018

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

Scott, R. M. (2014). Post hoc versus predictive accounts of children's theory of mind: A reply to Ruffman. *Developmental Review*, *34*(3), 300–304. https://doi.org/10.1016/j.dr.2014.05.001

Scott, R. M. (2017). The Developmental Origins of False-Belief Understanding. *Current Directions in Psychological Science*, *26*(1), 68–74. https://doi.org/10.1177/0963721416673174

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. https://doi.org/10.1111/j.1467-8624.2009.01324.x

Scott, R. M., & Baillargeon, R. (2014). How fresh a look? A reply to Heyes. *Developmental Science*, *17*(5), 660–664. https://doi.org/10.1111/desc.12173

Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*, *21*(4), 237–249. https://doi.org/10.1016/j.tics.2017.01.012

Scott, R. M., He, Z., Baillargeon, R., & Cummins, D. (2012). False-belief understanding in 2.5-year-olds: Evidence from two novel verbal spontaneous-response tasks. *Developmental Science*,

15(2), 181–193. https://doi.org/10.1111/j.1467-7687.2011.01103.x

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. https://doi.org/10.1016/j.cogpsych.2015.08.003

Scott, R. M., Roby, E., & Baillargeon, R. (n.d.). How sophisticated is infants' theory of mind? In O. Houdé & G. Borst (Eds.), *Cambridge handbook of cognitive development.* Cambridge, England: Cambridge University Press.

Selcuk, B., Brink, K. A., Ekerim, M., & Wellman, H. M. (2018). Sequence of theory-of-mind acquisition in Turkish children from diverse social backgrounds. *Infant and Child Development*, *27*(4), 1–14. https://doi.org/10.1002/icd.2098

Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-months-olds really attribute mental states to others? A critical test. *Psychological Science*, *22*(7), 878–880. https://doi.org/10.1177/0956797611411584.Do

Shahaeian, A., Peterson, C., Slaughter, V., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Developmental Psychology*, *47*(5), 1239–1247. https://doi.org/10.1037/a0023899

Shepard, R. N., & Metzler, J. (1971). Mental Rotation of three-dimensional objects. *Science*, *171*, 701–703. https://doi.org/10.1126/science.171.3972.701

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. https://doi.org/10.1037/0033-295X.84.2.127

Shweder, R., & Sullivan, M. (1993). Cultural psychology: Who needs it? *Annual Review of Psychology*, *44*(1), 497–523. https://doi.org/10.1146/annurev.psych.44.1.497

Sinigaglia, C., & Rizzolatti, G. (2011). Through the looking glass: Self and others. *Consciousness and Cognition*, *20*(1), 64–74. https://doi.org/10.1016/j.concog.2010.11.012

Sodian, B. (2011). Theory of mind in infancy. *Child Development Perspectives*, *5*(1), 39–43. https://doi.org/10.111/j.1750-8606.2010.00152.x

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, *13*, 907–912. https://doi.org/10.1111/j.1467-7687.2009.00946.x

Southgate, V., Johnson, M. H., El Karoui, I., & Csibra, G. (2010). Motor system activation reveals infants' on-line prediction of others' goals. *Psychological Science*, *21*(3), 355–359. https://doi.org/10.1177/0956797610362058

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*(7), 587–592. https://doi.org/10.1111/j.1467-9280.2007.01944.x

Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*,

130(1), 1–10. https://doi.org/10.1016/j.cognition.2013.08.008

Sternberg, S. (2004). Reaction-time experimentation. *Psychology, 600, 301.* Retrieved from http://www.psych.upenn.edu/~saul/rt.experimentation.pdf

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, *18*(7), 580–586. https://doi.org/10.1111/j.1467-9280.2007.01943.x

Surtees, A. D. R., Apperly, I. A., & Samson, D. (2013a). Similarities and differences in visual and spatial perspective-taking processes. *Cognition*, *129*(2), 426–438. https://doi.org/10.1016/j.cognition.2013.06.008

Surtees, A. D. R., Apperly, I. A., & Samson, D. (2013b). The use of embodied self-rotation for visual and spatial perspective-taking. *Frontiers in Human Neuroscience*, *7*(November), 698. https://doi.org/10.3389/fnhum.2013.00698

Surtees, A. D. R., Apperly, I. A., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, *150*, 43–52. https://doi.org/10.1016/j.cognition.2016.01.014

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

Surtees, A. D. R., Samson, D., & Apperly, I. A. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, *148*, 97–105. https://doi.org/10.1016/j.cognition.2015.12.010

Taumoepeau, M., & Ruffman, T. (2006). Mother and infant talk about mental states relates to desire language and emotion understanding. *Child Development*, *77*(2), 465–481. https://doi.org/10.1111/j.1467-8624.2006.00882.x

Taumoepeau, M., & Ruffman, T. (2008). Stepping stones to others' minds: Maternal talk relates to child mental state language and emotion understanding at 15, 24, and 33 months. *Child Development*, *79*(2), 284–302. https://doi.org/10.1111/j.1467-8624.2007.01126.x

Thillay, A., Lemaire, M., Roux, S., Houy-Durand, E., Barthélémy, C., Knight, R. T., … Bonnet-Brilhault, F. (2016). Atypical brain mechanisms of prediction according to uncertainty in autism. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00317

Thoermer, C., Sodian, B., Vuori, M., Perst, H., & Kristen, S. (2012). Continuity from an implicit to an explicit understanding of false belief from infancy to preschool age. *British Journal of Developmental Psychology*, *30*(1), 172–187. https://doi.org/10.1111/j.2044-835X.2011.02067.x

van der Wel, R. P. R. D., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition*, *130*(1), 128–133. https://doi.org/10.1016/j.cognition.2013.10.004

Wang, B., Hadi, N. S. A., & Low, J. (2015). Limits on efficient human mindreading: Convergence across Chinese adults and Semai children. *British Journal of Psychology*, *106*(4), 724–740.

https://doi.org/10.1111/bjop.12121

Wellman, H. M. (2014). *Making minds: How theory of mind develops.* Oxford University Press.

Wellman, H. M. (2018). Theory of mind: The state of the art. *European Journal of Developmental Psychology*, *15*(6), 728–755. https://doi.org/10.1080/17405629.2018.1435413

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of Theory-of-Mind development: The truth about false belief. *Child Development*, *72*(3), 655–684.

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind understandings in Chinese children. *Psychological Science*, *17*(12), 1075–1081. https://doi.org/10.1111/j.1467-9280.2006.01830.x

Wellman, H. M., Fang, F., & Peterson, C. (2011). Sequential progressions in a theory-of-mind scale: Longitudinal perspectives. *Child Development*, *82*(3), 780–792. https://doi.org/10.1111/j.1467-8624.2011.01583.x

Wellman, H. M., Kushnir, T., Xu, F., & Brink, K. A. (2016). Infants use statistical sampling to understand the psychological world. *Infancy*, *21*(5), 668–676. https://doi.org/10.1111/infa.12131

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Whiten, A. (2013). Humans are not alone in computing how others see the world. *Animal Behaviour*, *86*(2), 213–221. https://doi.org/10.1016/j.anbehav.2013.04.021

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *3*, 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34. https://doi.org/10.1016/S0010-0277(98)00058-4

Xu, F., Carey, S., & Welch, J. (1996). Infants' ability to use object kind information for object individuation. *Cognition*, *70*, 137–166. https://doi.org/10.1016/S0010-0277(99)00007-4

Yott, J., & Poulin-Dubois, D. (2016). Are infants' theory-of-mind abilities well integrated? Implicit understanding of intentions, desires, and beliefs. *Journal of Cognition and Development*, *17*(5), 683–698. https://doi.org/10.1080/15248372.2015.1086771

Zawidzki, T. W. (2013). *Mindshaping: A new framework for understanding human social cognition*. Cambridge, MA: MIT Press.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37–63. https://doi.org/https://doi.org/10.1016/S0885-2014(96)90027-1

Zeman, S. (2017). Confronting perspectives: Modeling perspectival complexity in language and cognition. *Glossa: A Journal of General Linguistics*, *2*(1), 6. https://doi.org/10.5334/gjgl.213