



The
University
Of
Sheffield.

Can Infants Reason About Beliefs?

Francesco Antilici

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

The University of Sheffield
Faculty of Arts and Humanities
Department of Philosophy

February 2018

ABSTRACT

At what point in development does the capacity to reason about what people think emerge? While developmental psychologists have been investigating this question for more than thirty years, the evidence they have gained so far is conflicting. On the one hand, the results of traditional, direct false-beliefs tests, which involve asking participants how a person with a false belief will act, suggest that most children under four years of age are still unaware that beliefs can be false. On the other, false-belief tests using indirect measures, such as, for example, looking times or anticipatory looking, suggest that even infants ascribe false beliefs to other people. As many have noted, these results pose a deep developmental puzzle.

In this work, I defend the claim that infants can already reason about beliefs. On the one hand, I argue that alternative interpretations of indirect false-belief tests fall short of the mark. On the other, I argue that the fact that young children fail direct false-belief tests can be explained in either of two ways, both of which are compatible with the claim that the capacity to reason about beliefs emerges early on. The first option is to maintain that young children fail because of performance difficulties. This type of position has been defended by other authors, but I'll argue that the particular proposal I put forward (which I call the *processing-time account*) offers a better account of the evidence. In contrast, the second option (which I call the *hybrid approach*) is one that, to the best of my knowledge, no one else has defended so far. This consists in arguing that direct and indirect false-belief tests recruit distinct cognitive systems, each of which can independently sustain the ability to reason about beliefs, but which follow different developmental trajectories.

After laying out these two options, in the last chapter I will also consider which is best supported by the evidence.

CONTENTS

I. INTRODUCTION	10
1. A Puzzle about Development	11
1.1 Direct False-Belief Tests	11
1.2 Indirect False-Belief Tests	13
2. Putting the Puzzle in Context	18
2.1 Other Indirect Mentalising Studies	19
2.2 Other Direct Mentalising Studies	24
3. The Nativism-Constructivism Debate	27
3.1 What is the debate about?	27
3.2 Nativism and the Early-emergence Approach	28
3.3 Constructivism and the Late-emergence approach	29
4. Wider Implications	30
4.1 Consciousness	30
4.2 Communication	31
4.3 Sociomoral Abilities	32
5. Plan of the Thesis	32
II. ALTERNATIVE INTERPRETATIONS OF INDIRECT FALSE-BELIEF STUDIES	34
1. Low-Level Proposals	35
1.1 The Three-way Association Account	35
1.2 The Low-Level Novelty Account	38
1.3 Summary	41
2. Behavioural Proposals	41
2.1 The Behaviour-Rule Account	41
2.2 The Affordance-based Account	47
2.3 The Enhanced Teleological-Reasoning Account	49

2.4 The Association Account	51
2.5 Summary	52
3. Mentalistic Proposals	53
3.1 The Desire-Awareness Account	53
3.2 The Perceptual Access Reasoning Account	55
3.3 The Minimal Theory of Mind Account	56
3.4 Summary	60
4. Conclusions	61
III. ALTERNATIVE INTERPRETATIONS OF DIRECT FALSE-BELIEF STUDIES	62
1. Additional Evidence for Late-Emergence Accounts	63
1.1 Developmental Progression	63
1.2 Executive Functioning	63
1.3 Language Ability	67
1.4 Experiential Factors	68
1.5 Deaf Children	70
1.6 Social Competence	71
1.7 Brain Maturation	73
2. Performance Accounts	74
2.1 The ToMM/SP Account	75
2.2 The Expanded Processing Demands Account	84
2.3 The Triple-Load Account	87
2.4 The Pragmatic Perspectives Account	91
2.5 The Pragmatic Development Account	95
3. Conclusions	100
IV. THE PROCESSING-TIME ACCOUNT	101
1. The Account	102
1.1 Processing Time	102
1.2 Factors Affecting RTP	104
1.3 Factors Affecting TAP	107

2. Evidence for Performance Accounts	108
2.1 Executive Functioning	109
2.2 Response-Generation Processes	112
2.3 Pragmatic Interpretation	113
2.4 Summary	116
3. Explaining Away Evidence for Late-Emergence Accounts	116
3.1 Brain Maturation	116
3.2 Executive Functioning	117
3.3 Language Ability	121
3.4 Social Competence	123
3.5 Developmental Progression	125
3.6 Experiential factors	127
3.7 Deaf Children	127
3.8 Summary	128
4. Conclusions	129
V. THE HYBRID APPROACH	130
1. The Hybrid Approach	131
1.1 A New Approach	131
1.2 The Late-Emergence Approach and the Distinct-System View	132
1.3 Explaining Evidence for Late-Emergence Accounts	135
1.4 An Argument for the Hybrid Approach	135
1.5 The Hybrid Approach and the Nativism-Constructivism Debate	136
2. Two Mentalising Systems	138
2.1 Evidence for Two Mentalising Systems	138
2.2 The Argument from Cognitive Demands	144
2.3 Explaining Evidence of Signature Limits	149
2.4 The Argument from Higher-Order States	153
3. Conclusions	156
VI. SHOWDOWN: HYBRID APPROACH V. PROCESSING-TIME ACCOUNT	158

1. One or Two Mentalising Systems?	159
1.1 Parsimony	159
1.2 Evidence for Two-Systems Accounts	159
1.3 Misalignment of Processing Features	161
2. Same System or Distinct Systems?	167
2.1 Is the Distinct-Systems View <i>Ad Hoc</i> ?	167
2.2 Evidence for Performance Accounts	170
2.3 Evidence from Neuroscience	172
3. Conclusions	175
VII. CONCLUSIONS	176

I

INTRODUCTION

Mentalising (also known as folk psychology, theory of mind or mindreading) is the capacity to reason about what other people think, want or feel; it is one of the pillars around which human social life is built. Humans seem to spend a significant portion of their waking life wondering about the mental lives of their con-specifics, ruminating over each other's past actions or trying to anticipate their next moves.

In this work, I focus on belief-reasoning (i.e., the capacity to reason about beliefs) and its development. While developmental psychologists have been investigating this question for more than thirty years, the evidence they have gained so far is conflicting. On the one hand, the results of traditional, direct false-beliefs tests, which involve asking participants how a person with a false belief will act, suggest that most children under four years of age are still unaware that beliefs can be false. On the other, false-belief tests using indirect measures, such as, for example, looking times or anticipatory looking, suggest that even infants ascribe false beliefs to other people. As many have noted, these results pose a deep developmental puzzle. Does an understanding of belief emerge early, already being present in infancy, or does it emerge later on, around the fifth year of age? This is the topic of the present work.

In §1, I review the results of direct and indirect false-belief tests. Then, to provide some context, in §2 I offer a brief overview of the rest of the evidence on the development of mentalising. In §3, I summarise the implications of our topic for the debate between nativists and

constructivists. In §4, I discuss its implications for other topics in philosophy and cognitive science. Finally, in §5, I lay out the plan for the rest of the thesis.

1. A Puzzle about Development

1.1 Direct False-Belief Tests

1.1.1 The Original False-Belief Test

The first false-belief test was carried out by Hanz Wimmer and Josef Perner (1983). Children were told a story about Maxi, a boy who really liked chocolate. In the story, Maxi's Mom comes back from shopping and she has bought ingredients to bake a cake, including chocolate. Maxi helps his Mom put the ingredients away; they put the chocolate in the blue cupboard. After helping his Mom, Maxi goes outside to play. While Maxi is outside, his Mom uses some of the chocolate, then puts it away in a different cupboard, this one green. At this point Maxi's Mom realises she forgot to buy eggs, so she goes to her neighbour to ask whether she can have some of hers. While his Mom is away, Maxi is feeling hungry, so he comes back in the house to eat some chocolate. At this point children were asked where Maxi would look for the chocolate. Notice that answering correctly requires understanding that Maxi does not know his Mom put the chocolate in the green cupboard, and thus still thinks the chocolate is in the blue cupboard, where they had put it previously. Most of the three- and four-year-olds in Wimmer and Perner's study pointed to the green cupboard; about half of the five-year-olds pointed to the green cupboard, while the other half pointed to the blue cupboard; almost every six-, eight- and nine-year-old pointed to the blue cupboard. As we will see, most of the false-belief studies that have been carried out since have found better performance in four- and five-year-olds, though, crucially, not in three-year-olds.

1.1.2 Other False-Belief Tasks

The type of task used by Wimmer and Perner is often referred to as the unexpected-transfer or, alternatively, change-of-location task. There are a few other tasks that are commonly used in false-belief tests, detailed below.

Unexpected-contents task. The unexpected-content task was first used in a study by Perner, Susan Leekam and Wimmer (1987). In this study, children were shown a box of Smarties and asked what was inside. After children said the box contained Smarties, the experimenter opened it and allowed them to peek inside it, thus revealing that it contained crayons instead. At this point, the experimenter asked what another person, who had not yet looked inside the box, would think is in it. To answer correctly, participants must understand that the person in question, not having looked inside the box, will assume it contains Smarties.

Appearance-reality task. The appearance-reality task was introduced by John Flavell, Eleanor Flavell and Frances Green (1986). Children were shown a deceptive object (a sponge that looked like a rock) and asked what it looked like. After children identified the apparent nature of the toy (“it’s a rock!”), they were allowed to touch it, discovering its true nature. Children were then asked what the object really was (e.g., a sponge) and what it looked like to them now (e.g. a rock). Notice that, strictly speaking, this is not a false-belief task, since children are not asked about anybody’s beliefs. Nonetheless, several psychologists have argued that an understanding of misleading appearances presupposes an understanding of the representational nature of mind; as a result, the appearance-reality tests is often used in tests that include batteries of false-belief tests.

Unexpected-identity task. The unexpected-identity task, first used by Alison Gopnik and Janet Astington (1988), is a variation on the appearance-reality task. As in the latter, children are shown a deceptive object, say a sponge that looks like a rock, and asked what it is; after they say it is a rock, they are allowed to inspect it. Then, children are asked: (1) what the object really is, and what it looks like; (2) whether another person, who has seen but not touched the object, will take it to be a sponge or a rock; (3) what they themselves thought it was before touching it. To answer correctly, participants must understand that someone who has no touched the object has no way of knowing its true nature and will thus take it to be what it looks like.

The explicit false-belief task. In the explicit false-belief task, introduced by Henry Wellman and Karen Bartsch (1988), children are told a brief story about a girl, Jane, who wants to find her kitten. The kitten is in the playroom, but Jane thinks it’s in the kitchen. Children are then asked where Jane will go look for her kitten.

1.1.3 A Meta-Analysis of False-Belief Studies

In 2001, Wellman, David Cross and Julanne Watson published a meta-analysis of false-belief tests, which looked at the results of 178 separate studies, using a variety of tasks. The results of the meta-analysis show that children's performance improves steadily over the preschool years, with young three-year-olds being below-chance and four-year-olds being above-chance. Some task manipulations were found to increase children's performance, allowing them to pass the test a few months in advance; for example, children were found to perform better if they had played an active role in the change of location that resulted in the agent's false belief (some of these results will be discussed further in Chapter IV, §2). Nonetheless, young children still performed worse than their older peers. Similarly, performance was found to vary across different countries; American children, for example, performed worse than their Australian peers but better than the Japanese; still, in every culture performance increased with age, moving from below-chance to above-chance.

1.1.4 The Late-Emergence Approach

Many have taken the results of direct false-belief tests to show that belief-reasoning abilities only emerge around the age of four. Call this the late-emergence approach. For many years, the late-emergence approach was the majority view in developmental psychology. As we are about to see, however, recent findings challenge this point of view.

1.2 Indirect False-Belief Tests

1.2.1 Infants Pass a False-Belief Test

In a ground-breaking report published in 2005, Kristine H. Onishi and Renée Baillargeon showed that even 15-month-olds could pass a false-belief test when indirect measures were used (Onishi & Baillargeon, 2005). Since simply asking such young infants what someone would do was not an option, Onishi and Baillargeon adapted the unexpected-transfer task to the violation-of-expectation paradigm. Since infants tend to look longer at events that are, for them, unexpected, by comparing how long infants look at each of two possible events one can infer which is more in line with their expectations. As infants watched, an experimenter (E1) played with a toy watermelon slice, then put it in a green box and left. While E1 was absent, the toy moved from the green box

into a yellow box next to it. In the test phase, after E1 had come back, half the infants saw her reaching for the green box, where she had put the toy, while the other half saw her reaching for the yellow box, where the toy really was. Looking times were longer for the second group, suggesting that infants had understood E1 thought the toy was still in the green box, where she had left it, and thus expected her to reach there.

Since 2005, many experiments have been carried out in several different laboratories, replicating or extending the initial finding by Onishi and Baillargeon. Their result has been extended in three important ways: (1) first, evidence of belief-reasoning in infants has been obtained with a variety of experimental paradigms besides the violation-of-expectation method; (2) second, positive results have been obtained with much younger infants, as young as six months old; and (3) third, positive results have been obtained with a variety of belief-reasoning tasks.

1.2.2 A Variety of Experimental Paradigms

Evidence has been obtained with a variety of experimental paradigms, including violation-of-expectation, anticipatory-looking, interactive, preferential-looking and two EEG-based paradigms. Let us zoom in on each of these paradigms in turn.

Violation-of-expectation. Most indirect false-belief studies have employed the violation-of-expectation method, described above. These include: Kovács, Téglás, and Endress (2010), Luo (2011a), Onishi and Baillargeon (2005), Song and Baillargeon (2008); Song, Onishi, Baillargeon, and Fisher (2008), Scott and Baillargeon (2009), Scott, Baillargeon, Song, and Leslie (2010), Scott, Richman, and Baillargeon (2015), Surian, Caldi, and Sperber (2007), Träuble, Marinović, and Pauen (2010).

Anticipatory-looking. Several indirect false-belief studies have used the anticipatory-looking paradigm, which takes as a dependent measure where infants look first, and longer, in anticipation of the target agent's reach (Clements & Perner, 1994; He, Bolz, & Baillargeon, 2012; Senju, Southgate, & Snape, 2011; Southgate, Senju, & Csibra, 2007; Surian & Geraci, 2012). If, for example, participants look first and longer at the green box as opposed to the yellow one, this is taken to show that they expect the agent to look there.

Interactive. Several indirect false-belief studies have used the interactive paradigm (D. Buttelmann, Carpenter, & Tomasello, 2009; D. Buttelmann, Over, Carpenter, & Tomasello, 2014;

F. Buttelmann, Suhrke, & Buttelmann, 2015; Knudsen & Liszkowski, 2012a, 2012b; Southgate, Chevallier, & Csibra, 2010). In this paradigm, participants are expected to help a target agent, who has a false belief; the tasks are set up in such a way that, given what infants do in trying to help the agent, it is possible to infer whether they are aware of her false belief. This is best illustrated with an example. Consider the experiment by David Buttelmann and colleagues (2009). An agent (E1) tries to retrieve a toy from the box where she had put it (i.e., the pink box). Unbeknownst to her, however, a second agent (E2) moved the toy to a second box (i.e., the yellow box.) and locked both boxes with pins. Thus, E1 is now struggling to open the pink box, even though the toy she is looking for is really in the yellow box. Infants are encouraged to help E1. If they are aware that E1 believes the toy to be in the pink box, they should realise she is trying to get the toy and should thus point or crawl to the yellow box, where the toy really is. If, on the other hand, they are not aware of E1's false belief, then they should assume E1 has her reasons for opening the pink box and should thus try to help her do that. Buttelmann and colleagues report that the sixteen- and eighteen-month-olds in their study approached the pink box more often than the yellow box, suggesting they knew E1 had a false belief.

Preferential-looking. One study has used the preferential-looking paradigm (Scott, He, Baillargeon, & Cummins, 2012). An experimenter narrated a false-belief story (of the unexpected-transfer type) while participants looked at a picture book depicting the events in the story. The story ended with a vague statement (e.g., "Sally is looking for her marble") and two pictures, each depicting one of the two possible outcomes (e.g., Sally looking either in the basket, where she had put the marble, or in the box, where the marble actually was.) The two-year-olds in this study looked mostly at the picture depicting Sally looking in the basket, suggesting they expected her to look there, consistently with her false belief.

Predictive neural activation. A recent study by Victoria Southgate and Angelina Vernetti has used a novel paradigm involving EEG which, for lack of a better term, I will call predictive neural activation paradigm (Southgate & Vernetti, 2014). It is well known that, when people expect someone to engage in a goal-directed action, their motor cortex exhibits heightened activation (Muthukumaraswamy, Johnson, & McNair, 2004); furthermore, in infants, motor cortex activation has been shown to correlate with a decrease in the amplitude of the alpha rhythm over their motor cortex, which can be detected using EEG (Southgate, Johnson, El Karoui, & Csibra, 2010).

Building on these previous findings, Southgate and Verneti modified the unexpected-transfer task so that the event to be predicted was whether the agent would or would not reach for the object in a box; they then used EEG to look at whether participants exhibited alpha suppression. Specifically, in the “FB-ball” condition, the agent (E1) watched as a ball rolled into a box; then, while E1 was away, the ball rolled out. Thus, in this condition, E1 thought the ball was in the box while in fact it was not. In the “FB-no-ball” condition, the opposite happened: E1 watched as the ball rolled out of the box, then, while E1 was away, the ball rolled back into it. Thus, in this case, E1 thought the box was empty while in fact was not. The six-month-olds in this study expected E1 to reach in the FB-ball condition but not in the FB-no-ball condition, consistently with what her false beliefs.

Object representation. Another study, by Dora Kamps, Eugenio Parise, Gergely Csibra and Kovács Ágnes (2015) also introduced a novel paradigm involving EEG. A previous study had found that, when six-month-old infants witnessed the occlusion of an object as opposed to its disappearance, an increased gamma activation could be detected over the temporal areas (Kaufman, Csibra, & Johnson, 2003). Kaufman and colleagues took this increased activation to be the signature of a sustained object representation in the infant brain. In other words, according to Kaufman and colleagues, the increased temporal activation signals that infants are representing the object to be still there, albeit occluded. In their first study, Kamps and colleagues found that the same activation could be detected in eight-month-olds when the object was occluded from an agent’s perspective but not from the infant’s. Building on this finding, in their second study, Kamps and colleagues had the object disappear while occluded from the agent but not from the infants, thus making the agent’s belief that the object was still there false. Increased gamma activation was detected even in this case, suggesting that infants were representing the agent’s false belief.

Notice that all these different paradigms (violation-of-expectation, anticipatory-looking etc.) have something in common: at no point are participants asked what the target agent thinks, or what she will do. Rather, experimenters infer what participants think the target agent thinks in a more oblique, “indirect” way based on their looking behaviour, or the way they try to help the agent, or the patterns of activity in their brain. Consequently, I use the term “indirect” to describe these false-belief tests and distinguish them from the more traditional type of test that we saw in §1.1.

1.2.3 A Variety of Belief-Reasoning Tasks

In a typical unexpected-transfer task, the target agent's (e.g. Maxi) belief about the location of the object is based on her having last seen the object in that location. In contrast, some indirect false-belief tests have used tasks where the agent's belief is the result of either testimony or inference from other beliefs. Let us see a couple of examples.

- Song et al. (2008). This study involved two agents (E1 and E2), a toy (a ball) and two containers (a blue box and a red mug). Like Onishi and Baillargeon (2005), the study included a “familiarisation” trial, where E1 put the ball in the box, and a “belief-induction” trial, where E1 was absent and E2 moved the ball from the box into the mug. However, this study also included an “intervention” trial, where E2 told E1 either “the ball is in the mug” (informative-intervention condition) or “I like the mug” (uninformative-intervention condition). In the informative-intervention condition, infants expected E1 to reach for the mug, while in the uninformative-intervention condition they expected her to reach for the box, which is where E1 had put the ball in the first place.
- Song and Baillargeon (2008). This study involved two experimenters (E1 and E2) and two toys (a toy skunk and a doll with blue hair). During the familiarisation trials, E1 always reached for the doll, thus displaying her preference. During the belief-induction trial, two boxes were introduced: one had a plain lid (plain box), the other had a lid with a tuft of blue hair (like the doll's) sticking out from underneath (hair box). As E1 is away, E2 puts the skunk in the hair box, and the doll in the plain one. Infants expected E1 to reach for the hair box, seemingly understanding that she would mistake the tuft of blue hair to be the doll's.

1.2.4 Earliest Evidence of Belief-Reasoning Abilities

While most indirect false-belief studies have investigated the belief-reasoning abilities of infants in their second year of life, some have looked at much younger infants. Specifically, the two EEG studies mentioned in §1.2.2, by Southgate and Vennetti (2014) and Kamps et al. (2015), obtained positive results with six and eight-month-olds, respectively; a violation-of-expectation study by Kovács et al. (2010) found evidence of belief-reasoning in seven-month-olds; and a violation-of-expectation study by Luo (2011a) found evidence in ten-month-olds. Importantly, none of these

studies found negative results with younger infants. This is important because it means that infants have provided evidence of possessing belief-reasoning abilities as early as they have been tested.

1.2.5 The Puzzle

In contrast with the results of direct false-belief tests, these new findings suggest that belief-reasoning abilities are already present in infants. Call this the early-emergence approach. This conflicting pattern of results gives rise to the following puzzle: are belief-reasoning abilities already present in infants (as maintained by the early-emergence approach) or do they emerge later, around the age of four (as maintained by the late-emergence approach)? Investigating this issue will keep us busy for the rest of this work.

2. Putting the Puzzle in Context

Before we begin to discuss our puzzle, it is important to put the problem in context. To this end, in this section we are going to broaden our focus and take a look at the evidence concerning the development of other key folk psychological notions, besides belief. We will look first at the data from mentalising studies using indirect paradigms, in §2.2, and then move on, in §2.3, to those using direct methodologies.

2.1 Other Indirect Mentalising Studies

2.1.1 Intention

A few months after birth, infants already seem able to correctly identify the goal of simple actions, like grasping and reaching. That infants see action as directed to goals was shown, most notably, in a seminal visual-habituation study by Amanda Woodward (1998). Woodward habituated 5- and 9- month-old infants to a hand reaching for and grasping one of two toys, which were placed on two platforms next to each other. After the habituation phase, the two objects were moved so that they switched places. In the test events, the hand either moved along the same path, thus ending in contact with a different object (new-goal event), or reached for the same object as before, though moving along a different path (new-path event). Nine-month-olds and, to a lesser extent, five-month-olds, looked longer at the new-goal event, showing that they considered this event more novel than the other. This, in turn, suggests that the infants had categorised the reaching movement shown during the habituation phase in terms of its goal (e.g. grasping a particular toy) rather than merely in terms of its physical properties (e.g., its trajectory).

A violation-of-expectation study by Luo (2011b), using a similar task, suggests that an understanding of goal-directed action is already in place in three-month-olds. In the orientation event, a box moved back and forward, ostensibly self-propelled; this was done so that the infants would recognise the box as agentive. Then, in the familiarisation trials, the box stood at the centre of the stage, with a cone on its left and a cylinder on its right, before “reaching” for the cone; this event was repeated several times. In the test phase, the cone was now at the box’s right and the cylinder at its left, and the box reached either for the cone again (old-goal event) or for the cylinder (new-goal). Consistently with what Woodward had found, the three-month-olds in this study looked longer at the new-goal event (see also: Sommerville, Woodward, & Needham, 2005).

A series of studies by Gergely Csibra, Gyorgy Gergely and colleagues provide evidence that infants expect agents to choose means that are adequate and efficient to bring about their goals (Csibra, 2008; Csibra, Biró, Koós, & Gergely, 2003; Csibra, Gergely, Biró, Koós, & Brockbank, 1999; Gergely, Nádasdy, & Csibra, 1995). In the visual-habituation study by Csibra (2008), for example, infants were habituated to computer-generated events involving a self-propelled box

moving around an obstacle to reach a box on the other side of the room. In the test phase, the obstacle was removed, and the box either travelled around a similar path as before, making a detour around the spot where the obstacle had been (detour-path event), or followed a straight line to the other box (straight-path event). The six-month-olds in this study looked longer at the detour-path event, suggesting that they expected the agent to adjust her action to the changed situational constraints.

The results of visual-habituation and violation-of-expectation studies receive further corroboration by those of imitation studies, which show that infants do not simply mimic the agent's bodily movements but take into account her goal (Carpenter, Call, & Tomasello, 2005; Gergely, Bekkering, & Király, 2002; Hamlin, Hallinan, & Woodward, 2008; Meltzoff & Moore, 1997; Schwier, Van Maanen, Carpenter, & Tomasello, 2006; Zmyj, Daum, & Aschersleben, 2009). Consider for example Meltzoff (1995). Eighteen-month-olds watched as an experimenter attempted to perform an unfamiliar action involving a tool, such as, for example, pushing the button on the surface of a box using a stick. Since none of the attempts were successful, the infants in this study did not witness the outcome the agent intended to bring about. However, when they were given the chance to handle the tools themselves, instead of mimicking the exact movements of the agent most of the infants acted so as to bring about the outcome that those movements were aimed at; for instance, they used the stick to push the button.

2.1.2 Preference

Variations on the task introduced by Woodward (1998) have been used to investigate whether infants can attribute preferences to an agent. A natural interpretation of the findings reported by Woodward is that infants take the fact that the agent always picked the same of two objects as a sign that she had a preference for that one over the other. Luo and Baillargeon (2005) put this interpretation to the test with 5-month-olds, using a self-propelled box task similar to the one described above. In the familiarisation trials, the box “reaches” repeatedly for the cone on its right, while in the test event it reaches either for the cone, now on its left, or for the cylinder, on its right. In the no-preference condition, only the cone was present during the familiarisation trials, while in the preference condition, similarly to Woodward's original task, both objects were present throughout. If infants ascribe preferences, they should look longer at the new-goal event in the

preference condition but not in the no-preference condition. This is because the box reaching for the cone during the familiarisation trials only conveys a preference for the cone if the cylinder is also present. In the no-preference condition, infants should not be able to tell which of the two objects the box liked best and should thus have no expectations. On the other hand, if infants do not ascribe preferences but simply expect the box to repeat the same action it performed in the familiarisation trials, then they should look longer at the new-goal event in both conditions. Luo and Baillargeon found a significant difference in looking times only in the preference condition, thus suggesting that five-month-old infants possess an understanding of preference.

An early study by Repacholi and Gopnik (1997) looked at whether infants would be able to ascribe to other people preferences opposite to theirs. An experimenter tasted two foods, one which children typically like (crackers) and one which they typically dislike (broccoli). The experimenter tasted the foods and, using typical facial and verbal expressions, conveyed whether she liked them or not. In the “match” condition, the experimenter liked the crackers but disliked the the broccoli; in the “mismatch” condition, she liked the broccoli but disliked the crackers. After tasting the foods, the experimenter asked children if they could give her “some”. In the mismatch condition, most of the eighteen-month-olds passed the experimenter crackers in the match condition and broccoli in the mismatch condition. Most of the fourteenth-month-olds, however, gave her crackers in both conditions.

2.1.3 Seeing

A few violation-of-expectation studies suggest that infants understand which objects agents can or cannot see from where they stand (Luo & Baillargeon, 2007; Luo & Beck, 2010; Luo & Johnson, 2009). As mentioned above, the study by Luo and Baillargeon shows that infants do not expect the agent to reach again for the same object she reached for in previous trials when a second object is introduced in the test phase. A follow-up study by the same authors shows that, in ascribing a preference, six-month-old infants take into account only the objects the agent can see (Luo & Baillargeon, 2007). During the familiarisation trials, an experimenter sat at the centre and back of a stage, which had a box on the left side and a football on the right side. An occluder stood between the box and the experimenter. In the hidden-box condition, the occluder was tall enough to hide the box from the experimenter’s view; in the visible-short-box condition, the box was the

same size but the occluder was shorter; in the visible-tall-box condition, the occluder was the same size as in the hidden-box condition, but the box was now taller. As in previous studies, in the familiarisation trials the agent repeatedly reached for one of the two objects (the football). In the test trials, the occluder was removed and the objects were switched of position: the box was now on the right and the football on the left. The experimenter reached either for the football again (old-goal event) or for the box (new-goal event). The infants looked longer at the new-goal event in the visible-short-box and visible-tall-box conditions, but not in the hidden-box condition; they thus seemed to understand that, if the experimenter could not see the box, her reaching for the football could not be taken to express a preference.

Support for the claim that infants understand seeing is provided also by studies on gaze following. Although three-month-old infants already follow an agent's gaze (D'Entremont, Hains, & Muir, 1997), it is not until the tenth month that they start taking into account the eyes specifically, and not just the orientation of the head (Brooks & Meltzoff, 2005). Twelve-month-olds seem to understand the referential aspect of looking: they can identify which object the agent is looking at, and if a barrier prevents them (but not the agent) from seeing the object, they will lean and move in order to gain sight of it (Caron, Kiel, Dayton, & Butler, 2002; Moll & Tomasello, 2004; Wellman, Phillips, Dunphy-Lelii, & LaLonde, 2004). Furthermore, they understand that an agent's sight of an object can be impeded by an opaque barrier located between her and the object, and, by the fourteenth month, they understand that a blindfold can also prevent seeing (Brooks & Meltzoff, 2002). Interestingly, whereas twelve-month-olds usually follow the gaze of blindfolded agents as if they could see, they correct this behaviour if they are given the chance to wear the blindfold themselves, experiencing its effect on seeing; conversely, if eighteenth-month-olds are given the chance to wear a trick blindfold, which does not prevent seeing, they will start to follow the gaze of blindfolded agents more frequently (Meltzoff & Brooks, 2008). By fourteen months, infants are also capable of using gaze direction in order to identify the object an agent is 'emoting' about (Repacholi, Meltzoff, & Olsen, 2008) and can use this insight in order to predict which object the agent may grasp (Phillips, Wellman, & Spelke, 2002).

2.1.4 Knowledge

Some of the false-belief studies mentioned in section 1 included, as controls, conditions where the agent lacked the crucial piece of information as opposed to having it wrong (He, Bolz, & Baillargeon, 2011; Knudsen & Liszkowski, 2012a; Scott & Baillargeon, 2009; Scott et al., 2010). Consider for example the experiment by Zijing He and colleagues (2011). Two agents (E1 and E2) are playing the following game: E2 hides a toy in one of two boxes and E1 has to point to where it is. There were three conditions. In the knowledge condition, E1 is present as E2 hides the toy. In the ignorance condition, E1 is not present as E2 hides the toy. In the false-belief condition, E1 sees E2 hide the toy in one box, but E2 then moves it to the other box when E1 is absent. In the knowledge condition, He and colleagues found longer looking times when E1 pointed to the wrong box; in the false-belief condition, in contrast, they found longer looking times when E1 pointed to the correct box; and in the ignorance condition they found no significant difference in looking times.

That infants correctly predict the behaviour of agents that are unaware as opposed to wrong is also suggested by studies of informative pointing. For example, an early study by Daniela O'Neill (1996) shows that, if two-year-old toddlers need the help of an adult in order to retrieve a toy, they point more often to the location of the toy if the adult was not present, or had her eyes closed, as the toy was being put there. In a series of experiments, Ulf Liszkowski and his colleagues have shown that even twelve-month-old infants can use pointing to provide to others the information they need (Liszkowski, Carpenter, Striano, & Tomasello, 2006; Liszkowski, Carpenter, & Tomasello, 2007, 2008). For example, if an agent accidentally and unknowingly drops an object or misplaces it, and then starts looking for it, 12-month-old infants will point to the current location of the object; if the agent accidentally misplaces two objects, but is looking only for one of them, infants will point to that object specifically; if the agent misinterprets their message, infants will keep pointing and show signs of dissatisfaction.

Lastly, an experiment by Tomasello and Haberl (2003) suggests that infants keep track of which objects an agent is already acquainted with. Tomasello and Haberl let 12- and 18- month-old infants play with two new toys, together with two agents (Sally and Anne.) After a while Sally leaves the room and Anne and the infant start playing with a new toy. When Sally comes back, she looks at the three toys (placed all together) and asks, “Wow! That’s cool! Can you give it to me?”

Infants give to Sally the new toy more often in this condition than in a control condition in which Sally was present throughout the play.

2.2 Other Direct Mentalising Studies

2.2.1 Desires

Children seem display at least a basic understanding of desire as early as they have been tested with direct methodologies (Cassidy, 1998; Cassidy et al., 1995; Joseph & Tager-Flusberg, 1999; Lillard & Flavell, 1992; Moses, Coon, & Wusinich, 2000; Rakoczy, Warneken, & Tomasello, 2007; Wellman & Liu, 2004; Wellman & Woolley, 1990). For example, the first experiment in the study by Kimberly Cassidy, Maura Cosetti, Ressa Jones and colleagues (1995) suggests that young three-year-olds already know that other people can have different likes and dislikes from them. The three-year-olds in this study were shown four foods and asked to choose one they really liked (say, candy) and one they really did not like (say, broccoli). The children were then told a story about a boy (call him Tom). While Tom is at the grocery shop with his Mom, he tells her he really likes broccoli and hates candy (the opposite food preferences from the child). The next day, Tom's Mom comes back from shopping; she has bought broccoli and candy and tells Tom he can pick one to eat. Which will Tom pick? Most of the three-year-olds in this study answered correctly (e.g., saying Tom would eat the broccoli.)

Similarly, the second study in Cassidy et al. (1995) shows that three-year-olds can also infer preferences from past experiences. For example, in "implicit past-experience" trials, children were told about a child who had almost drowned while swimming at sea, but who had had fun having a picnic at the park. The next day, her mom asks her whether she would like to go on a picnic or swimming at the beach. Children were asked what the child will say. Even in this case, three-year-olds were above-chance.

2.2.2 Seeing

Concerning vision, Flavell and colleagues have introduced an important distinction between level-1 and level-2 perspective taking (Flavell, Everett, Croft, & Flavell, 1981; Masangkay et al., 1974). Level-1 perspective-taking consists in understanding which objects another person can see

from where she stands. Level-2 perspective-taking, in contrast, consists in understanding how an object will look to the other person.

Children pass level-1 perspective taking tests as early as they have been tested using direct methodologies. For example, Wellman, Ann Phillips and Thomas Rodriguez (2000) used a screen and two objects, each placed on a different side of the screen. After being allowed to inspect the set-up, children were sat on one side of the screen and asked which objects a person sitting on the other side could see. Most of the two-year-olds in this study judged correctly that the person could only see the object placed on her side of the screen.

In contrast, direct level-2 perspective taking tests are typically passed around the fourth birthday. In a typical level-2 perspective-taking tasks, children sit at a table with a picture of a turtle on it. Children are asked whether the experimenter, sitting opposite to them, sees the turtle as being right side up or upside down. Typically, three-year-olds are at chance (Masangkay et al., 1974).

2.2.3 Emotions

Children seem capable of ascribing simple emotions like happiness and sadness as early as they have been tested with direct methodologies (Hadwin & Perner, 1991; Rakoczy et al., 2007; Stein & Levine, 1989; Wellman et al., 2000; Wellman & Woolley, 1990). For example, in the task used by Hannes Rakoczy et al. (2007), children are told that two characters (Susi and Tom) are at a lake, on a boat together. Susi says: “we should go to that house”, but Tom says: “No, we should go to that tree!”; then, the wind blows the boat to one of the two locations, and children are asked whether Susi is happy or sad and whether Tom is happy or sad. Most of the three-year-olds in this study answered correctly.

For another example, consider the perception-to-emotion task used by Wellman and colleagues (2000). Participants were asked to judge how a person (Ann) would feel upon receiving, as a present, either an empty box or a box containing a desirable toy, and how they themselves would feel upon receiving the desirable toy as a present. All the three-year-olds in this study said both they and Ann would feel happy upon receiving the desirable toy; most of them also said Ann would be sad upon receiving the empty box.

2.2.4 The ToM Scale

The “ToM Scale” is a series of direct mentalising tests that children typically pass in a specific order; it was designed by Wellman and Liu (2004) to map out the specific developmental progression followed by children. The scale includes the following seven tasks.

1. *Diverse Desires.* Children are asked which of two snacks they prefer; for example, whether they like cookies or carrots best. The experimenter then tells them a short story involving a person (Mr. Jones) with the opposite preference, who is hungry, wants to have a snack, and must choose between cookies and carrots. Children are asked which food the character will choose.

2. *Diverse Beliefs.* Children are told about a person (Linda) who is looking for her cat, which could be in either of two locations (the garage or the bushes). Children are asked where they think the pet is. Whatever they say, they are told Linda thinks her cat is in the other location. Children are then asked where Linda will look for her cat.

3. *Knowledge Access.* Children are allowed to peek inside a container (e.g., a box or a drawer) to see what is inside; then, they are asked whether another person (Polly), who has not looked inside the container, knows what is in it.

4-5. *False Belief.* Wellman and Liu (2004) included two false-belief tasks in their Scale: an unexpected-contents task and an explicit false-belief task (both of which were described in §1.1.2).

6. *Belief-Emotion.* The belief-emotion task is a variation on the unexpected-contents task. Children are told that a boy (Teddy) loves Cheerios (an American brand of cereal) and are shown a Cheerios box with rocks inside it. Children are then asked how Teddy will feel when he is first given the box (happy, because he loves Cheerios) and how he will feel after he is allowed to look inside it (sad, because there are rocks, not Cheerios, inside the box).

7. *Real-Apparent Emotion.* Participants are told a story about a boy (Matt) who is being made fun of; the character does not find the joke funny but laughs all the same because he does not want to be called a baby. Participants are asked whether the character is happy or sad and whether she looks happy or sad.

Wellman and Liu (2004) and Wellman, Fang, and Peterson (2011) show that most children in Western societies pass the tests in this order.

3. The Nativism-Constructivism Debate

Why should one care about whether belief-reasoning abilities emerge early or late? A good reason to care about it is because of its implications for the debate between nativism and constructivism (other reasons will be discussed in §4). After clarifying what this debate is about in §3.1, in §3.2 and §3.3 we will consider how the early-emergence and late-emergence approaches, respectively, relate to nativism and constructivism.

3.1 What is the debate about?

The question of how much of our knowledge is innate may be as old as philosophy itself – it was first brought under the spotlight by Plato, and then re-examined centuries later in the works of philosophers like Descartes, Locke, Hume and Leibniz (Cowie, 1999; Samet, 2008; Simpson, Carruthers, Laurence, & Stich, 2005). Empiricism, the doctrine that all knowledge is learned from experience, came out as a clear winner from this second round of the debate, and was the dominating view for the best part of the two centuries that followed. The innateness controversy is currently experiencing a second revival, however, thanks to a large extent to Chomsky’s seminal work in linguistics (e.g. Chomsky, 1959). Crucially, for the first time in history, cognitive science has now given us the tools we need to settle the question experimentally.

At first blush, nativism is simply the claim that most of our knowledge is innate, while constructivism (the modern-day incarnation of empiricism) is the claim that most of our knowledge is learned (or “constructed”) from experience. Indeed, it is common to see nativism defined precisely along these lines (e.g., Prinz, 2002; Samet & Zaitchik, 2017). On closer inspection, however, the debate appears to be more nuanced (Carruthers, Laurence, & Stich, 2007; Margolis & Laurence, 2012; Pinker, 1997). A better way to bring the disagreement between nativists and constructivists into focus may be to look at the type of cognitive mechanisms that, according to each type of position, allow organisms to acquire the psychological traits they possess. Nativists about a given psychological trait tend to argue that the trait in question is acquired thanks to the contribution of domain-specific cognitive mechanisms (sometimes called *modules*) that contain a good deal of innate structure. In contrast, a constructivist would argue that the trait is acquired thanks to a domain-general learning mechanism.

Notice that one can be a nativist about certain traits but a constructivist about others. In addition, a cognitive mechanism can be specialised and innate to varying degrees, and in principle these two features may vary independently of each other. As a result, nativism and constructivism should not be understood to be clearly demarcated, mutually exclusive alternatives. A good way to visualise the debate is to imagine specialisation and innateness as two knobs that can be turned up or down, independently of each other. Turn them both all the way up, and you get radical forms of nativism; turn them both all the way down, and you get radical forms of constructivism; in between, lots of intermediate positions, some of which will be hard to classify as belonging to either camp.

3.2 Nativism and the Early-emergence Approach

The early-emergence approach has been defended, pretty much without exception, by researchers belonging to the nativist camp. The reason for this should be obvious: the early emergence of a trait is good evidence that the trait is either innate or acquired thanks to a specialised mechanism, allowing the organism to acquire the trait quickly, with relatively little exposure to the stimuli. Let us see some examples of early-emergence accounts in the literature.

3.2.1 Alan Leslie

Alan Leslie (1987) was among the first cognitive scientists to defend a nativist position about mentalising, and his account has been an inspiration for many of the early-emergence account that followed. He argued that representing propositional attitudes (like pretend or believe) requires the contribution of a special-purpose, domain-specific cognitive mechanism, which he called *Theory of Mind Mechanism* (ToMM for short). Leslie argued that ToMM was innate and would typically come online by the second year of life. This, notice, was several years before the results of the first indirect false-belief test were published; Leslie's claim was based mainly on the observation that young children engage in pretend play which, he argued, requires representing that one is pretending. In a later publication, Scholl and Leslie (1999) argued that ToMM was likely to have many of the properties traditionally associated with modules, including informational encapsulation, localised neural implementation, and disposition to selective impairment. At the same time, however, Scholl

and Leslie stressed that nothing in their position precluded the possibility of children acquiring new knowledge about the mind through learning (see also: Leslie, 2000a).

3.2.2 Renée Baillargeon

Based on the results of several indirect false-belief studies, Baillargeon, Scott, and He (2010) argue that humans possess an innate mentalising system, allowing them to interpret behaviour in terms of the underlying mental states. They argue that this system comprises two subsystems that may come online at different times. The first subsystem (SS1) would allow one to represent motivational states (like desires and intentions) and reality-congruent informational states (like true belief or absence thereof) but not reality-incongruent informational states (like false belief); to represent the latter, the second subsystem (SS2) would be needed. This is because, Baillargeon and colleagues argued, representing false beliefs requires not just “masking” the information the agent does not have but maintaining two contradictory representations of reality (one’s own and the agent’s).

3.3 Constructivism and the Late-emergence approach

The late-emergence approach has been defended mainly, although not exclusively, by constructivists. Let us look at some examples.

3.3.1 Alison Gopnik and Henry Wellman

Gopnik and Wellman (1992, 1994) defend one of the most influential constructivist accounts to date, the theory-theory. This claims that children possess a domain-general learning mechanism which they use to construct and then progressively revise theories about the world, in a way not dissimilar from how scientists go about formulating and revising their theories. During the preschool years, Gopnik and Wellman argue, children would gradually move from a desire-perception theory, in which mental states are conceived non-representationally, to a representational belief-desire psychology. In many cases, two-year-olds’ desire-perception psychology will lead to incorrect predictions and fail to offer sensible explanations. As it happens in the case of scientific theories, however, children will not immediately abandon it; rather, they will try to dismiss the counterevidence or to elaborate auxiliary hypothesis. Three-years-olds would be in such a transitional stage,

entertaining the possibility of false beliefs only as an auxiliary hypothesis to account for otherwise recalcitrant data. Finally, around the age of four, children would acquire a fully representational conception of the mind, which would allow them to pass direct false-belief tests.

3.3.2 Josef Perner

Another influential version of the theory-theory was proposed by Perner (1991). Like Gopnik and Wellman, Perner maintains that children acquire a representational theory of mind around the age of four. However, he argues that this process of theory change is enabled by a development in children's ability to entertain mental models of the world. One-year-olds, according to Perner, are only capable of entertaining a single model, which is a representation of the current situation. During the second year, children would then acquire the capacity to entertain multiple models, thus being able to simultaneously represent not only the current situation but also past, future and hypothetical situations. In Perner's eyes, this is what allows two-year-olds to engage in pretend play, which, he argues, requires behaving as if a possible situation were actual. Finally, around the age of four, children would acquire the capacity to meta-represent, by which Perner means representing the representational relation itself. This, Perner argues, requires building a meta-model, which represents that a certain situation (e.g. the situation shown in a picture, or the situation presented by a belief) represents another situation (e.g. reality.), either truly or falsely. The capacity to entertain meta-models would allow four-year-olds to acquire a concept of representation, opening the door to a representational understanding of the mind.

4. Wider Implications

We have seen in the last section that the question of how early belief-reasoning abilities emerge has important implications for the debate between nativism and constructivism. In this section, we take a quick look at other ramifications of this issue, specifically at its connections with consciousness, communication and moral judgement.

4.1 Consciousness

The most mystifying and intriguing aspect of conscious experience is its phenomenal character, the fact that there is something it is like for the subject to have a conscious experience —

something which seems to be impossible to put into words or communicate to anyone who has not had a similar type of experience (Block, 1995). Thus, a blind person will never know what it is like to see a rainbow, and we will never know what it is like for a bat to perceive the environment through echolocation (Nagel, 1974). Phenomenal character has been argued to pose the hardest challenge to cognitive science, forever resisting all attempts to be explained in physical terms (Chalmers, 1995; Levine, 1983). Nonetheless, several philosophers have argued that consciousness can be explained, and the favoured approach among these philosophers has been to try and reduce phenomenal consciousness to representational content (Dretske, 1995; Lycan, 1996; Tye, 1995). These representational approaches divide in two broad categories: first-order and higher-order. Higher-order representationalists maintain that for a state to possess phenomenal character, it must itself be represented by a higher-order state (Carruthers, 2000; Lycan, 1996; Rosenthal, 1997); first-order representationalists, in contrast, deny this (Dretske, 1995; Tye, 1995).

An important implication of higher-order representationalism is that only creatures that are capable of entertaining higher-order states can be conscious in the phenomenal sense. Notably, the results of direct false-belief tests have often been taken to show that children younger than four cannot entertain higher-order states. Fred Dretske (1995), for example, uses this to mount an objection to higher-order theories: since the evidence suggests that children younger than four cannot entertain higher-order states, it follows from higher-order representationalism that they must not be phenomenally conscious; and yet, surely, they are. Of course, as they say, one philosopher *modus ponens* is another's *modus tollens*. Thus, Carruthers (2000) argues that if it is true that children younger than four cannot entertain higher-order states, then the right conclusion to draw is not that higher-order representationalism is false but that children younger than four are not phenomenally conscious.

4.2 Communication

Understanding what someone means by an utterance or gesture seems to require representing what she intended to communicate by means of it. Indeed, in a seminal paper entitled "Meaning", Paul Grice argued that 'meaning something by utterance U' should be analysed as 'intending U to produce some effect in the audience by means of the recognition of this intention' (1957, p. 385). Grice's analysis has inspired several influential theories of communication (e.g. Schiffer, 1972;

Sperber & Wilson, 1986). These so-called “neo-Gricean” theories have the implication that understanding what a speaker means requires being able to entertain complex higher-order representations (Breheny, 2006; Thompson, 2014a). An important implication of this is that if children under four cannot ascribe mental states, then they have no way of understanding what people mean by their utterances. Thus, the development of communication appears to be bound to that of mentalising.

4.3 Sociomoral Abilities

The ability to make moral judgements seems to presuppose the capacity to ascribe mental states. Thus, for example, to decide whether you should blame Tom for hitting you on the head, you need to consider whether he did so intentionally or accidentally. Surprisingly, the effect seems to go in the opposite direction as well, since when people (including children from the age of four, see Leslie, Knobe, & Adam, 2006) find the consequence of an action to have a negative valence they are much more likely to judge that it was intentional (Knobe, 2003).

In recent years, developmental psychologists have begun using indirect methodologies to investigate what infants understand of the social and moral sphere. There is now a growing body of findings suggesting that infants have specific expectations concerning fairness, reciprocity, authority, ingroup-outgroup dynamics and other sociomoral concepts (Baillargeon et al., 2015). As Baillargeon and colleagues note, this provides converging evidence that infants possess mentalising abilities, since a sociomoral abilities presuppose mentalising abilities. On the other hand, this means that how we interpret the evidence from indirect false-belief studies will affect how we interpret the evidence from indirect sociomoral studies.

5. Plan of the Thesis

In Chapter II, we will consider how late-emergentists have proposed to explain the results of indirect false-belief tests. In Chapter III, on the other hand, we will consider how early-emergentists have proposed to explain the results of direct false-belief tests. In Chapter IV, I will present my own, early-emergentist explanation of direct false-belief results, called the *processing-time account*. In Chapter V I will explore a new option, alternative to both early-emergence and late-emergence

accounts, which I call the *hybrid approach*. Finally, in Chapter VI, I will consider which offers the best account of all the evidence – the processing-time account, or the hybrid approach.

II

ALTERNATIVE INTERPRETATIONS OF INDIRECT FALSE-BELIEF TESTS

Late-emergentists have argued that the results of indirect false-belief tests provide no evidence that infants can reason about what other people think. This is because, they claim, there are ways infants could pass these tests without being aware of other people's beliefs. In the literature, one can find many different proposals about how infants could do this, varying significantly in respect to the level of intellectual sophistication they presuppose.

On one end of the spectrum, we find very frugal accounts, maintaining that quite basic abilities or dispositions, such as, for example, a preference for low-level novelty, would be sufficient for passing an indirect false-belief test (Heyes, 2014; Perner & Ruffman, 2005). In the following, I will refer to these as *low-level* interpretations. On the opposite end, we find richer accounts, claiming that some mentalising abilities would be required, though, crucially, not a proper understanding of belief (Apperly & Butterfill, 2009; Hedger & Fabricius, 2011; Wellman, 2010, 2014). I will refer to these alternative interpretations as *mentalistic*. And as one might expect, several proposals occupy the middle ground between these two extremes, maintaining that infants can predict behaviour

without however ascribing any mental states (De Bruin & Newen, 2012, 2014; de Bruin, Strijbos, & Slors, 2011; Perner, 2010; Perner & Ruffman, 2005; Zawidzki, 2011). These I will call *behavioural* interpretations¹.

Low-level interpretations will be discussed in §1; behavioural interpretations, in §2; and mentalistic interpretations, in §3. As we will see, none of these alternative interpretations can explain all indirect false-belief results, which means that the face-value, early-emergentist interpretation is the best we've got.

1. Low-Level Proposals

Low-level accounts maintain that indirect false-belief results can be explained without attributing to infants any understanding of behaviour or any capacity to predict behaviours as such. Mainly two lean accounts have been proposed: Josef Perner and Ted Ruffman's (2005) three-way association account, which will be discussed in §1.1, and Cecilia Heyes' (2014) low-level novelty account, which will be discussed in §1.2.

1.1 The Three-way Association Account

Perner and Ruffman (2005) argue that the differences in looking times found by Onishi and Baillargeon (2005) can be explained on the assumption that infants form three-way associations between agents, objects and locations. After the change of location, half the infants in this study were shown the agent (E1) reaching for the green box (green-box event), while the other half were shown E1 reaching for the yellow box (yellow-box event). Since both groups had already seen E1 reaching for the green box during the familiarisation trials, both would already have formed a three-way association between E1, the green box and the toy. Thus, only the infants in the yellow-box event would form a new association, this time involving E1, the yellow-box and the toy. According to Perner and Ruffman, this might well be the reason why the infants in the yellow-box group had longer looking times: not because they were expecting E1 to reach for the green box, but simply

¹ Notice that behavioural interpretations do not deny that the infants themselves possess mental states – what they deny is that infants would ascribe mental states to other people (or even themselves.)

because, having never seen E1 reaching for the toy in the yellow box, they were in the process of forming a new three-way association.

This proposal seems to be ruled out by one of the control conditions in Onishi and Baillargeon (2005). The condition described above is the FB-Green condition. In the TB-Yellow condition, E1 stayed and witnessed the object moving from the green box into the yellow box. This flipped the looking times: the infants now looked longer when E1 reached for the green box as opposed to the yellow box. In terms of three-way associations, however, we would expect to see similar looking times in the two conditions, since in neither case E1 reaches for the yellow box until the test phase. Thus, in the TB-Green condition as in the FB-Green condition, infants would still have to form a new association in the yellow-box event but not in the green-box event.

To circumvent this difficulty, Perner and Ruffman suggest that three-way associations may be formed not just based on reaching but also based on looking. E1 looked at the toy as it was going in the yellow box; this may have led infants to associate E1 with the yellow box and the toy already during the belief-induction phase. Notice, however, that this falls short of fixing the issue. For, now, by the time they get to the test phase, infants will have formed both associations: E1-Toy-Yellow Box and E1-Toy-Green Box. Thus, neither in the case of the green-box event nor in the case of the yellow-box event would infants form a new association, which means three-way associations cannot explain the difference in looking times².

Another suggestion made by Perner and Ruffman (2005) may be argued to solve this problem. Perner and Ruffman maintain that parts of the brain (“neurons in non-frontal regions”) code for the recency of the stimuli. Infants may thus look longer at one event compared to the other not because one of the two involves an entirely novel agent-object-location triad, but because one event involves a less recent triad compared to the other, i.e., one that was seen less recently. In the TB-Yellow condition, E1 looked at the toy in the yellow box after she reached for the toy in the green

² Oddly, as they try to explain the TB-Yellow and FB-Yellow conditions, Perner and Ruffman (2005) refer to “E1-toy-green box” as a new combination of elements, which would induce infants to form a new association. Perner and Ruffman seem to forget, however, that infants already saw E1 reaching for the toy in the green box during the familiarisation trials. Thus, the combination is not really new, and no new association would have to be formed. (Although see the next footnote.)

box. Thus, the E1-Toy-Yellow Box triad was more recent than the E1-Toy-Green Box one, thus eliciting less looking time³.

Well, let's suppose that this works, and thus that Onishi and Baillargeon (2005) can be explained in terms of three-way associations. Perner and Ruffman's account still seem to fare poorly with plenty of other studies. The classic experiment by Woodward (1998), described in Chapter I, §2.2.1, seems to offer a very elegant illustration of its difficulties. An agent (E1) reaches repeatedly for one of two objects (always in the same location). In the test phase, the two objects are in reversed positions, and E1 reaches either for the same object as before (new-path event) or for the other one (new-goal event). Note that in either case participants witness a novel agent-object-location combination: in the new-path event E1 is reaching for the same object as before, but at a new location, whereas in the new-goal event she is reaching for a new object, but at the old location. Notice that neither triad would be "less recent" than the other; they would both be entirely novel. The three-way association account would thus predict equal looking times. In fact, however, the infants in this study looked longer at the new-goal event, as one would expect if they attributed to her a preference for the object she grasped in previous trials.

Crucially, it is not an accident that the three-way association cannot account for all the evidence; indeed, the opposite would have been truly surprising, for there is now a large number of mentalising studies employing looking times as a measure. For the three-way association account to work, in each of these studies the event that is not consistent with the agent's mental states (at which infants look longer) would always have to involve a novel or, at least, less-recent agent-object-location triad (thus resulting in the formation of a new three-way association). Since these two properties are not causally related (as the study by Woodward goes to show), there is no reason why they should be reliably co-instantiated, other than sheer coincidence.

The three-way association account suffers from another problem as well, just as threatening as the first. Not all indirect mentalising studies have used looking times as a measure. Some of the studies that have used other paradigms, such as, for example, interactive studies, are clearly problematic for the three-way association account. Consider for example the study by D. Buttelmann et al. (2009), which was described in Chapter I, §1.2.1. An agent (E1) tries to retrieve a toy from

³ Or, perhaps: infants simply have a bad memory, only remembering the most recent combination of agent, object and location. This seems to fix the problem too.

the box where she had put it. Unbeknown to her, however, a second agent (E2) has moved the toy to a second box and locked both boxes with a pin. Thus, E1 is now struggling to open the first box, which does not actually contain the toy she is looking for. Most of the infants in this study tried to help her by pointing or crawling to the second box; they thus seemed to understand that E1 was only trying to open the first box because she thought the toy was in there. The problem for the three-way association account here is that infants' behaviour in this and similar studies appears to be the outcome of an episode of practical reasoning; that is, it appears to be motivated by a desire to help the agent coupled with a belief that the agent is looking for the toy. That infants associate the experimenter with the toy and one of the boxes cannot be the reason they point or crawl to the other box; while associations can of course influence behaviour, as in episodes of priming, they are not the right kind of state to either inform or motivate action.

1.2 The Low-Level Novelty Account

The low-level novelty account, defended by Heyes (2014), maintains that infants enjoy looking at events that involve novel combinations of colours, shapes and other low-level, perceptible properties. This simple fact, Heyes argues, can explain most indirect false-belief results. Consider, for example, the study by Onishi and Baillargeon (2005). In the FB-Green condition, E1 is absent as the toy “crawls” out of the green box into the yellow one. All the infants will have seen E1 reaching for the green box during the familiarisation trials; thus, to the infants in this condition, E1 reaching for the yellow box will seem more perceptually novel, explaining why infants looked longer at the yellow-box event.

The low-level novelty account seems to be ruled out by the TB-Yellow condition in Onishi and Baillargeon's study. In this case, E1 stayed as the object moved from one box to the other; it is still true, however, that E1 never reached for the yellow box. Thus, the yellow-box event should still look more novel. In this case, Heyes claims that the toy moving into the yellow box should make E1 reach into the yellow box look less perceptually novel, since the two events are perceptually similar — in both cases there is movement towards the yellow box.⁴

⁴ Notice that this actually does not fix the issue. The yellow-box event may not look novel since infants saw the toy moving the yellow box; but the green-box event would not look novel either, since infants also saw E1 reaching

Alright; does this not mess up the explanation for the FB-green condition, however? For in that condition, as well, infants saw the object move into the yellow box. Here Heyes maintains that the fact that E1 reappeared right after the toy moved might have caused retroactive interference. That is, the infants in the FB-Green condition were so captured by E1's reappearance that they forgot what had just happened (the toy moving from the green to the yellow box). Hence, E1 reaching for the yellow box still looked more perceptually novel to them. (Since, in the TB-Yellow condition, E1 stayed as the toy moved, there was no retroactive interference in that case.)

Heyes acknowledges that not all indirect false-belief studies can be explained in terms of perceptual novelty. In some cases, she claims, we must take into account "imaginal novelty". Consider, for example, the experiment by Song and Baillargeon (2008), which was described in Chapter I, §1.2.3. The study involved two experimenters (E1 and E2) and two toys (a toy skunk and a doll with blue hair). During the familiarisation trials, E1 always reached for the doll, thus displaying her preference. During the belief-induction trial, two boxes were introduced: one had a plain lid (plain box), the other had a lid with a tuft of blue hairs (like the doll's) sticking from underneath (hair box). In the false-perception condition, as E1 was away, E2 put the skunk in the hair box, and the doll in the plain one. The true-perception condition was similar except that E1 stayed as E2 put the objects in the boxes. In the test phase, E1 reached either for the plain box (plain-box event) or for the hair box (hair-box event).

Now, the true-perception condition seems to pose a problem for Heyes. Infants saw E1 reaching for a doll with blue hair during the familiarization trials. During the test phase, neither object is visible, but the hair box has a tuft of blue hair. The less novel event for them should thus be E1 reaching for the hair box, instead of the plain one. Heyes suggests that this problem can be solved by appealing to what infants are likely to imagine, instead of what they perceive. If infants imagine the doll in the plain box and the skunk in the hair box, then E1's reaching for the plain box will be the less novel event. (In the false-perception condition, however, E1's disappearance and re-appearance causes retroactive interference, thus preventing the infants from imagining the contents of the boxes. Consequently, they do expect her to reach for the hair box.)

for the green box during the familiarisation trials. Thus, neither should look novel, which would predict equal looking times.

Can all indirect false-belief studies be explained in terms of low-level novelty? For this to be the case, the event that is inconsistent with the mental states of the agent would always have to be the one infants either perceive or imagine as more novel; as in the case of the three-way association account, however, there is simply no reason this should be the case, aside from sheer coincidence. The study by Woodward (1998) provides, once again, a neat counter-example. The new-goal event is similar to previous events in that the hand moves along the exact same path as before, but also dissimilar in that the hand ends up touching a different toy. On the other hand, the new-path event is similar to previous events in that the hand ends up touching the same toy as before, but also dissimilar in that the hand is now moving along a different path. Overall, both are novel in some respects but not in others. On Heyes' account, then, one would thus expect to see roughly similar looking times; and yet, Woodward (1998) found that infants looked longer at the new-goal event.

Like the three-way association account, the low-level novelty account suffers from a second problem, as well: it does not seem to apply to interactive studies. To illustrate, consider again the interactive study by D. Buttelmann et al. (2009). The fact that infants associate E1 with the toy and one of the boxes cannot explain why infants crawl to the other box; and neither can the fact that infants prefer novel stimuli. In fact, that infants exhibit any helping behaviour at all should be problematic for Heyes. Plainly, to even try and help someone achieve their goals, one must first ascribe some *goals* to them. If infants conceived of E1 simply as an interesting but mindless configuration of shapes and colours, trying to help her would not make much sense⁵.

⁵ In discussing the study by Buttelmann and colleagues, Heyes mentions that the infants may crawl to the box where the toy is simply because they want to play with the toy, and not because they are trying to help the agent. The study by Southgate, Chevallier, and Csibra (2010) seems to rule out this suggestion. The study used a design similar to the one by Buttelmann and colleagues, except that it involved two toys. E1 puts one toy in box A and the other in box B. Then, as E1 is away, E2 switches them. When E1 comes back, she points to Box A, saying to the infant, "can you give me the Sefo?" ("Sefo" is a made-up word.) Infants retrieve the object that is in Box B, apparently understanding that the Sefo is the object E1 wrongly takes to be in Box A. Now, the problem for Heyes is that there are two toys. The authors tried to match the toys to make them look equally novel and interesting. Furthermore, they counterbalanced both which toy was in which box and which box E1 pointed at. Thus, if infants simply reached for the toy they liked, we would expect to see roughly half of them reach for box A and the other for Box B, which is not what Southgate and colleagues found.

1.3 Summary

Both the proposals we have considered in this section can explain the results of the study by Onishi and Baillargeon only to the extent that, in this study, the event that is consistent with the agent’s mental states happens to be one that is more novel along some dimension or other. There is no reason why these two properties should be aligned in all indirect mentalising studies, however, and, indeed, we have seen at least one counterexample. In addition, neither Perner and Ruffman’s three-way association account nor Heyes’ low-level novelty account had an explanation for indirect false-belief studies using an interactive paradigm. Thus, overall, we can consider these proposals ruled out by the evidence.

2. Behavioural Proposals

In this section we will discuss those alternative proposals of indirect false-belief studies that attribute to infants some ability to predict behaviour as such, without however ascribing to them any understanding of the mind. Perner’s behaviour-rule account will be discussed in §2.1; the affordance-based account put forward by Leon de Bruin, Derek Strijbos and Marc Slors, in §2.2; Tadeusz Zawidzki’s enhanced teleological-reasoning account in §2.3; and, finally, the association account defended by de Bruin and Albert Newen, §2.4.

2.1 The Behaviour-Rule Account

2.1.1 The “search-where-last-seen” rule

Several authors have suggested that infants may predict behaviour by relying on superficial, non-mentalistic generalisations, sometimes called “behaviour rules” (Gallagher & Povinelli, 2012; Perner, 2010; Perner & Ruffman, 2005; Ruffman, 2014; Ruffman, Taumoepeau, & Perkins, 2012; Sabbagh, Benson, & Kuhlmeier, 2013). Specifically, Perner and Ruffman (2005) suggest that the task used by Onishi and Baillargeon (2005) could be passed using the generalisation that people look for things where they last “saw” them (i.e., where they were last on their direct line of sight.) This works because in the study by Onishi and Baillargeon (as in the typical unexpected-transfer task) the box where E1 takes the object to be is also the box where the toy was last in her

unobstructed line of sight. So long as this is the case, knowing the latter piece of information is enough to predict which box E1 will reach for; one does not actually need to ascribe a belief to her.

This is an important weak spot in Onishi and Baillargeon's study⁶, one that (as we will see in the rest of this chapter) many alternative interpretations have tried to exploit. Nonetheless, many of the indirect false-belief studies published since 2005 have used different belief-reasoning tasks, where the agent does not take the object to be where she has last seen it (e.g., Scott & Baillargeon, 2009; Scott et al., 2010; Scott et al., 2015; Song & Baillargeon, 2008; Song et al., 2008; Träuble et al., 2010) Consider, for example, the study by Song et al. (2008), which was described in Chapter I, §1.2. In the informative-intervention condition, E1 leaves a ball in a blue box, then goes away; while E1 is absent, E2 moves the ball into a mug; when E1 comes back, E2 tells her the ball is now in the mug. If infants were using the search-where-last-seen rule, they would expect E1 to reach for the blue box, which is where, presumably, E1 last saw the ball. Instead, the infants in Song and colleagues' study expected E1 to reach for the mug, thus seeming to understand that E1 would update her belief based on E2's testimony. Or, to see a second example, consider again the false-perception condition in the study by Song and Baillargeon (2008), where E1 has not seen the object she is looking for (i.e., a doll) in either of the boxes. In this case, if infants were using the search-where-last-seen rule, they should not have any expectations about what E1 will do. Instead, the infants in this study expected E1 to reach for the hair box, thus anticipating that E1 would take the blue hair coming out of the box to be the doll's. These studies (together with the others mentioned above) decisively show that infants are not using the search-where-last-seen rule.

2.1.2 Behaviour Rules Redux: The Shortcut Argument

Despite this, Perner (2010) argues that there are other behaviour rules that could allow infants to pass the tests mentioned above. In fact, he claims, all the mentalising tests that have been used with infants are necessarily open to behaviour-rule interpretations. The argument that supports this bold claim was put forward by Povinelli and Vonk (2003), and was originally directed at

⁶ It is worth noting that Onishi and Baillargeon are not to be blamed for this — first, because, as mentioned above, this weak spot was already present in the original unexpected-transfer task, which Onishi and Baillargeon simply adapted to the violation-of-expectation paradigm; and second, because all studies are inevitably open to some alternative interpretation or other, which must be ruled out in follow-up studies.

mentalising studies in apes and other non-human species, but can be easily extended to infants. The argument takes its move from the observation that since mental states cannot be directly observed (you cannot look in my head to see what I'm thinking) one has no choice but to infer them from behavioural cues and other observable features of the situation. Now, the problem is that if the agent's action can be predicted based on the mental states attributed, and those mental states can be inferred based on the events observed, then it should always be possible to take a shortcut, as it were, and predict the agent's action directly from the events observed, skipping the intermediate, mentalistic step. Call this the shortcut argument.

To see the argument in action, consider again the study by Song et al. (2008) mentioned above. To pass this task through mentalising, one would have to infer where E1 thinks the ball is based on what happened during the belief-induction trials — specifically, based on where she last saw the ball and on what E2 told her. It seems, however, that one could also take a shortcut, thus expecting E1 to look where E2 told her the ball was, without ascribing to her the corresponding belief. Put in the form of a rule, the shortcut would look something like this:

BR1: If an agent does not have a direct line of sight to an object, then, if she was told the location of the object, she will look for it there, otherwise she will look for it where it was last in her direct line of sight.

Of course, this particular rule cannot explain all indirect false-belief studies; it cannot explain, for example, the study by Song and Baillargeon (2008). Nonetheless, the shortcut argument shows (or at least purports to show) that for any given set of indirect false-belief tests, there must be at least one behaviour rule that could be used to pass them.

2.1.3 The Objection from Learning

The shortcut argument has generated a good deal of discussion in the literature (Andrews, 2005; Buckner, 2014; Butterfill & Apperly, 2013; Carruthers, 2013; Clatterbuck, 2015; Fitzpatrick, 2009; Low & Wang, 2011; Lurz, 2011; Spaulding, 2011; Thompson, 2015) — too much discussion, in fact, for me to attempt to summarise it here. What I will do, instead, is focus on the most serious problem I take the shortcut argument to have; this is that infants may not be able to learn the required behaviour rules.

First of all, notice that the fact that a test could, in theory, be passed by certain means does not entail that it is plausible that a given category of participants pass it by those means. To illustrate, consider a math test in which one must add pairs of positive integers smaller than 1000, picked at random. In theory, one might pass such a test without knowing how to add. After all, the number of possible questions, though very large (10^6), is finite. Certainly, however, no one would seriously suggest that children would be able to pass it this way, since there is no way the average child (or, for that matter, the average adult) could memorise the answers to 10^6 questions!

Going back to our matter, the shortcut argument simply shows that if infants knew the required behaviour rules, they could pass indirect mentalising tests without possessing any mentalising abilities. It is still an open question, however, whether infants can be expected to know the required behaviour rules, specifically whether they could learn them over the first few months of their life. Peter Carruthers has put this point in the form of a challenge to supporters of the behaviour-rule approach:

[...] it is important to note that the behavior-rule and infant-mindreading hypotheses assume very different explanatory burdens. This is easiest to see if it is assumed that behavior rules need to be learned over the course of the first few months of the infant's life. The challenge, then, is to show, in respect of each proposed rule, that infants have adequate opportunities to learn it in the time available. This seems implausible with respect to a number of the behavior rules that would be needed to explain some of the more recent infancy results. But in any case it is a challenge that behavior-rule theorists have not yet taken up. (Carruthers, 2013, pp. 150-151).

Why doubt whether infants could learn the required behaviour rules? Well, when one actually gets down to finding the behaviour rule that could be used to pass a given test, the result is often a very ad hoc rule – a rule so closely tailored to the particular experiment that it would be useless in any other situation. And if that is indeed the case, then it is hard to see how infants could have learned it, unless they had the opportunity to observe that specific situation, to which only the rule applies, enough times in the past. Given that some of the experiments depict fairly uncommon false-belief scenarios, there are legitimate grounds to doubt whether infants could have learned the corresponding behaviour rules. The same problem does not arise on the supposition that infants

rely on mentalistic generalisations, which are supposed to be very general and should thus apply across different situations⁷.

It will help to consider a concrete case. Take the study by Scott et al. (2015). The experiment starts with several familiarisation trials, during which it is established that the target agent (call her E1) likes rattling toys: E1 shakes several objects, one after the other, storing those that rattle in a box and throwing the rest in a bin. In the test trial, E1 grabs a new toy and shakes it, noting that it rattles. Before either storing it or throwing it away, however, E1 is called away for a little while, leaving the toy on a table. Taking advantage of E1's absence, a second agent (E2), who was present throughout, picks from the bin one of the toys E1 had discarded during the familiarisation trials, which looks identical to the one E1 just shook but does not rattle, and puts it in its place, thus stealing the rattling toy. When E1 came back, the seventeen-month-old infants in this study expected her to store the non-rattling toy instead of putting it back in the bin, thus (apparently) understanding that E1 would be fooled by the fact that the two toys looked the same.

Now, let us try to come up with a behaviour rule that could be used to pass this test. The action we want to predict is whether E1 will store the toy or throw it in the bin. What E1 will do depends on whether she takes the toy to be one that rattles when shaken or not. What E1 believes must be inferred, in turn, from what happened earlier in the experiment. If, for example, the toy rattled when E1 shook it, then E1 is likely to think the toy is a rattling one. As a first attempt, consider the following rule:

BR2. If an agent has put all the objects that rattled when shook in one location, and all the objects that did not rattle when shook in another location, then she will continue to do so with the next toy she will shake.

BR2 leaves room for improvement, since it predicts that, upon her return, E1 will throw the toy in the bin instead of storing it (since she shook it during the familiarisation trials and it did not

⁷ Can supporters of the behaviour-rule approach avoid the problem by maintaining that behaviour rules are innate? Not quite, for, as Carruthers (2013, p. 151) points out, if a rule is innate then it must have been selected, which, in turn, means that it must have posed an adaptive advantage for our ancestors. If, however, the rule at issue is so ad hoc that it only applies to a very uncommon situation, it is very unlikely that it could have improved the fitness of our ancestors. In comparison, it seems much easier to make a case for the adaptiveness of having folk psychological generalisations which apply across situations.

rattle). To fix this problem, we need to add to the rule a clause about what happens when a rattling toy is replaced with one that looks the same. Thus:

BR2'. If an agent has put all the objects that rattled when shook in one location and all the objects that did not rattle when shook in another location, then she will continue to do so with the next object she will shake. If, however, the object the agent has shaken is replaced with one that looks the same while the agent is not present, then she will do with the replacement object what she would have done with the object she shook.

BR2' seems to make the right prediction: when E1 comes back, she will do with the replacement toy what she would have done with the toy she shook, that is, storing it in the box. However, BR2' is still less than perfect, since it contains some unnecessary details. An improved version might be:

BR2''. If an agent has performed action A1 with all the objects that turned out to have property P1 when tested, while performing action A2 with all the objects that turned out to have property P2 when tested, then she will continue to do so with the next object. If, however, an object the agent has already tested is replaced with one that looks the same while the agent is not present, then the agent will do with the replacement object what she would have done with the object she had already tested.

BR2'' seems to be about the best one can do without introducing mental states. One can argue about the details, of course, but it is hard to think of a behaviour rule that would be considerably more general than BR2'' and still make the correct prediction. It seems, for example, that any behaviour rule would have to specify that the replacement object must look the same as the object it replaces, for the agent would not conflate the two otherwise. Similarly, it is important that the replacement happens while the agent is not present, and after the agent has already tested the object; for, again, the agent would not be fooled otherwise.

Now, if BR2'' is indeed the behaviour rule that one must use to pass the test by Scott et al. (2015), then it is reasonable to doubt whether infants would ever be able to learn it. For it seems unlikely that the average infant will have opportunity to observe, during the first seventeen months of her life, the type of situation BR2'' describes a sufficient number of times to memorise its outcome. In fact, I am not sure whether I myself have ever observed an agent in this exact predicament! (Note that a slightly different situation would lead infants to learn a different rule; to learn BR2'',

infants would have to have observed the specific sequence of events that satisfies the conditions of the rule, and this seems incredible.)

2.2 The Affordance-based Account

de Bruin et al. (2011) defend an affordance-based interpretation of indirect false-belief results. An affordance is what an object, or more generally an environment, affords to an agent. According to De Bruin and colleagues, infants understand that different objects afford different actions. For example, an infant might realise that a pacifier affords putting-it-in-the-mouth. After several failed attempts to fit a ball in her mouth, however, the infant might begrudgingly conclude that it does not afford the same action to her. Infants would also understand that an object may afford different actions to different agents. The infant in our example might realise that while the ball does not afford putting-it-in-the-mouth to her, it might afford putting-it-in-the-mouth to the dog, who has a bigger mouth. Furthermore, infants would be sensitive to the behavioural expressions of goal-directedness and would exploit these behavioural cues in predicting whether agents will take advantage of the affordances available to them. For example, if big brother is wetting his lips looking at the cake, an infant might consider him acting on the corresponding affordances (e.g. stuffing a big slice of cake in his mouth) as more likely than if he had made a disgusted face. Finally, infants would know that “an observed change in affordance does (or does not) imply a change of the agent's future action if the change is (or is not) visually or otherwise accessible to the agent” (2011, p. 512). Two points are important here. First, infants would keep track not just of what actions are afforded to agents about also of which of these affordances are “visually or otherwise accessible to the agent”. This notion of accessibility is meant to play a role similar to that which perception plays in our folk psychology; instead of perceiving (or not perceiving) an object *O* at a location *L*, agents will access (or not access) the affordances provided by *O*-at-*L*. Infants will expect agents to act only on the affordances that have been accessible to them. If a ball is within the dog's reach, our infant will be aware that it affords grasping-with-the-mouth to it; if, however, the dog does not have a direct line of sight to the ball, that affordance will not be accessible to it, and thus the infant will not expect the dog to act on it. Second, a change in affordances would only affect an agent's actions if it is accessible to the agent. Suppose our infant sees Mom put the cake in the fridge; then, while Mom is away, big brother eats the cake. Since the cake is no more, it does not

afford to Mom all the actions it used to afford to her, such as grasping, holding, eating and so on. Thus, there has been a change in the affordances available to Mom. However, Mom has had no way to access this change, since it happened while she was away. Thus, the fact that the affordances have changed should not lead infants to alter their expectations concerning Mom's behaviour. It is just as likely now that Mom will go to take the cake out of the fridge, as it was before big brother ate it. The bottom line is that what matters in predicting behaviour are not the actual affordances, but the affordances agents have, or have had, access to.

We can see how this would work for a standard false-belief test, like Onishi and Baillargeon (2005). Since E1 put the toy in the green box, not only will the toy-in-the-green-box afford several actions to her, like grasping and playing with, but these affordances will be accessible to E1. Thus, infants will consider it likely that she might act on them. When the watermelon slice moves to the yellow box, there is a change in affordances, but one E1 has no access to. While the toy-in-the-yellow-box affords grasping to E1, E1 has no access to that affordance and thus infants will consider it very unlikely that she will act on it. At the same time, infants will not alter the likelihood that E1 will (try to) reach for the toy in the green box. Thus, infants will be surprised when E1 reaches for the yellow box as opposed to the green one.

Notice, however, that this explanation works because it exploits the same weak spot as the search-where-last-seen rule considered in §2.1.1: in the study by Onishi and Baillargeon (2005), the location where E1 takes the object to be is the same location where the object was last on her line of sight — and thus, the same where the affordances related to the object were last accessible to her. We have seen in §2.1.1, however, that other indirect false-belief tests are not open to this type of interpretation. For example, consider again the informative-intervention condition in Song et al. (2008), where infants expect E1 to look for the ball inside the mug, despite having last seen it in the box, because E2 told her that the ball was in the mug. That E2 tells E1 that the ball is in the cup should not change the fact that E1 last accessed ball-related affordances in the box; thus, infants should expect E1 to reach for it there.

How serious is this problem? It seems the issue could be fixed by tinkering with the conditions on which agents can access affordances. Perhaps agents do not need a direct line of sight to O at L to access the affordances provided by that object at that location; perhaps, they could also access the affordances provided by O-at-L if they were told that O is at L. Fair enough. It seems,

however, that unless we grant infants an understanding of belief, some of the rules that govern accessibility of affordances will have to be quite ad hoc. This is the same problem that afflicts behaviour rules (discussed in §2.1.3). Since we have already discussed this issue there, I will not reiterate it here.

2.3 The Enhanced Teleological-Reasoning Account

The teleological reasoning account was put forward by Csibra, Gergely and colleagues (Csibra & Gergely, 1998; Gergely et al., 1995) as a possible interpretation of indirect goal-attribution studies (some of which were discussed in Chapter I, §2.2.1). Intentional actions, Csibra and colleagues argue, are caused by internal mental states, like intentions or desires, but justified by external goals. For example, what caused the chicken to cross the road is (say) the desire to cross the road, which is of course internal; what justifies it, however, is the goal of being on the other side of the road, which is just an external state of affairs. Csibra and colleagues argue that infants may rely on this non-mentalistic notion of goal as they try to predict what another agent will do. Specifically, infants may expect agents to choose means that are efficient to bring about their goals given the constraints of the situation. If, for example, infants have been habituated to a large ball jumping over a barrier to reach a smaller ball on the other side, they will expect the large ball follow a straight line when, in the test events, the barrier is removed (as in Gergely et al., 1995).

Csibra and colleagues never suggested that infants could use teleological reasoning to pass indirect false-belief studies. There would seem to be an obvious obstacle to doing this, for on the teleological reasoning account infants expect agents to pursue their goals in the most efficient way given the way the world actually is, rather than the way the agent takes to world to be. Thus, in an unexpected-transfer type of scenario, the teleological reasoning account predicts that infants will expect the agent to look for the toy at its actual location.

Despite this problem, Zawidzki (2011) has argued that an “enhanced” version of teleological reasoning, incorporating the notion of “information access”, can explain indirect false-belief results:

Among the “situational constraints” ... relative to which infants make judgments of means-ends rationality, older infants include information access. That is, using entirely behavioral cues, like gaze direction, older infants keep track of the

information to which their interpretive targets have access, and incorporate this information into the situational constraints on judgments of means-ends rationality. Zawidzki (2011, pp. 492-493).

What Zawidzki seems to have in mind is that only the events that happened while the agent was present (and oriented in the right direction, etc.) should be included among the situational constraints. If, for example, big brother eats the cake while Mom is away, the fact that the cake has been eaten will not be included among the situational constraints to be taken into consideration in predicting (and/or evaluating for rationality) what Mom will do. Thus, the infants would still judge that going to the fridge would be a rational means for Mom to get the cake.

Crucially, Zawidzki maintains that only the information an agent has had immediate access to would be included in the situational constraints. If big brother says, “I want some cake!”, the infant would only expect him to go get it from the fridge if he had immediate access to the information that the cake was in the fridge. If Mom just told him there was cake in the fridge, that piece of information will be included in the situational constraints for big brother’s action; if Mom told him an hour ago, instead, that information will not be included in the situational constraints.

This feature of Zawidzki’s account differentiates it from Perner’s (2010) behaviour-rule account, but it is not clear whether this improves the account or weakens it. Consider, for example, Onishi and Baillargeon (2005). In the FB-Green condition, E1 last saw the toy in the green box, but this happened during the familiarisation trials, several seconds before the test trial. If, as Zawidzki writes, infants “seem to take into account only information to which targets have access immediately before making their decisions”, (2011, p. 496) then they should not take into account the information that the toy was in the green box; at the same time, infants should not take into account the information that the toy is now in the yellow box either, for E1 was not present during the change of location. As a result, it seems infants should not expect E1 to reach for either of the boxes.

At the same time, Zawidzki’s account still seems to inherit many of the problems that afflicted the behaviour-rule account. Thus, consider for example the study by Scott et al. (2015), where E2, taking advantage of E1’s momentary absence, replaces the rattling toy which E1 just shook with one that looks the same but does not rattle. Since, when E1 comes back, she can see

the non-rattling toy, that the toy is there should be something she has access to; in which case, the infants should expect E1 to discard the toy again, as she already did during the familiarisation trials⁸.

2.4 The Association Account

De Bruin and Newen (2012, 2014) claim that infants can create special representations of agents' object-directed behaviours, which they (somewhat misleadingly) refer to as “associations”. Infants would be able to represent two main types of object-directed behaviours: handling the object (“motor-based associations”) and looking at it (“perception-based associations”). Let us see how this would work in practice. Consider Onishi and Baillargeon (2005). De Bruin and Newen (2012) argue that, as infants see E1 reaching for the green box, they would form a motor-based association of her reaching for the green box. Infants would deactivate this association while E1 is absent and reactivate it as E1 comes back. At this point, infants would expect E1 to re-enact the same behaviour captured in their association, namely reaching for the green box. Thus, when they see her reaching for the yellow box, their expectations will be violated (and/or they will have to create a new motor-based association of E1 reaching for the yellow box.)

This explanation seems to be ruled out by the TB-Yellow condition, where E1 stays as toy moves from the green box to the yellow one. Infants now expect E1 to reach for the yellow box instead of the green box again. However, it is still true that, in the familiarisation trials, E1 reached for the green box and not for the yellow one. Thus, infants would still have that motor-based association of E1 reaching for the green box, and if that were what they based their expectations on, they should now look longer when E1 reached for the yellow box (which is the opposite of what Onishi and Baillargeon found).

In response to this objection, de Bruin and Newen are likely to point out that infants would also create a perception-based association as they see E1 looking at the object going into the yellow box. Based on this, infants would then expect E1 to reach for it there. Two observations are in order. First, should infants not expect E1 to repeat the object-direct behaviour represented by the

⁸ Perhaps, Zawidzki may object, E1 has access to the information that a toy is there, but not to the information that it is one of the non-rattling toys she already discarded. Even granting this, however, the infants should not expect E1 to store the toy with the rattling ones — for E1 cannot have access to the information that the toy rattles, since the toy does *not* rattle. And yet, Scott and colleagues found that the infants in their study looked longer when E1 discarded the toy compared to when she stored it.

perception-based association, namely looking at the toy in the yellow box? Why do they expect her to reach for it? (After all, they do not have a motor-based association of E1 reaching for the toy in the yellow box.) And second: infants should still have a motor-based association of E1 reaching for the green box; thus, should they not expect E1 to reach for the green box every bit as much as they expect her to reach for the yellow one?

These objections may seem pedantic; perhaps, when infants create a perception-based association of E1 looking at the toy in the yellow box, they would erase the previous, motor-based association of E1 reaching for it in the green box; and, perhaps, when they have an association of E1 behaving towards an object in a certain location, infants would expect her to reach for it there, regardless of whether the association was motor-based or perception-based. Fair enough. Notice, however, that for all intents and purposes this is tantamount to claiming that infants expect agents to look for things where they last saw them. We have already considered (and rejected) this hypothesis in §2.1.1⁹.

2.5 Summary

As we have seen, the most promising behavioural proposals are those that try to exploit the fact that, in a typical unexpected-transfer task, the location where the target agent takes the desired object to be is the same where the object was last in her unobstructed line of sight. Several studies have, shown, however, that infants can predict what the agent will do even when this is not the case. It seems that providing a behavioural explanation of these other studies would require attributing to infants very ad hoc and tailor-made generalisations – generalisations that infants may not have occasion to learn. Thus, overall, while behavioural proposals cannot be ruled out as confidently as low-level proposals can, the evidence still speaks against them.

⁹ And yet, De Bruin and Newen (2012) seem positive that their account can explain the study by Song and Baillargeon (2008). This is what they suggest: “the operating system only selects the visual information about the scene that the infant shares with the other agent (the tuft of blue hair on the plain box that belongs to the doll) and represents this information as an anticipation of the agent’s reaching behaviour” (p.11). It is difficult to make sense of this suggestion. For one, the tuft of blue hair sticks out of the hair box, not the plain one, and it is not the doll’s, it simply looks like it. In any case: the information the infants share with the other agent is simply that there is a tuft of blue hair coming out of the hair box. Since E1 has never reached for that tuft of blue hair, why should infants expect her to reach for it now? Unless, of course, they understand that E1 takes that tuft to be the doll’s — but that would be tantamount to ascribing to her a false belief!

3. Mentalistic Proposals

In this section we will discuss those alternative proposals that ascribe to infants at least *some* understanding of the mind (though not of beliefs as such, of course.) There are three mentalistic proposals that I know of: Henry Wellman’s desire-awareness account (discussed in §3.1), Joseph Hedger and William Fabricious’ perceptual-access reasoning account (§3.2) and Ian Apperly and Stephen Butterfill’s (§3.3) minimal theory of mind account. Together with Perner’s behaviour-rule account, Apperly and Butterfill’s has been among the most influential proposals in the field and will thus be discussed in more detail.

3.1 The Desire-Awareness Account

According to Henry Wellman (2010), indirect false-belief results do show that infants possess a relatively sophisticated understanding of the mind. Specifically, in his eyes, there is sufficient evidence to conclude that infants ascribe to other agents desires, emotions, perceptual experiences and even states of knowledge or ignorance. However, Wellman draws the line at false belief:

[indirect false-belief results] help underwrite the description ... that infants achieve a desire – emotion – perception understanding of persons that encompasses an initial sense of knowledge and ignorance (or more precisely, awareness and unawareness) ... however, I do not believe the data demonstrate an infant recognition of false belief (and hence of an internal world of mental contents¹⁰. (Wellman, 2010, p. 267).

Wellman further elaborates this interpretation of indirect false-belief studies in his recent book (Wellman, 2014, pp. 175-184). He provides several (putative) explanations of indirect false-belief studies in terms of his account. Consider, for example, Onishi and Baillargeon (2005). Since the infants in the study are aware that E1 did not see the move from one box to the other, Wellman maintains they will assume E1 does not know its current location — without, however, going so far as to ascribe a false belief to him. Furthermore, Wellman maintains that infants expect agents that are ignorant about the location of an object to search for it

¹⁰As this quotation suggests, Wellman seems to assume that, unless infants know that beliefs can be false, they cannot conceive of the mind as “an internal world of mental contents”. What should we make of this claim? It is not clear why desire, knowledge and the rest would not qualify as “mental contents”. All these types of mental states are typically taken both to be “internal” and to possess intentional content. What more could one need to see the mind as an “internal world of mental contents”?

randomly (Wellman, 2014, p. 179). Thus, when infants see E1 reaching for the object in the correct location, infants would be surprised: was E1 just lucky, or did she know the toy was in there?

There seems to be a problem with this interpretation: strictly speaking E1's reaching for the correct location is consistent with a random search. That one is searching randomly does not necessarily mean that they will look in the wrong place; it simply means that the probability that one will find what they are looking for on a first attempt is less than one. Given that there are two boxes, the probability that E1 will reach for either should be fifty-fifty. Thus, neither event should be more surprising than the other.

One can wonder whether this rejoinder is fair, however. Intuitively, E1 reaching for the right location on a first attempt does seem more surprising than her reaching for the wrong location. True, when probability theory is taken into account, this intuition turns out to be misguided. Can we really expect infants to be skilled at probabilistic reasoning, however? Besides, even if we grant that infants can correctly rate the probability of the two events, the obstacle would appear to be easy to circumvent: Wellman could just claim that infants expect ignorant agents to search in the wrong location (a possibility that has been raised by several authors).

The problem for Wellman goes deeper than this, however. Several indirect false-belief studies have included controls to ensure infants are not simply ascribing ignorance (He et al., 2011; Knudsen & Liszkowski, 2012a; Scott & Baillargeon, 2009; Scott et al., 2010; Southgate et al., 2007). Consider, for example, He et al. (2011). In this experiment, two agents (E1 and E2) are playing the following game: E2 hides a toy in one of two boxes and E1 must point to where it is. There were three conditions. In the knowledge condition, E1 is present as E2 hides the toy. In the ignorance condition, E1 is not present as E2 hides the toy. In the false-belief condition, E1 sees E2 hide the toy in one box, but E2 then moves it to the other box as E1 is absent. In the knowledge condition, He and colleagues found longer looking times when E1 pointed to the wrong box; in the false-belief condition, in contrast, they found looking times when E1 pointed to the correct box; and in the ignorance condition they found no significant difference in looking times. Thus, the two-year-olds in this study do seem to expect ignorant agents to search randomly, looking equally regardless of which box E1 points to. Crucially, however, they did not do this in the false-belief condition. The fact that there was a difference in

looking times between the ignorance condition and the false-belief condition is a problem for Wellman since, according to him, both should be ignorance conditions!^{11,12}

3.2 The Perceptual Access Reasoning Account

In contrast both with nativist accounts and with the received constructivist view, Hedger and Fabricius (2011) argue that children only acquire an understanding of false beliefs around the age of six. Up to that point, they rely on what the authors call perceptual access reasoning or PAR for short. Specifically, Hedger and Fabricius argue that children base their expectations on two generalisations:

- (1) sees (didn't see) -> knows (doesn't know);
- (2) knows (doesn't know) -> will get it right (wrong).

In other terms, not perceiving leads to not knowing and not knowing leads to error. Infants, in contrast, would lack the concept of knowledge, and thus would not be able to rely on PAR. Instead, infants would use an abridged version, which Hedger and Fabricius call "Rule A":

- (A) See/not see —> Get it right/ Get it wrong.

Despite the difference in details, it should be clear that this proposal fails for the same reason Wellman's does: experiments like He et al. (2011) show that infants expect that an agent who has not seen an object at its current location is equally likely to reach for either of two possible locations. Furthermore, experiments like

¹¹ At one point in his book, Wellman suggests that infants may, in some cases, expect ignorant agents to look for toys where they last saw them (Wellman, 2014, p. 180). He does not clarify when infants would do this as opposed to expecting a random search; and he does not appeal to this explanation when interpreting Onishi and Baillargeon (2005). In any case, this is very similar to the search-where-last-seen rule we discussed in §2.1.1 and will thus inherit its problems.

¹² Oddly, when discussing the ignorance condition in Scott and Baillargeon (2009), Wellman (2014) claims that equal looking times are consistent with his account, since he maintains that infants expect ignorant agents to search randomly. Wellman does not seem to notice the inconsistency in his position. If his account predicts equal looking times, then it should do so across studies, and across both ignorance and false-belief conditions. However, Scott and Baillargeon (2009) did not find equal looking times in the false-belief condition, nor did Onishi and Baillargeon (2005) in their FB-Green and FB-Yellow conditions. To explain these results, Wellman must claim that ignorant agents will search in the wrong location, but if he does so then he cannot explain equal looking times in ignorance conditions. You can't have it both ways!

Scott et al. (2015) show that infants do not expect agents who can see the object they are looking for to “get it right”.

3.3 The Minimal Theory of Mind Account

3.3.1 The mToM Account

Apperly and Butterfill (2009) argue that infants have a simplified version of adult folk psychology. Crucially, while infants’ “minimal theory of mind” (as the authors call it; mToM henceforth) would posit internal states that causally interact with each other, it would not posit any intentional/representational states. Thus, infants’ minimal theory of mind could not include any reference to beliefs, which are intentional states par excellence, nor in fact to any other type of propositional attitude. The minimal theory of mind would, however, posit a non-intentional type of state playing a belief-like causal role, guiding and informing action. Apperly and Butterfill refer to this non-intentional counterpart of belief as registration. The minimal theory of mind would also include a non-intentional counterpart to perception, which the authors refer to as encountering¹³. The differences between registering and encountering on the one hand and believing and perceiving on the other, concern not only the content of these states but also their causal role. Let us have a closer look.

Contents. Both beliefs and perceptual experiences are intentional states, which can potentially take any proposition, no matter how complex, as their object. Registration and encountering, in contrast, possess no intentionality; they do not represent anything at all. Instead of putting agents in relation with propositions, as beliefs and perceptual experiences do, registration and encountering put them in a more direct, non-intentional connection with objects. One can only register and encounter objects and their locations; there is no such thing as registering or encountering that. This has two quite important implications that are worth spelling out.

(a) Registrations and encounterings must always involve an agent, an object and a location. For example, Tom can register his mug on the table; he cannot, however, register his mug simpliciter, without registering its location; and he cannot register any of the mug's other properties, such as, for example, its size, shape or colour.

¹³ What about conative states like desire and intention? Does mToM include a counterpart to these states as well? Butterfill and Apperly (2013, pp. 613-614) argue that this would not be necessary. They do acknowledge that infants possess some understanding of goal-directedness, often being able to infer the goal an agent is trying to accomplish. However, they argue that this understanding need not be based on the ascription of conative states to the agent. To see a sequence of bodily movements as goal-directed infants only need to represent those movements as having the function of bringing about a certain outcome.

The upshot is that beliefs (/perceptions) that concern anything other than the location of an object will have no matching registration (/encountering). My seeing that the bird is in the tree, for example, will have as its counterpart an encountering of the bird in the tree; my belief that all bachelors are men, on the other hand, will have no corresponding registration.

(b) Beliefs, like all other propositional attitudes, not only represent objects but seem to always represent them under some mode of presentation or other. Beliefs about identity are often taken to offer the clearest illustration of this fact. For example: how can Aunt May swear she would kill Spider-man next time she meets him, while at the same time resolving to bake a cherry pie for him? Has she gone senile? Not quite: she is simply thinking about him under two different modes of presentations, namely “Spider-man” in one case and “my nephew” in the other, all the while ignoring that, so to speak, one person wears both masks. In the following, I will refer to this feature of propositional attitudes as their *intensionality*, with an *s*. The reason this is important for us is that registrations and encounterings, not being intentional states, cannot be intensional either. Aunt May's registration (/encountering) of her nephew cannot represent him one way or the other, since, after all, it does not represent him at all. If Aunt May encounters Spider-man swinging over 5th Avenue, she must also encounter her nephew in the exact same location. It follows that, if an infant were keeping track of what Aunt May encounters and registers as opposed to what she sees and believes, the infant would be unable to understand why Aunt May shouts angrily at Spider-man as he swings by above her, instead of, say, trying to feed him cherry pie.

Causal role. Beliefs interact with other mental states, often in very complex ways. For example, beliefs ordinarily interact with other beliefs in inference, and with desires, intentions and emotions in decision-making. In contrast, registrations have a rather streamlined causal role; they do not take part in reasoning, either of the theoretical or of the practical kind. The only way one can register an object is by encountering it; more precisely, agents register objects in the location where they last encountered them. Thus, if John last encountered his car in the garage, he will register it there, and will continue to do so even upon hearing the alarm go off and despite his wife's attempts to inform him that the car has been stolen.

In addition, beliefs are not just causes of action but also reasons for action, and thus possess a normative dimension, being subject to the norms of rationality and truth. For example, given the evidence one has, there are things one ought to believe and, generally speaking, one ought to believe the truth; furthermore, given what one believes and wants, there are things one ought to do. In the case of registrations, this normative dimension would be completely absent.

Let us see an example of the mToM account in practice; consider (Onishi & Baillargeon, 2005). The infants in this study expected E1 to reach for the green box, which is where she thought the toy was, instead of the yellow box, which is where the toy really was. Notice that E1's belief concerns nothing more than the location of the toy and will thus have a matching registration. Furthermore, notice that the location where E1 takes the toy to be is also the location where she last saw it; thus, that will also be the location where she last encountered it, and thus the location where she registers it. Notice, finally, that while E1's belief is intensional, its intensionality can be safely ignored in predicting her action. E1 must be thinking of the toy in some way; perhaps as "the thing I put in there" or as "my dearest possession." How she thinks of it, however, is irrelevant; what matters is what she believes of it ("that it is in the green box".) For all these reasons, ascribing to E1 a registration of the toy in the green box (together with the goal of retrieving it) will suffice for predicting her action.

3.3.2 Problems

On reflection, it should not be surprising that the mToM account can explain the results of Onishi and Baillargeon (2005), since it exploits the same vulnerability targeted by the behaviour-rule and affordance-based accounts (§2.1.1 and §2.2, respectively). Indeed, the location where one registers an object is the same where one last encountered it, which is the same location where the object was last in one's unobstructed line of sight. Given this, we should expect the mToM account to face similar problems as Perner and Ruffman's (2005) search-where-last-seen rule; problems that should by now be familiar. To repeat once again: not in all indirect false-belief studies the target agent believes the object to be where she last saw it.

It may be argued, however, that this does not speak against the hypothesis that infants reason about simplified, belief-like states. Perhaps, infants do not reason about a type of state as simple as Apperly and Butterfill originally suggest, but this falls short of showing that they reason about *belief*. The question we need to address, then, is: can the mToM account be expanded so as to explain all indirect false-belief results, while still remaining sufficiently distinct from the face-value interpretation?

Consider again the study by Scott et al. (2015). While E1 is away, E2 replaces the rattling toy which E1 just shook with one that looks the same but does not rattle. When E1 comes back, will she put the toy in the bin (where she has been putting all the non-rattling toys), or in the box (where she has stored all the rattling ones so far)? We immediately stumble upon an obstacle: to predict what E1 will

do with the toy, infants would have to ascribe to her a registration of the toy on the table as one that rattles when shaken. That is, after all, what determines what E1 will do with the toy: whether she takes it to be one that rattles or one that doesn't. The problem, however, is that according to the mToM account agents can only register objects and their locations; their other properties, such as, for example, whether they rattle when shaken, cannot be registered. E1 can register the toy on the table, but she cannot register that it rattles when shaken. Ascribing the first registration without the second will not suffice, however, to figure out whether E1 will keep the toy or throw it away. Lacking such a crucial piece of information, infants should be unable to predict E1's action. Thus, based on mToM, one would expect infants to judge both possible outcomes as equally likely¹⁴.

Suppose, now, that we modify mToM by allowing properties other than location to be registered. Call this revised version of the account mToM*. With this change in place, it is now possible for E1 to register the toy as one that rattles when shaken. Remember, however, that registrations must always be caused by a corresponding encountering. The only way E1 could register the toy as rattling is if she encountered it as rattling; and since the toy in question does not rattle, she could not possibly have encountered it as rattling. In fact, E1's belief that the toy on the table is one that rattles when shaken is based on her assumption that that is the toy she previously shook. In other words, the belief in question is formed as the result of an inference from some of E1's other beliefs. So, again, we face a serious obstacle: infants would still be unable to ascribe to E1 a registration of the toy as a rattling one because E1 has never heard that toy rattle, and that is the only way she could form such a registration according to mToM*.

Suppose, now, that we modify mToM* even further, to allow registrations to interact with other registrations in something analogous to inference; call this mToM**. We would still not be out of the woods. E1's belief that the toy on the table is a rattling one isn't inferred from any old belief; it is inferred from the belief that that is the toy she previously shook; in other words, from a belief about the identity of the toy¹⁵. This is important because beliefs about identity are paradigmatically intensional. The belief that

¹⁴ In fact, since E2 took the non-rattling toy from the bin containing the toys E1 had already discarded, infants might expect him to put it back there.

¹⁵ Commenting on other studies which seem to show that infants can ascribe false beliefs about identity (Scott & Baillargeon, 2009), Butterfill and Apperly (2013) suggest that these studies can be explained on the assumption that infants ascribe registrations about types, as opposed to identity (2013, p. 626). Thus, instead of ascribing to E1 the belief that the toy on the table is *the one that rattles* (a belief stating the identity of two token objects), infants could ascribe the belief that the toy on the table is one that rattles (a belief about the type instantiated by the token object). Notice, however, that this does solve the problem in the case at hand. This is because, as explained above, the reason

(1) the toy on the table is the toy on the table

is not nearly as informative as the belief that

(2) the toy on the table is the one E1 previously shook,

even though both claim that one and the same object is identical to itself (the difference is in the way the object is represented: as the toy on the table, or as the toy E1 previously shook.) Notice that (2), but not (1), would allow E1 to infer that the toy on the table is one that rattles when shaken, given that the toy she previously shook did in fact rattle. As we have seen, however, there is no way of translating beliefs like (2) in registration-speak. The closest one can get is the registration of the toy as identical to itself, and this clearly will not do. Thus, even on mToM**, infants would still be unable to ascribe to E1 a registration of the toy as one that rattles. Even though registrations can now be “inferred” from other registrations, we still lack a suitable registration from which E1 could infer that the toy on the table is one that rattles when shaken.

It seems the only way to solve this third problem would be to concede that registrations, like beliefs, are intensional (and thus, intentional)¹⁶; call this mToM***. It would now be possible for E1 to register that the toy on the table is the same she previously shook, from which she could then infer that the toy on the table rattles when shaken. Notice, however, that in this way we would have effectively erased, one by one, all the main differences between registration and belief. Like belief, “registration” is now an intentional mental state that can take part in inference and is imbued with intensionality. In our attempts to explain Scott and colleagues' finding, we seem to have ended up turning mToM into the face-value interpretation it was supposed to replace!

3.4 Summary

The fact that infants expect agents that do not know the location of an object to behave differently from agents that have a false belief about its location rules out both the desire-awareness account and the perceptual-

E1 believes that the toy is a rattling one is that she thinks that that is the same toy she previously shook. It seems that this *second* belief cannot be replaced by one about types of objects.

¹⁶ There is a couple of studies suggesting, instead, that infants possess no understanding of intensionality (Low, Drummond, Walmsley, & Wang, 2014; Low & Watts, 2013). These studies will be discussed in Chapter V, §2.3.

access reasoning account. On the other hand, the fact that infants can predict what an agent will do even when the agent does not take the object to be where she last saw it speaks against the mToM account. In principle, the mToM account can be modified and expanded to account for this fact. I have argued, however, that this would bring the mToM account dangerously close to the face-value interpretation, erasing the main differences between the two.

4. Conclusions

None of the alternative interpretations that we have considered in this chapter offer a satisfactory account of indirect false-belief results. This provides evidence that the face-value, early-emergentist interpretation may be on the right track. Of course, the evidence is less than conclusive but, arguably, it is strong enough to take the early-emergence approach on as a working hypothesis. The question that we need to address, now, is whether the early-emergence approach can explain why young children tend to fail direct false-belief tests. This will be taken up in the next chapter.

III

ALTERNATIVE INTERPRETATIONS OF DIRECT FALSE-BELIEF TESTS

In this chapter, we are going to discuss whether early-emergence accounts can explain the evidence that, *prima-facie*, appears to support late-emergence accounts. The main piece of evidence supporting late-emergence accounts is, of course, the finding that most children under four years of age fail direct false-belief tests. There is, however, a good deal of additional evidence supporting late-emergence accounts, which must be taken into account too. We will thus start in §1 by briefly reviewing these additional findings. Then, in §2, we will discuss the main proposals early-emergentists have defended in their attempts to explain away the evidence for the late-emergence position.

1. Additional Evidence for Late-Emergence Accounts

1.1 Developmental Progression

As mentioned in Chapter I, studies employing the ToM scale have shown that most children pass the following direct mentalising tests in a specific order: diverse-desire test first, then diverse-belief test, then knowledge-access test, then false-belief test, then finally the hidden-emotion test (Wellman et al., 2011; Wellman & Liu, 2004). Interestingly, Iranian and Chinese children have been found to follow a slightly different progression, passing the knowledge-access test before the diverse-belief test (Shahaecian, Peterson, Slaughter, & Wellman, 2011; Wellman, Fang, Liu, Zhu, & Liu, 2006). Wellman and colleagues (2011) argue that this developmental progression reflects a process of “conceptual change,” where children revise their theories of mind based on the evidence they are exposed to, including both their first-hand observations and what adults say about the mind. On this type of account, the fact that Iranian and Chinese children follow a slightly different progression from Western children is simply a reflection of the differences between the respective cultural environments. Iranian and Chinese cultures, being more collectivist, are likely to put more stress on common knowledge as opposed to the individuality of beliefs, explaining why the knowledge-access test is passed before the diverse-belief test.

1.2 Executive Functioning

Executive functioning is an umbrella term grouping those higher cognitive processes that are involved in managing, supervising and overriding other cognitive processes. Among the processes that psychologists typically include in the “executive” category, inhibitory control, working memory and cognitive flexibility are perhaps the main ones (Diamond, 2013). Inhibitory control is the ability to inhibit impulses and automatic responses and resist temptations; working memory is the ability to hold information in mind and manipulate it; cognitive flexibility is the ability to shift between different mental sets or consider different aspects of a problem.

Several studies show that performance on direct false-belief tests correlates positively with performance on tests of inhibitory control and working memory (Carlson & Moses, 2001; Carlson,

Moses, & Breton, 2002; Davis & Pratt, 1995; Duh et al., 2016; Gordon & Olson, 1998; Keenan, Olson, & Marini, 1998). Carlson and Moses (2001), for example, tested a group of three- and four-year-olds using an inhibitory-control battery, a mentalising battery, and a test of language ability. The authors found that mentalising tests, including direct false-belief tests, were robustly correlated with tests of inhibitory control; after age and language ability were controlled for, there was still a significant correlation between the mentalising aggregate and the inhibitory control aggregate, of moderate size ($r=.41$). A recent meta-analysis by Devine and Hughes (2014), which looked at over a hundred studies, confirms that the correlation with executive functioning is robust, although it suggests it may be weaker than initially assumed: for studies that controlled for age and verbal ability, the correlation coefficient was only .22.¹

What explains this correlation? Typically, two types of explanation are taken into account. According to “expression” accounts, performance on executive functioning and direct false-belief tests correlates because executive functioning is required to express or display one's belief-reasoning abilities. According to “emergence” accounts, in contrast, the two correlate because executive functioning is required to learn about the mind and about beliefs specifically, which in turn is required for passing false-belief tests. Notice that these two types of account are not mutually exclusive: it is possible that executive functioning plays both roles, and that both contribute to the correlation.

A recent experiment by Lindsey Powell and Susan Carey (2017) provides some of the best evidence for expression accounts so far. Powell and Carey took advantage of the phenomenon of ego depletion: tasks that are taxing on executive functioning can lead to its depletion, leading to poor performance in any subsequent tasks that also rely on this cognitive function. If executive functioning is required to express one's mentalising abilities, as expression accounts maintain, then performing an executive functioning task before a false-belief test should negatively affect performance on the latter. To test this hypothesis, Powell and Carey gave four and five-year-olds either

¹ Devine and Hughes point out that the studies that report larger effect sizes are those that use aggregate ToM and EF scores, as opposed to focusing on the relation between specific tests. The explanation Devine and Hughes provide is that using aggregate scores enhances variance, which can lead to larger effect sizes. This is important because it suggests that the small effect size found is not due to a lack of specificity in the analysis; in fact, the effect would have been smaller had Devine and Hughes restricted their analysis to studies that focused on the relation between (say) DFB and IC tests.

an executive functioning task (experimental condition) or a “filler” control task (control condition) immediately before a false-belief test. They found that, as predicted by expression accounts, performance on the false-belief test was significantly better in the latter (control) condition compared to the former (experimental) condition.

Other findings speak against a pure expression account and in favour of an expression-emergence hybrid, however (Benson, Sabbagh, Carlson, & Zelazo, 2013; Carlson, Claxton, & Moses, 2015; Sabbagh, Xu, Carlson, Moses, & Lee, 2006). Let us take a quick look at these studies.

- The study by Mark Sabbagh and colleagues (2006) shows that although Chinese preschoolers perform significantly better than age-matched American children on an executive-functioning battery, they do not perform significantly better on a direct false-belief battery, even though the two variables correlate within each group. That is: Chinese children who score higher on tests of executive functioning compared to their Chinese peers tend to also score higher on direct false-belief tests, and the same is true for American children of the same age; however, when the two groups are pooled, the correlation disappears, since Chinese children score higher than their American peers on executive-functioning tests but not on direct false-belief tests. This suggests that there is some other variable besides executive functioning that is negatively affecting Chinese children's performance on direct false-belief tests. Sabbagh and colleagues point to the fact that Chinese children are much less likely to have siblings and will thus have less opportunities to engage in conversations about the mind (the evidence on siblings and mental-state talk will be discussed in §1.4). On the assumption that engaging in this type of social interaction helps children learn that beliefs can be false, the fact that Chinese children are less exposed to this type of stimulus could balance out the beneficial effect of their superior executive abilities. At the same time, since inhibitory control is also required to learn that beliefs can be false, performance on tests of inhibitory control correlates with performance on false-belief tests within each group, where exposure to mental state talk is kept constant or at least varies to a lesser degree.
- The study by Stephanie Carlson and colleagues (2015) shows that performance on tests of inhibitory control correlates with performance on two mentalising tests that do not appear

to make demands on executive functioning, namely the think-know¹⁷ task (Moore, Pure, & Furrow, 1990) and the sources of knowledge¹⁸ task (O'Neill & Gopnik, 1991). The reason Carlson and colleagues are confident that these mentalising tests are not taxing on inhibitory control is that young children tend to answer randomly in these tests, which suggests that their poor performance is not the result of a failure to inhibit a prepotent response (in which case one would expect them to consistently give the wrong answer, as it happens in direct false-belief tests.) Notably, these direct mentalising tests were found to correlate with tests of inhibitory control as strongly as a direct false-belief test. As the authors point out, this is consistent with emergence accounts but not with (pure) expression accounts. If the reason performance on false-belief tests correlates with performance on inhibitory-control tests is that false-belief tests load heavily on inhibitory control, as expression accounts suggest, one would not expect the correlation to hold for mentalising tests that do not load on inhibitory control. On the other hand, if the reason the correlation exists is that inhibitory control is required to learn about the mind, as emergence accounts maintain, then it is not surprising that the correlation extends to all mentalising tests independently of their task demands.

- The study by Jeannette Benson and colleagues (2013) shows that three-year-olds score on a battery of executive-functioning tests predicts how much they will benefit from subsequent training on a direct false-belief test. This is just as emergence accounts predicts: having better executive functioning helps children acquire the relevant folk psychological generalisations during training, which then results in better direct false-belief performance in the post-test.

¹⁷ “Two identical monkey puppets provided information regarding the location of a sticker hidden in one of two boxes (blue and red). The experimenter wore one puppet on each hand and animated the puppets with identical voices. For four trials, monkeys alternately used low certainty (think) or high certainty (know) terms. For example, Monkey A said, “I think the sticker is in the red box,” and Monkey B said, “I know the sticker is in the blue box.” [...] Children received a point each time they chose the box indicated by the “know” monkey...” (Carlson et al., 2015, p. 189).

¹⁸ “Children were shown a red cardboard tunnel with a green cloth flap covering the openings at both ends. They were told that different objects would be placed inside the tunnel and that they would learn the identity of the object by looking inside and seeing it, putting their hands inside and feeling it, or being told what was inside. [...] After learning the identity, children were then asked how they knew the identity of the object” (Carlson et al., 2015, p. 189).

1.3 Language Ability

Many studies show that children's direct false-belief performance correlates positively with their language abilities (e.g., Astington & Jenkins, 1999; Hughes, 2005; Ruffman, Slade, Rowlandson, Rumsey, & Garnham, 2003). To illustrate, Astington and Jennifer Jenkins (1999) carried out a longitudinal study where they tested a group of three-year-olds three times over a period of a few months, using both a mentalising battery (which included an unexpected-transfer task, an unexpected-contents task and an appearance-reality task, all direct) and a test of general language ability (which assessed both semantics and syntax, both receptive and expressive.) Astington and Jenkins found that children's early language ability predicted their later mentalising score, even after controlling for their age and earlier mentalising score (the correlation coefficient was in the order of .3, indicating a weak to moderate correlation.) A meta-analysis by Karen Milligan, Astington and Lisa Dack (2007), which looked at more than a hundred studies, found that, after controlling for age, language ability was still significantly correlated with performance on mentalising tests, including false-belief tests, again with a correlation coefficient in the order of .3. Notably, all aspects of language ability covered in the meta-analysis (general language, semantics, syntax, receptive vocabulary and memory for complements) were found to correlate with mentalising, and while these individual correlations differed in strength, the differences were mostly non-significant.

Two training studies provide further evidence that improving children's language ability brings about a corresponding improvement in their performance on direct mentalising tasks, as opposed to merely correlating with it (C. Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003). Courtney Hale and Helen Tager-Flusberg (2003), for example, trained a group of three and four-year-olds either on a direct false-belief test, or on a sentential complement test,² or on a relative-clause control. While children in the false-belief training group only improved their direct false-belief performance, and children in the relative-clause group only improved their performance on the relative-clause control, children in the sentential-complement training group improved both their direct false-belief performance and their sentential-complement performance. Notably,

² The sentential complement test, based on Jill de Villiers and Jennie Pyers (2002), consisted of telling a story about a character who said she was doing something while she was doing something else, and then asking children what the character said. This type of test is meant to assess children's facility with verbs that take complement clauses (e.g., believe that, say that, etc.) which are supposed to correspond to reality but not always do.

sentential-complement training was found to be as effective as false-belief training in improving children's direct false-belief performance.

Late-emergentists have defended two types of explanations for this correlation, which closely map the “expression” and “emergence” accounts considered for the correlation with executive functioning in the last section (§1.2). The first maintains that natural language provides the required representational medium for mentalising (P. A. de Villiers & de Villiers, 2012). The second maintains that participating in conversations about mental states allows children to learn about the mind, and that children's ability to take part in or follow such conversations is mediated by their language abilities (Milligan et al., 2007).

1.4 Experiential Factors

There is evidence that being exposed to the type of stimuli that would allow one to learn about the mind does improve children's performance on direct false-belief tests. The stimuli in question include participating in training sessions on false-belief scenarios, being exposed to conversations about the mind and interacting with a sibling or friend.

1.4.1 Training

Several studies show that children's direct false-belief performance can be improved through training (Clements, Rustin, & McCallum, 2000; Hofmann et al., 2016; Lecce, Bianco, Demicheli, & Cavallini, 2014; Melot & Angeard, 2003; Perner & Kloo, 2003; Rhodes & Wellman, 2013; Slaughter & Gopnik, 1996). Virginia Slaughter and Gopnik (1996), for example, gave three- and four-year-olds two training sessions over a period of two weeks. Children were divided in three conditions: a false-belief condition, a coherence condition, and a control condition. Children in the false-belief condition were trained on a false-belief task with feedback; children in the coherence condition were trained on tasks that involved ascribing desires or perceptual experiences different from their own; finally, children in the control condition were trained on control tasks that involved reasoning about the contents of a bag or a Tupperware container. In the retest after training, children were also given the appearance-reality, think-know and sources-of-knowledge tests, plus a number conservation task that was used as a further control. Slaughter and Gopnik found that the false-belief and coherence training improved children's performance on the false-belief, appearance-reality,

think-know and sources-of-knowledge tests, and did so more than the control training. In contrast, false-belief and coherence training did not improve children's performance on the number conservation task. Accordingly, Slaughter and Gopnik take their finding to show that false-belief and coherence training both helped children construct a better theory of mind.

1.4.2 Mental-state Talk

There is evidence that being exposed to conversations about mental states can also lead to an improvement in direct false-belief performance (Adrian, Clemente, & Villanueva, 2007; Adrian, Clemente, Villanueva, & Rieffe, 2005; Guajardo & Watson, 2002; Ruffman, Slade, & Crowe, 2002; Slaughter, Peterson, & Mackintosh, 2007). For example, Ted Ruffman, Lance Slade, and Elena Crowe (2002) conducted a longitudinal study with a group of three- and four-year-olds that were tested at three points in time, roughly six months apart. At each time-point, the testing comprised a mentalising battery (which included a false-belief test), a test of language ability, and a “picture test” where mothers were asked to describe pictures of people engaging in varying activities to their children. Ruffman and colleagues coded the language both mothers and children used in the picture test, noting in particular the amount of references they made to thoughts, feelings and desires. They found that the frequency with which mothers referenced mental states predicted children's performance on mentalising tests at later times; this remained true even after controlling for a number of potential mediators, such as children's performance on both mentalising and language tests and the frequency with which they talked about mental states at earlier time points. In contrast, children's earlier mentalising performance did not predict their mothers' mental state utterances at later times.

1.4.3 Siblings

There is evidence that children with at least one sibling tend to perform better on direct false-belief tests, compared to those with no siblings (Hughes & Ensor, 2005; McAlister & Peterson, 2006, 2007, 2013; Perner, Ruffman, & Leekam, 1994; Peterson, 2000; Ruffman, Perner, Naito, Parkin, & Clements, 1998). For example, Anna McAlister and Candida Peterson (2006) tested one-hundred-twenty-four children between the ages of three and five using a mentalising battery, which included two direct false-belief tests. Children with at least one “child” sibling (operationalised as a sibling between the ages of one and twelve) tended to score higher than those with no child

siblings on both mentalising and executive functioning tests; in both cases, the correlation coefficients were weak (in the order of .2) but significant.

The standard, late-emergentist explanation for this correlation is that having a sibling provides children with more opportunities for social interaction, which in turn provides them with more evidence that people can have false beliefs (Perner et al., 1994). Consistent with this interpretation, other variables that are likely to enhance opportunity for social interaction have also been found to correlate with direct false-belief performance, including the number of older children and adults children interact with daily (Lewis, Freeman, Kyriakidou, Maridaki-Kassotaki, & Berridge, 1996), the number of days children spend playing with peers (Shahaeian, Henry, Razmjooe, Teymoori, & Wang, 2015) and whether children have classmates of different ages (Wang & Su, 2009). It is possible, as well, that having a sibling gives children more opportunities to engage in conversations about the mind. Supporting this hypothesis, Jane Brown, Nancy Donelan-McCall and Judy Dunn (1996), after analysing a sample of four-year-olds' conversations recorded at home, found that four-year-olds were more likely to talk about their own or someone else's mental states when they interacted with their friends and siblings as opposed to their mothers, while a study by Jenkins, Sheri Turrell, Yuiko Kogushi and colleagues (2003), which looked at how parents interacted with their children at home, found that four-year-olds with an older sibling were more exposed to mental-state talk than children without an older sibling.

1.5 Deaf Children

Deaf children are known to perform much worse on direct false-belief tests than typically developing children. For example, Peterson and Michael Siegal (1995) report that the deaf children in their study, with ages ranging between eight and thirteen, were 17% correct on a standard unexpected-transfer task. In a follow-up study, Peterson and Siegal (1999) found that deaf children's direct false-belief performance was highly dependent on whether their parents were also deaf; specifically, deaf children of deaf parents performed like typically developing children, while deaf children of hearing parents were severely impaired, as Peterson and Siegal (1995) had found. A study by Russell et al. (1998), which looked at deaf children aged from four up to sixteen years, found that their direct false-belief performance improved with age, suggesting that deaf children are not irreversibly impaired in mentalising but simply exhibit a developmental delay. Later studies

confirmed this finding and extended it to other direct mentalising tests. Specifically, studies employing the ToM scale (discussed in Chapter 1, §2.2.4 and in this chapter, §1.1) show that deaf children of hearing parents follow a similar developmental progression though the five items of the ToM scale as typically developing children and deaf children of deaf parents, but on a timetable delayed by several years (Peterson, Wellman, & Liu, 2005; Peterson, Wellman, & Slaughter, 2012). Notably, an intervention study by Wellman and Peterson (2013) shows that this delay can be alleviated through training on a direct false-belief task, which was found to improve deaf children's performance not just on direct false-belief tests but across the ToM scale.

What is the reason deaf children of hearing parents have poorer direct false-belief performance than deaf children of deaf parents? Parents of deaf children are unlikely to be proficient in sign language unless they are themselves deaf; as a result, deaf children of hearing parents typically do not acquire sign language at home and, as a result, are delayed in their language development. As discussed in §1.3, some late-emergentists have argued that language is required for representing beliefs and other propositional attitudes; since deaf children of hearing parents typically have language delays, it makes sense that they should be correspondingly delayed in their mentalising abilities (P. A. de Villiers & de Villiers, 2012). In addition, especially during the first years of life, deaf children will have a much harder time communicating with their parents when their parents have normal hearing as opposed to being deaf, and this can be expected to be all the more likely when abstract topics, like thoughts and feelings, are concerned. As discussed in §1.4, some late-emergentists have argued that participating in conversations about mental states helps children learn that people can have false beliefs. Since deaf children of hearing parents are deprived of this opportunity for learning, it should take them longer to learn that beliefs can be false (Peterson & Siegal, 1995, 1999).

1.6 Social Competence

There is robust evidence that children's performance on direct false-belief tests correlates with their performance on several tests of social competence. In particular, preschoolers who perform better on direct false-belief tests are more likely to tell a lie to conceal a transgression (Talwar & Lee, 2008), are better at keeping a secret and at playing hide-and-seek (Ding, Wellman, Wang, Fu, & Lee, 2015; Peskin & Ardino, 2003), score higher on social competence tests (Razza & Blair,

2009), tend to be more popular in the classroom and have more friends (Cassidy, Werner, Rourke, Zubernis, & Balaraman, 2003; Fink, Begeer, Peterson, Slaughter, & De Rosnay, 2015; Slaughter, Imuta, Peterson, & Henry, 2015), are better at producing persuasive arguments (Kołodziejczyk & Bosacki, 2015; Peterson, Slaughter, & Wellman, 2017; Slaughter, Peterson, & Moore, 2013) and more likely to engage in pro-social behaviour (Imuta, Henry, Slaughter, Selcuk, & Ruffman, 2016). Not surprisingly, such findings are taken to support a late-emergence account: children who pass direct false-belief tests know that beliefs can be false and can exploit this knowledge in their social interactions, while children who fail direct false-belief tests, not knowing that beliefs can be false, cannot (Sabbagh et al., 2013; Wellman, 2014). Let us zoom in on some of these findings.

- Joan Peskin and Vittoria Ardino (2003) investigated the relationship between three- to five-year-olds' ability to keep a secret and play hide-and-seek and their mentalising skills. Mentalising was measured using two direct false-belief tests. In the hide-and-seek task, participants had to play two rounds of hide-and-seek with an experimenter. In the keeping-a-secret task, participants had to keep from spoiling a surprise (i.e., keep from telling the experimenter that someone had bought a birthday cake for her.) The correlation coefficient between the mentalising aggregate and the social (hide-and-seek + keeping-a-secret) aggregate was .42 once age was partialled out.
- Peterson, Slaughter and Wellman (2017) investigated the link between mentalising and persuasion skill in a large sample of children ranging from three to twelve years of age, which included typically developing children, deaf children of hearing parents and children with autism spectrum disorder. Mentalising was measured using the ToM Scale plus an additional battery of direct false-belief tests, while persuasion skill was measured using several peer persuasion tasks. To illustrate, one of these tasks involved convincing a puppet (“Matty”) to do something he was not inclined to do, like brushing his teeth or eating broccoli. Matty would resist children's persuasion attempt, at which point children were encouraged to try again; the whole procedure was repeated three times. Children received one point for each “high-level” argument they used, which the authors define as arguments that “had to supply new substantive information aimed at shifting Matty's initial mental state of resistance”. Peterson and colleagues found that, in all three groups (typically developing, deaf and autistic) children's performance on direct mentalising tasks was

correlated with their performance on peer persuasion tasks. In typically developing children, for example, direct mentalising performance independently predicted 16% of the variance in peer-persuasion performance, over and above age and language.

- Rachel Razza and Clancy Blair (2009) conducted a longitudinal study of the relationship between mentalising and social competence. In the initial testing (T1), the mean age was 5 years and two months; the second testing (T2) was a year later. Mentalising was measured using a battery of direct false-belief tasks, while social competence was measured using the Preschool and Kindergarten Behavioural Scales, which is a standardised report compiled by teachers, which is meant to provide an assessment of children's typical class behaviours; it includes items pertaining to social cooperation and social interaction. Razza and Blair found a bidirectional link between direct false-belief performance and the social competence score: direct false-belief performance at T1 was correlated positively with social competence at T2, and social competence at T1 was correlated positively with direct false-belief performance at T2.

1.7 Brain Maturation

Neuroscientific investigations suggest that several cortical areas are involved in mentalising, including the temporo-parietal junction (TPJ), the medial prefrontal cortex (mPFC), the precuneus (PC) and the superior temporal sulcus (STS) (Koster-Hale & Saxe, 2013). The TPJ, in particular, has been argued to be a core component of the mentalising network (e.g., Saxe and Kanwisher, 2003; Samson et al., 2004). A couple of very suggestive studies have found that children's performance on direct false-belief tests correlates with the maturation of their mentalising network, including the TPJ (Sabbagh, Bowman, Evraire, & Ito, 2009; Wiesmann, Schreiber, Singer, Steinbeis, & Friederici, 2017). This seems to be an important finding, since it provides independent evidence that whether children pass or fail a direct false-belief test depends, at least in part, on their mentalising abilities. Let us take a look at these studies.

- Sabbagh and colleagues (2009) tested a group of four-year-olds using electroencephalogram (EEG). The premise of the study was that resting alpha activity increases in amplitude and coherence over the preschool years, due to maturational changes that take place in the brain. The increase in alpha coherence, in particular, is thought to reflect the extent

to which the neurons in a given region are organised, thus firing in a more synchronous manner. Sabbagh and colleagues recorded alpha rhythm while children were at rest, and then analysed their data (using a technique called sLORETA) to estimate current density (which can be taken as a measure of alpha coherence) over different cortical regions. Sabbagh and colleagues found that current density over the right TPJ and the dorsal mPFC was moderately correlated with performance on direct mentalising tasks, including two direct false-belief tests. The authors suggest a late-emergentist interpretation of this finding: the right TPJ and dorsal MPFC become more functionally organised as children acquire a “representational” theory of mind.

- Charlotte Wiesman and colleagues (2017) used diffusion-weighted magnetic resonance imaging (dMRI) to investigate white matter maturation in a group of three- and four-year-olds. White matter is made up of axons connecting the neurons in the cortex (the “grey” matter). Diffusion-weighted MRI can be used to compute fractional anisotropy, an index that is used to estimate how water molecules are diffused in biological tissue; a value of 0 corresponds to equal diffusion in every direction while a value of 1 corresponds to diffusion along just one dimension. Given that white matter is made up of bundles of axons disposed in parallel fashion, it is expected to exhibit relatively high levels of fractional anisotropy. Notably, as the brain matures the axons that make up white matter grow and get more myelinated (essentially, more insulated), leading to increases in fractional anisotropy during childhood. Now, Wiesmann and colleagues found that children who had higher levels of fractional anisotropy in the white matter surrounding some of their mentalising areas (including, among others, the right TPJ and PC) tended to perform better on direct false-belief tests. Notably, like Sabbagh and colleagues, Wiesmann and colleagues also suggest a late-emergentist interpretation of their finding, according to which changes in white matter maturation correspond to the acquisition of a representational theory of mind.

2. Performance Accounts

Early-emergentists argue that young children, despite possessing belief-reasoning abilities, fail to display them in the context of direct false-belief tests because of performance difficulties. So

far, early-emergentists have pointed to two types of factors that could be negatively affecting young children's performance in direct false-belief tests. The first is a limitation in the processing resources young children have at their disposal. Several authors have argued that direct false-belief tests make high demands on processing resources, such as executive functioning, that are especially poor in young children (Baillargeon et al., 2010; Carruthers, 2013; Leslie et al., 2005; Setoh et al., 2016). Call this a processing-load type of account. The second type of performance-limiting factor early-emergentists have appealed to has to do with children's capacity to correctly interpret the linguistic stimuli used in the test. Specifically, children are argued to have difficulties with the process of pragmatic interpretation, having to do not with decoding the literal meaning of an utterance but with figuring out what the speaker means by it (Helming, Strickland, & Jacob, 2014; Helming et al., 2016; Westra, 2016; Westra & Carruthers, 2017). Call this a pragmatic type of account. In the following, we will discuss five main proposals, the first three of which fall into the processing-load category while the last two belong to the pragmatic category.

2.1 The ToMM/SP Account

As mentioned in Chapter I, §3.2.1, Alan Leslie was among the first to defend an early-emergence account; he argued that the cognitive mechanism underlying the capacity to represent mental states, which he called ToMM (short for “Theory of Mind Mechanism”), was innate and would typically be up and running by the second year of life, if not earlier. To explain why most two and three-year-olds had trouble with direct false-belief tests, Leslie and some of his colleagues developed the *ToMM/SP account*, according to which ascribing false beliefs requires the contribution of a further cognitive mechanism besides ToMM, called selection processor or SP for short (Friedman & Leslie, 2004; Leslie, 2000b; Leslie, Friedman, & German, 2004; Leslie & Polizzi, 1998; Leslie & Thaiss, 1991; Roth & Leslie, 1991, 1998). On this account, the reason SP is required is that ToMM is likely to have a true-belief bias, ascribing true beliefs by default. In order to ascribe a false belief, then, one would first need to override the default attribution, and this is where SP comes in. SP is supposed to be a domain-general process playing an “executive” role, having to do with managing other cognitive processes and overriding their responses when needed. Now, if SP is not strong enough to carry out the override, the true-belief default will prevail, resulting in the ascription of a true belief. This, according to the ToMM/SP account, is precisely what happens in young children:

even though both their ToMM and their SP are already functioning, their SP is still too weak to override the true-belief bias, thus leading them to failure.

2.1.1 Indirect false-belief tests

While the results of indirect false-belief studies support Leslie's claim that ToMM is innate and early-emerging, they are also in clear tension with his ToMM/SP account. This is because these studies show that even infants can ascribe false beliefs, which suggests one of two things: either SP is not required to ascribe a false belief, or SP is required but already sufficiently developed in infancy.

Could it not be, however, that some feature of direct false-belief tests makes the true content more salient, thus making it more difficult to inhibit? An early study by Clements and Perner (1994), which combines direct and indirect methodologies, seems to rule out this possibility. Clements and Perner essentially carried out an anticipatory-looking unexpected-transfer task with the typical direct question added at the end. Clements and Perner found that many of the children who had just looked at one location in anticipation of the agent's return tended to point to the other location when asked where the agent would go, thus displaying both responses in short succession.

2.1.2 Executive functioning

In discussing evidence that children's performance on direct false-belief tests correlates with their performance on tests of inhibitory control, (Friedman & Leslie, 2004; Leslie, 2000b) suggests that 'SP' and 'inhibitory control' may refer to the same cognitive mechanism/capacity. On this assumption, the ToMM/SP account would be consistent with an expression account of the correlation, which maintains that inhibitory control and direct false-belief performance correlate because inhibitory control is required to express one's belief-reasoning abilities. So far, so good; however, there are two considerations that speak against this explanation of the correlation.

- First, remember that the meta-analysis by Devine and Hughes (2014) found the correlation between performance on direct false-belief tests and performance on executive functioning tests (including tests of inhibitory control) to be quite low for studies that controlled for language ability and age ($r = .22$, which corresponds to about 5% of the variance.) Since

on the ToMM/SP account a lack of inhibitory control is the main reason young children fail direct false-belief tests, one would expect the correlation to be much stronger.

- Second, the ToMM/SP account cannot explain the evidence for emergence accounts of the correlation (discussed in §1.2), according to which inhibitory control and direct false-belief performance correlate because inhibitory control helps children learn that beliefs can be false. Consider, for example, the study by Sabbagh and colleagues (2006), which found that Chinese children did not perform better than American children on a direct false-belief test despite outperforming them on a test of inhibitory control. This is problematic if we assume (as the ToMM/SP account does) that a lack of inhibitory control is the main factor holding back young children's direct false-belief performance, since in this case one would expect an improvement in the former to result in an improvement in the latter. Or consider the study by Carlson et al. (2015), which found that children's performance on the think-know and sources-of-knowledge tests correlated with an inhibitory-control test as strongly as their performance on a direct false-belief test. This is problematic because neither the think-know test nor the sources-of-knowledge test requires ascription of false beliefs; as a result, on the ToMM/SP account, one would not expect SP to be recruited in these tests.

2.1.3 Language Ability

We have said that, on the ToMM/SP account, children's inhibitory control is the main factor determining failure or success on a direct false-belief test. This, however, presupposes that children correctly understand the linguistic stimuli in the test. If there is a sizable portion of two-, three- and four-year-olds for which these linguistic stimuli are still somewhat challenging, then language ability should also be acknowledged as an important factor in predicting failure or success. Thus, a possible explanation for why children's performance on tests of language ability correlates with their direct false-belief performance is that children with better language ability are less likely to misinterpret the linguistic stimuli in direct false-belief tests. Several considerations speak against this type of explanation, however.

- First, it is worth noting that some two-and-a-half-year-olds can pass indirect false-belief tests that involve linguistic stimuli comparable to those used in direct false-belief

tests. In the first experiment reported by Scott et al. (2012), for example, children were narrated an unexpected-transfer type of story as they looked at a picture book illustrating the events in the narration, and their looking behaviour suggested they predicted where the character in the story would look (this experiment was described in Chapter I, §1.2.2). In a second experiment, children watched as an experimenter was given a direct, unexpected-transfer task, and looked longer when the experimenter answered incorrectly. While the unexpected-transfer task was enacted by two other experimenters instead of narrated, the test question was similar to that typically used in direct false-belief tests (i.e., “When Sally comes back, she is going to need her toy again. Where will she think it is?”) Thus, the first experiment suggests that, by two years and a half, most children are able to understand the narration used in direct expected-transfer tasks, while the second experiment suggests that they are also able to interpret the test question used in those tasks.

- Language and direct false-belief performance are found to correlate not just in younger children (two- and three-year-olds) but also in older ones, up to seven years of age at least (Milligan et al., 2007). Given that, as we have just seen, even the youngest children seem already capable of interpreting the linguistic stimuli in direct false-belief tests, it seems unlikely that some four- or five-year-olds would still be unable to do so, in a way that leads them to failure. (This last qualification is important: to fail, it is not sufficient that children misinterpret some of the linguistic stimuli in the test -- they must do so in a way that causes them to miss some of the critical bits, such as, for example, where the agent put the marble.)
- If the correlation with language is simply a result of the language demands made by direct false-belief tests, one would expect the correlation to be stronger for those tests that involve more (or more complex) linguistic stimuli. Milligan and colleagues, however, found that the four direct mentalising tasks they took into account in their meta-analysis (unexpected-transfer false-belief task, unexpected-content false-belief task, deception task, and belief-emotion task) all correlated with language ability to a similar extent (i.e., no significant difference was found). Nonetheless, as the authors note, these tasks vary considerably in their language demands. For example, the unexpected-transfer task includes a narrative, while the unexpected-content task does not. According to Milligan and colleagues, the fact

that all these tasks correlate with language ability to a similar extent suggests that task-dependent language demands do not fully explain the correlation (Milligan et al., 2007, p. 637).

2.1.4 Developmental Progression

On the ToMM/SP account, the fact that some direct mentalising tasks are passed earlier than others must be a reflection of the extent to which these tasks recruit SP. This seems to work well for the first two steps in the ToM scale, the diverse-desire test and the diverse-belief test.

In the diverse-desire task, the experimenter introduces children to a toy figurine called “Mr. Jones”, who wants something to eat. Then children are asked, we have a carrot and a cookie, which one do you like best? After children answer, the experimenter says that Mr. Jones likes the other type of food best and asks children which one Mr. Jones will pick. Now, to pass this type of test, children must ascribe to Mr. Jones a preference different from their own. On the ToMM/SP account, this does not require SP, since there is no “own-preference or own-desire bias” parallel to the true-belief bias. Furthermore, the test clearly does not involve ascribing a false belief. Thus, the diverse-desire test should be much easier than typical false-belief tests.

In the diverse-belief task, instead, children are introduced a toy figurine called “Linda” and are told that she is looking for something (e.g., her cat), which may be in either of two locations (the bushes or the garage). Children are then asked to guess where Linda's cat is. Whatever children say, the experimenter will then say that Linda thinks the cat is in the other location and will ask them where she will look for it. Now, this test requires children to ascribe a belief that, from their point of view, is false, and thus requires them to use their SP to inhibit the true-belief bias. It thus makes sense that children should pass the diverse-belief test after the diverse-desire test. At the same time, since children simply guessed the location of the cat, their belief about its location should be less salient, making it easier to inhibit it, compared to typical direct false-belief tests. Thus, it makes sense that children should pass the diverse-belief test before they pass false-belief tests.

While the ToMM/SP account works well for these two tasks, explaining the rest of the developmental progression presents some serious obstacles.

- First, the ToM scale has a third step, between the diverse-belief test and the false-belief test: the knowledge-access test. In this task, children are shown the contents of a container (e.g., a toy dog in a drawer) and asked whether someone who has not looked inside it (e.g., “Polly”) knows what's in it. The problem is that passing this task does not require ascribing a false belief to Polly; a state of ignorance (i.e., lack of knowledge) will suffice. Thus, it is not clear why this test is passed after the diverse-belief test.
- Second, remember that cross-cultural studies have found that children in collectivist societies, such as China and Iran, pass the knowledge access test before the diverse-belief test. This cross-cultural variation is difficult to account for on the ToMM/SP account. It is possible, of course, that Iranian and Chinese children differ in their executive skill compared to children in Western countries (indeed, see Sabbagh et al. 2009 concerning Chinese children), but, as argued by Wellman et al. (2011), this would simply change how fast children progress through each of the steps, not the trajectory of their progression.
- Third, a recent study by Ceymi Doenyas, Melis Yavuz and Bilge Selcuk (2018) looked at longitudinal correlations between children's performance on the ToM scale and their performance on tests of inhibitory control. The study had two main findings. First, when examined cross-sectionally, only the knowledge-access test and the false-belief test correlated with inhibitory-control performance. Thus, while on the ToMM/SP account one would expect children's performance on the diverse-belief test to also correlate with inhibitory control (although to a smaller degree) no significant correlation was found in this study. Second, at T2, once age and language ability were controlled for, only direct false-belief performance was predicted by inhibitory-control performance at T1. This suggests that the raw cross-sectional correlation between children's knowledge-access performance and their inhibitory-control performance is not due to the inhibitory demands of the knowledge-access test. To sum up: while on the ToMM/SP account one would expect children's diverse-belief and knowledge-access performance to correlate with their inhibitory-control performance even after potential mediators were controlled for, this study found the opposite result.

2.1.5 Deaf children

The ToMM/SP account has a few potential explanations for the fact that deaf children of hearing parents tend to score significantly lower on direct false-belief tests than both typically developing children and deaf children of deaf parents. However, none of these potential explanations is supported by the evidence.

- The first and most obvious explanation is that deaf children of hearing parents may have poor inhibitory control, in which case they would not be able to resist the true-belief bias. The studies by Woolfe, Want, and Siegal (2002) and P. A. de Villiers and de Villiers (2012) allow us to rule out this possibility. Woolfe and colleagues (2002) found that deaf children of hearing parents did not perform significantly worse than deaf children of deaf parents on a test of executive functioning, while having significantly worse direct false-belief performance. De Villiers and De Villiers (2012) found that the younger group of deaf children in their study (aged 4-6) did not perform significantly worse than a group of typically developing children of the same age, while the older group (aged 6-8) did, although both groups had much poorer direct false-belief performance than typically developing children. Thus, although *some* deaf children of hearing parents appear to have impaired (or delayed) executive functioning, it seems this is not the reason why they tend to have worse direct false-belief performance.
- The second potential explanation is that the poor direct false-belief performance of deaf children of hearing parents is a consequence of their language delay. Two lines of evidence speak against this possibility, however. First, deaf children of hearing parents perform well on the false-photograph task, which makes similar language demands compared to typical direct false-belief tests (Peterson & Siegal, 1999; Woolfe et al., 2002). Second, deaf children of hearing parents perform just as poorly on direct false-belief tests with reduced language demands as they do on typical direct false-belief tests (P. A. de Villiers & de Villiers, 2012; Woolfe et al., 2002).
- The third and last possibility is that, in deaf children of hearing parents, ToMM takes longer to mature and come on-line. It may be suggested, for example, that the maturation of ToMM requires, in order to be triggered, exposure to certain social stimuli. The finding

that deaf children still follow the same developmental progression as other children speaks against this hypothesis, however. Remember that, on the ToMM/SP account, the developmental progression revealed by the ToM scale is to be explained by concurrent developments in inhibitory control. Furthermore, remember that, as noted above, many deaf children of hearing parents do not appear to be delayed in their inhibitory control. Accordingly, one would expect that deaf children of hearing parents, instead of following the same developmental progression as typically developing children, would start passing all the tasks in the ToM scale at the same time, as soon as their ToMM reaches the required level of maturation.

2.1.6 Training

On the ToMM/SP account, false-belief training cannot help by improving young children's belief-reasoning abilities, since these abilities are supposed to be already sufficiently developed to allow them to pass a direct false-belief test. There is two possible alternative explanations, however.

- First, training on a direct false-belief task may help children by giving them opportunity to practice inhibiting the true-belief bias. I know of only one study that has tested this hypothesis (Benson et al., 2013). Benson and colleagues report that, while children did have better inhibitory control in a post-test after undergoing false-belief training, how much children's inhibitory control improved did not predict how much their direct false-belief performance improved, which speaks against the hypothesis under consideration.
- Second, training on a direct false-belief task may have helped children by giving them practice interpreting the linguistic stimuli in the test. To the best of my knowledge, only one study speaks to this hypothesis (C. Hale & Tager-Flusberg, 2003). Hale and Tager-Flusberg report that training on a direct false-belief task did not improve children's performance on a sentential complements task. This does not completely rule out the hypothesis under consideration, however, since it is possible that training improved other aspects of children's language abilities, aspects that were not controlled for.

2.1.7 Mental-state talk

A potential explanation of the finding that children whose parents are more likely to talk about the mind tend to have better direct false-belief performance is that these children also have better inhibitory control. To the best of my knowledge, no study has investigated this possibility. Thus, this hypothesis cannot be completely ruled out. Still, *why* should children's inhibitory skill correlate with their parent's propensity to talk about the mind? There seems to be no obvious connection between the two, which may very well be why the hypothesis has not been tested.

2.1.8 Siblings

On the ToMM/SP account, the fact that children with siblings have better direct false-belief performance than children without siblings has two potential explanations, similar to those taken into account for deaf children: having a sibling may lead children to develop either better inhibitory control or better language ability or both, which then leads to better direct false-belief performance. Since many studies controlled for the potential mediation of these two factors, however, this possibility can be ruled out with some confidence. For example, McAlister and Peterson (2006), who included tests of language ability and executive functioning in their study, found that children's language ability, their executive functioning and whether they had at least one child sibling all made independent contributions to explaining variability in children's performance on a mentalising battery.

2.1.9 Social Competence

On the ToMM/SP account, a possible explanation of the finding that children who have better performance on direct false-belief tests tend to be more socially competent is that inhibitory control may be the mediating factor between these two variables. In other words, if inhibitory control is required to pass both direct false-belief tests and tests of social competence, then children who have better inhibitory control should perform better on both types of test, explaining why the two correlate.

Some of the studies mentioned §1.7 controlled for executive functioning, however, and their results speak against this hypothesis. For example, Peskin and Ardino (2003) included in their investigation two executive functioning tests, both of which are thought to measure children's

inhibitory control. Once age was partialled out, the executive functioning aggregate turned out to be significantly correlated with the mentalising aggregate, but not with the social aggregate. Furthermore, once both age and executive functioning were controlled for, the correlation between the mentalising and social aggregates remained significant ($r=.38$). Similarly, Razza and Blair (2009) included two executive functioning tests in their study, measuring inhibitory control, working memory and cognitive flexibility; they report that the bidirectional relationship found between direct false-belief and PKBS scores was independent of executive functioning.

2.1.10 Brain Maturation

Finally, the finding that children's performance on direct false-belief tests correlates with the level of maturation of their mentalising network is difficult to square with the ToMM/SP account. By two years of age, ToMM should already be sufficiently mature to allow children to pass a direct false-belief test; what two-year-olds would be lacking is a sufficiently developed SP. Consequently, the ToMM/SP account predicts that only the maturation of the neural areas implementing SP should correlate with children's direct false-belief performance. Both Sabbagh et al. (2009) and Wiesmann et al. (2017) included executive functioning tests in their studies, however, and found that maturation in the mentalising network predicted direct false-belief performance over and above executive-functioning. Wiesmann and colleagues also included a test of language ability, again finding that children's language ability did not account for the correlation. In addition, while some of the areas in the mentalising network may be argued to sub-serve inhibitory functions (e.g., see Frith & Frith, 2003) it is implausible to suggest that all of them do.

2.1.11 Summary

While being very influential, the ToMM/SP can explain almost none of the findings discussed in this or previous chapters. Let us consider, see now, whether any of the more recent proposals fare any better (see table 1).

2.2 The Expanded Processing Demands Account

Renée Baillargeon and her colleagues have developed their own processing-load account, which has kept evolving through the years (Baillargeon et al., 2013; Baillargeon et al., 2010;

Baillargeon et al., 2015; Setoh, Scott, & Baillargeon, 2016). To keep the exposition as simple as possible, I will be focusing here on their most recent proposal, advanced by Peipei Setoh, Rose Scott and Renée Baillargeon (2016) and known as the *Expanded Processing-Demands* (EPD for short) account.

Setoh and colleagues argue that direct false-belief tests recruit two additional processes compared to indirect false-belief tests: a response-generation process, which involves interpreting the question and holding the interpretation in mind while selecting a response, and inhibitory control, required to resist the temptation to point to the actual location of the object (reality bias). Notably, Setoh and colleagues maintain that the reality bias, instead of being intrinsic to belief-reasoning *per se* as Leslie and colleagues claim, would be triggered by the test question used in direct false-belief tests. Furthermore, Setoh and colleagues argue that response-generation processes are also likely to load on other executive resources, like working memory, adding to young children's difficulties. Thus, on the EPD account, inhibitory control is an important factor at play but is not the only one; children with good inhibitory control may still fail the test if they lack other executive resources.

2.2.1 Indirect false-belief tests

The claim that the reality bias is triggered by the test question allows Setoh and colleagues to solve one of the most pressing problems of the ToMM/SP account, namely its inability to explain how young children can simultaneously pass indirect false-belief tests and fail direct false-belief tests. Since a test question is present in direct false-belief tests but not in indirect false-belief tests, the reality bias, is triggered in the former case but not in the latter. Thus, consider again the experiment by Clements and Perner (1994), where children were first given an anticipatory-looking unexpected-transfer task and then, after their first looks were recorded, were asked the standard test question. Here, Setoh and colleagues can argue that up until the question is asked, no reality bias is triggered; thus, children can exercise their mentalising abilities unhindered and, having predicted that the character will look for the object where she left it, children will make their anticipatory looks toward that location. However, the test question then triggers the reality bias, which many young children will fail to inhibit, resulting in reality-based responses.

2.2.2 Executive functioning

The fact that inhibitory control is not the only factor at play allows Setoh and colleagues to explain another result that was puzzling for the ToMM/SP account, namely that by Sabbagh and colleagues (2006). This study found that Chinese children, despite outperforming American children on a test of inhibitory control, did not perform any better on a battery of direct false-belief tests. Setoh and colleagues argue that Chinese children may lack some other critical processing resource tapped by response-generation processes, thus balancing out their better inhibitory skill; indeed, they suggest that the processing resource in question may be working memory.

Despite this, other findings concerning the relationship between executive functioning and direct false-belief performance are just as puzzling for the EPD account as they were for the ToMM/SP account.

- First, on the EPD account, one would still expect to see a strong correlation between executive functioning and direct false-belief performance; instead, the meta-analysis by Devine and Hughes (2014) shows that the correlation is only weak.

- Second, while the EPD account can explain one of the studies supporting an emergence account of the correlation (i.e., the study by Sabbagh and colleagues mentioned above) it does not seem able to have an explanation for the other two (Carlson et al., 2015; Benson and colleagues, 2013). For example, the fact that young children tend to answer randomly on the think-know and sources-of-knowledge tests, instead of consistently giving the wrong response as they do in direct false-belief tests, suggests that no reality bias is present in these tests. Thus, on the EPD account, one would not expect to see children's performance on these tests correlate with their performance on tests of inhibitory control; and yet, this exactly what Carlson and colleagues found.

2.2.3 Other findings

In respect to the rest of the evidence we have been discussing in this chapter, the EPD account does not seem to have a better explanation than the ToMM/SP account; this should not be surprising given that the two accounts are, after all, very similar. Thus, for example, since language ability plays a similar role on both accounts (that of allowing children to interpret the linguistic stimuli in the test) the EPD account should not fare any better than the ToMM/SP account in

explaining the relationship between language ability and direct false-belief performance. Or, to consider another example, since on both accounts young children already possess the mentalising abilities that are required for passing direct false-belief tests, the finding that children's performance on these tests correlates with the maturation of their mentalising network should be just as puzzling for one account as it is for the other.

2.2.4 Summary

The EPD account marks an improvement compared to the ToMM/SP account, since it can at least explain why young children pass indirect false-belief tests while failing direct false-belief tests. Nonetheless, it fares no better than the ToMM/SP account with most of the other findings (see table 1).

2.3 The Triple-Load Account

Yet another processing-load account, known as the *triple-load account*, has been defended by Peter Carruthers (2013, 2016). Carruthers maintains that direct false-belief tests recruit the mentalising system more extensively than indirect false-belief tests do. The reason for this is, Carruthers argues, is that indirect false-belief tests only involve one mentalising process, namely that of inferring what the target agent thinks and wants and of predicting her action. In contrast, direct false-belief tests would involve two additional mentalising processes besides the one just mentioned: first, inferring the communicative intentions of the experimenter in order to correctly interpret her utterances; and second, calculating how the experimenter will interpret possible responses to the test question, in order to select a communicative act appropriate to convey the intended answer (This presupposes a neo-Gricean view of pragmatic processing, such as relevance theory; see for example Sperber & Wilson, 1986; Sperber & Wilson, 2002) The fact that direct false-belief tests involve three mentalising processes will result in a triple processing-load on executive functioning compared to indirect false-belief tests, which involve only one mentalising process. Thus, Carruthers suggests that two- and three-year-old children, who pass indirect false-belief tests while failing direct false-belief tests, do so because they possess enough executive resources to carry out one or two mentalising processes, but not three. As children grow older, either mentalising processes

become more efficient (and thus less resource-consuming), or the processing resource available to them increase, or a combination of both.

2.3.1 Experiential factors and deaf children

It is possible that mentalising processes become more efficient with practice. If this is correct, then experiential factors like receiving training on direct mentalising tests, having parents that are more likely to talk about the mind, and having a sibling would help young children not because they allow them to learn that beliefs can be false, but because they increase the efficiency of their mentalising processes through practice, thus reducing the processing load children must sustain to pass direct false-belief tests. It would then make sense that deaf children of hearing parents, who, during their first years of life, are all but deprived of opportunities to engage in conversations about the mind, should pass direct mentalising tests later than typically developing children.

2.3.2 Brain Maturation

It is possible that the processes of brain maturation tracked by estimated regional current-density (Sabbagh et al., 2009) and fractional anisotropy (Wiesmann et al., 2017) lead to, or correlate with, an increase in the efficiency of mentalising processes. If this hypothesis is on the right track, then the triple-load account should be able to explain why children who exhibit a higher level of maturation in their mentalising network tend to have better direct false-belief performance, since these children will require less executive resources to pass a direct false-belief test.

2.3.3 Executive functioning

Like the EPD account, the triple-load account maintains that children's inhibitory control is only one of the factors determining whether they will pass or fail a direct false-belief test. Thus, like the EPD account, the triple-load account can explain the results of the study by Sabbagh and colleagues (2006), who found that Chinese children outperformed their American peers on a battery of executive functioning tasks but not on a battery of direct false-belief test. In fact, the triple-load account is consistent with Sabbagh and colleagues' own suggestion that Chinese children's better inhibitory control may be balanced out by the fact that Chinese children do not have siblings, for having a sibling, as mentioned above, is likely to result in more mentalising practice.

The fact that the efficiency of mentalising processes also plays a role in determining whether children pass or fail a direct false-belief test may allow the triple-load account to explain a finding that was problematic for the EPD account, namely that the correlation between executive functioning and direct false-belief performance is relatively weak (Devine and Hughes, 2014). For example, some children may pass despite their relatively poor executive skill because, having had more opportunities to practice their mentalising skills, they possess more efficient mentalising processes.

Even so, the triple-load still seems to leave some of the findings in this area unexplained. Consider, for example the study by Carlson and colleagues (2015), showing that children's performance on the think-know and sources-of-knowledge test also correlate with their inhibitory control, and did so as strongly as their direct false-belief performance. (This finding is still problematic because Carruthers, like Leslie, maintains that inhibitory control would be recruited to ascribe a false belief; thus, it is surprising that tasks that do not require ascribing false beliefs correlate with inhibitory control as strongly as direct false-belief tests.)

2.3.4 Language Ability

While inferring speakers meaning may seem to fit the intuitive definition of “language ability”, tests of language ability typically only measure children's understanding of the vocabulary and grammar of their language (Milligan et al., 2007). Still, children with better language ability (in this restricted sense) may be in a better position to infer what the experimenter wants to know. This is because the literal meaning of a sentence is clearly an important datum in working out what the speaker meant to convey by uttering it. (Thus, for example, to recognise that by uttering “I am stuffed!” in response “do you want another slice of cake?” I implied “No” you must understand what “stuffed” means; if you think it means *empty*, then you might conclude that I implied “Yes”). Thus, we can expect that not being able to decode the literal meaning of the test question will make it more difficult for children to infer what the experimenter wants to know, possibly resulting in increased processing demands.

This type of explanation for the relationship between language ability and direct false-belief performance, while not being identical to the one considered in §2.1, where we discussed the ToMM/SP account, seems to suffer from similar problems. Thus, for example, it seems to clash

with the finding that different false-belief tasks, using test questions of varying degrees of complexity, all correlate with language ability to the same extent (Milligan et al., 2007).

The triple-load account can also provide a different account of the relationship between language ability and direct false-belief performance, however. Children's language abilities may affect the extent to which they benefit from exposure to mental-state talk and training. As mentioned in §2.3.1, by participating in conversations about the mind with their parents, children can get more practice at mentalising, and this can be expected to result in increased efficiency of their mentalising processes. Patently, however, this is only true so long as children understand what their parents are saying. Since language ability can mediate the beneficial effect of experiential factors independently of the language demands made by direct false-belief tests, this type of explanation is compatible with the finding that language correlates to similar extents with false-belief tasks that have different language demands (Milligan et al, 2007). Notice, however, that this type of explanation has a different problem: it cannot explain why receiving training on sentential complements improves children's direct false-belief performance.

To sum up, then, the triple-load offers two explanations: either language ability helps children in inferring what the experimenter wants to know, or it helps them by mediating the effect of practice. While each can explain some of the evidence, neither can explain all of it.

2.3.5 Social Competence

Like direct false-belief tests, many tests of social competence may involve predicting what other people will do while also processing linguistic stimuli. Accordingly, one can expect that passing these tests will also require carrying out several mentalising processes in a short amount of time, imposing a multiple processing load on executive functioning. It makes sense, then, that children who fail direct false-belief tests should also fail these social competence tests. On the other hand, some social-competence tests may be more frugal, involving one or two mentalising processes at most. The triple-load account predicts that children's direct false-belief performance should not correlate with this latter type of test.

The evidence seems to speak against this hypothesis. Playing hide-and-seek with an experimenter, for example, seems to only involve one mentalising process: anticipating where the other will look, or figuring out where the other may have hidden. In particular, since the game is supposed

to be played in silence, there will be no linguistic stimuli to process. The fact that children's direct false-belief performance correlates with their ability to play hide-and-seek should thus be puzzling for the triple-load account.

2.3.6 Developmental Progression

The triple-load account seems to have trouble explaining why children follow the developmental progression revealed by the ToM Scale. Notice that all the tests in the ToM Scale seem to involve three mentalising processes, no differently from direct false-belief tests. This is because they all require ascribing mental states to the target agent while interpreting the experimenter's question and selecting a response. Thus, the hypothesis that children pass the diverse-desire, diverse-belief and knowledge-access tests before the false-belief test because they involve fewer mentalising processes is not tenable.³

2.3.7 Summary

The triple-load account can explain several findings that were puzzling for the ToMM/SP and EPD accounts. These include those suggesting that experiential factors (exposure to mental-state talk, training etc.) can affect children's direct false-belief performance. Nonetheless, other findings remain puzzling, including most notably the correlation between direct false-belief performance and social competence, and the developmental progression revealed by the ToM Scale (see table 1).

2.4 The Pragmatic Perspectives Account

In contrast with all the early-emergentists discussed so far, Katharina Helming, Brent Strickland and Pierre Jacob (2014, 2016) defend a pragmatic type of account. (The authors do not name their account; let us call it the *Pragmatic Perspectives Account*). The key idea behind pragmatic accounts

³ It is true, of course, that the fact that all these direct mentalising tasks involve three mentalising processes does not entail that these processes must be equally taxing in each case. For example, participants' own belief seems to be less salient in diverse-belief test compared to the false-belief test, and may thus be easier to inhibit, resulting in a smaller load being placed on inhibitory control. The suggestion that the tasks in the ToM scale differ in their demands on executive functioning has already been discussed in section §2.1, however, and was found to face several problems; nothing Carruthers says suggests a possible solution to those problems.

is that young children answer incorrectly because they misunderstand what the experimenter is asking. According to Helming and colleagues, the reason this happens is that young children cannot take a third-person perspective on the target agent's action while maintaining a second-person perspective on the experimenter's question. Taking a third-person perspective on an action means observing it unfold in a detached way, without being invested in its outcome; taking a second-person perspective on an action, in contrast, means having a stake in its outcome and contributing to it, by either helping or hindering the agent. Young children, Helming and colleagues argue, are likely to find it difficult to assume and maintain both types of perspective at the same time. This, in turn, would make two possible misinterpretations of the test question very tempting for young children. The cooperative misinterpretation takes the test question to be about what the target agent should do as opposed to what she will do; the referential misinterpretation, in contrast, takes the test question to be about where the object actually is. Importantly, in both cases the correct response is the reality-based one.

2.4.1 Experiential factors

By participating in training sessions targeted at a direct false-belief tests, children may practice taking different types of perspectives at the same time, which will then improve their direct false-belief performance. That fact that training helps is thus consistent with the pragmatic perspectives account.

Participating in conversations about the mind may also involve taking both a third-person and a second-person perspective, at least in some cases. Accordingly, it is possible that children whose parents are more likely to discuss mental states will get more practice at taking different types of perspectives at the same time, which will then help them pass direct false-belief tests. The pragmatic perspectives account may thus be able to explain why these children tend to have better direct false-belief performance. Finally, if having a sibling leads to more mental-state talk within the family, then the pragmatic perspectives account can also explain why children who have a sibling should have better direct false-belief performance.

2.4.2 Deaf children

Since the pragmatic perspectives account can explain why participating in conversations about the mind helps, it can also explain why deaf children of hearing parents, who will have trouble participating in such conversations over their first years of life, tend to have very poor direct false-belief performance. In addition, given that deaf children are less exposed to linguistic stimuli, it is not implausible to suggest that they should also have relatively poor pragmatic understanding.

2.4.3 Brain Maturation

It is unclear how the pragmatic perspectives account could explain the finding that children with better direct false-belief performance tend to have a more mature mentalising network. This is because Helming and colleagues do not tell us why young children have trouble taking a second-person perspective on the target agent's action (e.g. Sally's looking for the marble) while maintaining a third-person perspective on the experimenter's question. Perhaps, it may be suggested, as children's mentalising abilities improve due to the maturation of their mentalising areas, maintaining those two perspectives should become easier.

2.4.4 Executive functioning

Since the pragmatic-perspectives account does not assign any role to executive functioning, it does not predict the correlation between executive functioning and direct false-belief performance. This problem may be fixed by maintaining that, in order to select the correct interpretation of the test question, children must first inhibit the cooperative and referential interpretations, which may initially be more salient. In other words, having good inhibitory control might facilitate pragmatic understanding. With this modification in place, the pragmatic-perspectives account may also be able to explain why the correlation with executive functioning is relatively weak. After all, inhibitory control would only help those children which entertain, as a possible interpretation, the one that happens to be correct; inhibiting the wrong interpretation is not enough; one must also select the correct interpretation. Perhaps, for most young children, the correct interpretation does not even come to mind as a potential candidate, in which case having good inhibitory control would be of little use. Notably, however, even this modified version of the account cannot explain the evidence for emergence accounts.

2.4.5 Language Ability

The pragmatic development account can provide essentially the same type of explanation(s) provided by the triple-load account: having better language ability may facilitate pragmatic interpretation, while also allowing children to benefit more from exposure to mental-state talk and training.

2.4.6 Developmental Progression

All the direct mentalising tests in the ToM scale seem to require taking both a third-person and a second-person perspective, at the same time. As a result, the pragmatic perspectives account predicts a different developmental progression than that which is typically observed in studies using the ToM Scale. For example, on the pragmatic perspectives account, the diverse-belief test should be as difficult as the false-belief test. Not only both tests require children to take a second-person perspective on the experimenter's communicative action and a third-person perspective on the agent's endeavour to find the object, but the referential and cooperative misinterpretation are possible in both cases and will lead to the same result – a reality-based response. The results obtained with the ToM Scale are thus difficult to square with the pragmatic perspectives account.

2.4.7 Social Competence

According to the pragmatic perspectives account, the children who have better direct false-belief performance are those that are better at simultaneously assuming both a third-person and a second-person perspective. Accordingly, the finding that children who have better direct false-belief performance tend to also score higher on many tests of social ability may be explained if these other tests also require children to assume inconsistent perspectives. This does not seem to be the case, however. Consider for example the ability to play hide-and-seek task without betraying one's hiding location, which has been found to correlate with direct false-belief performance (Ding et al., 2015; Peskin & Ardino, 2003). In playing hide-and-seek, children only need to take one perspective, namely a second-person perspective on the experimenter's action, who is looking for them; there is no need to *also* take a third-person perspective on anything else. Thus, on the pragmatic perspectives account, there seems to be no reason for children's performance on this test to correlate with their performance on direct false-belief tests.

2.4.8 Summary

The findings that were problematic for the triple-load account are just as problematic for the pragmatic-perspectives account. Furthermore, the triple-load account arguably provides a better account of many of the findings, including, for example, why children with better direct false-belief performance tend to have a more mature mentalising network (see table 1).

2.5 The Pragmatic Development Account

A second pragmatic account, called the *pragmatic development account*, is defended by Evan Westra (2016a). Westra maintains that the complex pragmatics of belief discourse make it difficult for young children to tell when adults are talking about beliefs; he discusses two types of considerations that support this claim. First, he argues that adults frequently leave beliefs implicit in their explanations and descriptions of behaviour. For example, suppose I showed you a picture of a person cracking a bottle of beer open, and asked you to explain this piece of behaviour; “she wanted to have a drink” would be a perfectly good answer. Of course, this explanation assumes that the person in the picture thought there was beer inside the bottle; but mentioning this would be odd from a pragmatic point of view. As this example illustrates, beliefs are often left implicit in conversation, and what beliefs the speaker takes the target agent to have is left for the audience to figure out. It is possible that young children simply fail to recognise that what someone believes is often taken for granted, which may in turn lead them to underestimate the frequency with which adults talk about beliefs. Second, Westra points out that the verb “think” is often used in an indirect manner, to convey not what someone believes but rather what is the case (as when one responds with “I think they are over there” to “I can't find the keys!”). Young children, Westra argues, may have a hard time telling literal, attributive uses apart from indirect ones, and may initially take every use of “think” to be indirect; this would in turn reinforce the (misguided) impression that beliefs are rarely talked about.

This type of account has a straightforward explanation for direct unexpected-contents tasks, where children are asked what someone who has not looked inside a container (e.g., a Smarties tube with crayons inside) will think is inside it. Since the question uses the verb “think”, if children take the indirect reading they will think the experimenter is asking what is inside the box, not what

the other person *thinks* is inside it. This would explain why young children tend to provide reality-based answers (e.g., “crayons!”).

This explanation does not extend to direct unexpected-transfer tasks, however, where children are typically asked where the target agent will look for an object, as opposed to where she thinks the object is. Here, Westra presents a slightly different explanation: he argues that, since children think that people rarely talk about beliefs, the possibility that the experimenter may want them to exhibit their awareness of the agent's belief will not be very salient to them. This, in turn, will lead them to prefer a different interpretation of the question; they might, for example, take it to concern what the agent should do or where the object is (similar to what Helming and colleagues propose).

2.5.1 Experiential factors and deaf children

One of the main motivations behind the pragmatic-development account is that of providing a better explanation of the evidence concerning training, mental-state talk, siblings and deaf children, which is problematic for the ToMM/SP and EPD accounts (as argued in §2.1-2). According to Westra, being exposed to conversations about the mind improves children's direct false-belief performance not because it allows them to learn that beliefs can be false, as late-emergentists suggest, but because it allows them to gain a better grasp of the types of situation where adults are likely to be talking about beliefs. This, in turn, should make children better able to interpret the test question in direct false-belief tests, thus improving their chances of answering correctly. Similarly, false-belief training that includes feedback improves children's direct false-belief performance not because it allows them to learn that beliefs can be false, as suggested by late-emergentists, but because it helps children recognise what the experimenter wants to know. In a nutshell, these factors help because they allow children to better understand the pragmatics of belief discourse (Westra, 2016a, p. 248). In light of this, we can then explain why deaf children of hearing parents tend to have poor direct false-belief performance: since they have trouble communicating during their first years of life, deaf children of hearing parents will be less exposed to conversations about the mind and will thus have a much harder time interpreting the question used in direct false-belief tests. Finally, having two or more children may induce parents to talk about the mind more often,

for example in settling disputes, which will then help their children acquire facility with the pragmatics of belief discourse (Westra & Carruthers, 2017, p. 171).

2.5.2 Developmental progression

Westra and Carruthers (2017) argue that the pragmatic-development account can explain the developmental progression revealed in the ToM scale. The reason the diverse-desire test is passed before the others, they argue, is that it does not involve beliefs, and thus it does not present the same difficulties of pragmatic interpretation as direct false-belief tests. After all, desire-talk is not nearly as infrequent as belief-talk, and while “want” can also be used in an indirect way, this indirect use is much less common than with “think”. We can thus expect that it should be much easier for children to recognise that the experimenter is asking about the agent's behaviour because she wants the child to show that she is aware of the agent's desires (Westra & Carruthers, 2017, p. 172).

The second step in the progression is the diverse-belief test. In this task, children are told that a person (e.g. Linda) is looking for something (e.g., her cat), which may be in either of two locations (the bushes or the garage) and are then asked to guess where the thing is. Whatever children say, the experimenter will then say that Linda thinks the cat is in the other location and will ask them where she will look for it. Here, children are likely to assume that “think” (in “Linda *thinks* the cat is in the other location”) is being used in an indirect manner, in which case they will take the utterance to state the actual location of the cat. If children think the cat is in the bushes, they will of course say that Linda will look there, regardless of whether they recognise that the experimenter is asking that question because she wants children to show that they are aware of Linda's beliefs. The reason this test is passed later than the diverse-desire test is that some children may go on believing their own guess about the location of the object, ignoring the experimenter's utterance about what Linda believes. Still, children can pass the test even if they do not correctly interpret the question, and this makes the diverse-belief test significantly easier than the false-belief test (Westra and Carruthers, 2017: 171-2).

The third step in the progression, which is the last before the explicit false-belief test, is the knowledge-access test. In this task, children are shown the contents of a container (e.g., a toy dog in a drawer) and asked whether someone who has not looked inside it (e.g., “Polly”) knows what's in it. Here Westra and Carruthers (2017: 172-3) present two complementary explanations. The first

is that since Polly does not have the goal of finding the toy dog, a cooperative interpretation of the question (i.e., “the experimenter wants me to help Polly find the object”) is ruled out; this should make the knowledge-access test easier than the false-belief test. At the same time, the factual interpretation (i.e., “the experimenter wants me to show that I am aware of the location of the object”) is still possible, and the correct interpretation should still seem unlikely to young children since it concerns what the agent thinks. Thus, the knowledge-access test should be more difficult than the diverse-belief test. The second explanation, instead, suggests that children's difficulties may be partly due to an experimental artefact. This is because young children are biased to answer all yes/no questions positively (Okanda & Itakura, 2008), and in this case the correct answer is negative (since Polly does *not* know what is inside the drawer.)

If this type of account is on the right track, how do we explain the fact that Iranian and Chinese children, differently from their Western peers, pass the knowledge-access test before the diverse-belief test? Westra and Carruthers point out that resisting the temptation to provide a positive answer is likely to require inhibitory control (Moriguchi, Okanda, & Itakura, 2008) and that Chinese children have been found to have better inhibitory control than their American peers (Sabbagh et al., 2006). Furthermore, they argue, Wellman's own proposal can be given a pragmatic twist: since collectivist societies, like Iran and China, put more emphasis on common knowledge than they put on diversity of opinion, children in these societies will be more likely to recognise that the experimenter is interested in Polly's lack of knowledge as opposed to Linda's diverse belief (Westra and Carruthers, 2017).

2.5.3 Social Competence

While the pragmatic development account seems to fare well with the evidence discussed so far, it does not seem to be able to explain why children who have better direct false-belief performance should be more socially competent. This is because the ability to recognise when people are talking about beliefs appears to be irrelevant to many tests of social competence. Consider for example the study by Peskin and Ardino (2003), which found that children's performance on a direct false-belief test correlated with their performance on a hide-and-seek test. Children with better direct false-belief performance were found to be less likely to betray their location while playing the hider (e.g., less likely to make noise while hidden, less likely to hide while the

experimenter was still looking, and so on). Now, in a game where one is not supposed to *speak*, what use could the ability to navigate the pragmatics of belief talk possibly be? Or consider, as a second example, the study by Peterson et al. (2017), which found that three- to twelve-year-olds who scored higher on a battery of direct false-belief tests were more likely to produce high level arguments in trying to convince a puppet (e.g., Matty) to do something it was not inclined to do (e.g., eating broccoli). Again, what use could being able to tell when people are talking about beliefs be, in trying to convince Matty to eat his broccoli?

2.5.4 Brain Maturation

In theory, at least, it is possible that having better mentalising skills may help children recognise when adults are talking about beliefs. If this assumption is correct, then the pragmatic perspectives account may be able to explain why children with a more mature mentalising network tend to have better direct false-belief performance.

2.5.5 Other Findings

On the pragmatic development account, language ability and executive functioning play essentially the same role as on the pragmatic perspectives account (discussed in §2.4.4–5).

2.5.6 Summary

The pragmatic development account is the only proposal, among the one we have considered so far, that provides an explanation of why children follow the developmental progression revealed in the ToM Scale. In this respect, the pragmatic development account certainly marks an improvement on the other proposals. Nonetheless, the account still cannot explain why children with better direct false-belief performance tend to be more socially competent. Furthermore, like all other

	ToMM/SP	EPD	Triple-Load	PP	PD
Indirect false-belief tests	✗	✓	✓	✓	✓
Developmental progression	✗	✗	✗	✗	✓
Executive functioning	✗	✗	✗	✗	✗
Language ability	✗	✗	✗	✗	✗
Experiential factors	✗	✗	✓	✓	✓
Deaf children	✗	✗	✓	✓	✓
Social competence	✗	✗	✗	✗	✗
Brain maturation	✗	✗	✓	✓	✓

Table 1. Summary of Chapter III. PP=Pragmatic Perspectives Account; PD= Pragmatic Development Account.

accounts considered so far, the pragmatic development account leaves some of the evidence concerning executive functioning and language ability unexplained. Finally, its explanation of why these children tend to have a more mature mentalising network seems less plausible than that provided by the triple-load account (see table 1 for a visual summary).

3. Conclusions

In this chapter, we have considered whether early-emergentists can explain away the evidence that is often claimed to support late-emergence accounts. We have considered five proposals: the ToMM/SP account (§2.1), the expanded processing demands account (§2.2), the triple-load account (§2.3), the pragmatic perspectives account (§2.4) and the pragmatic development account (§2.5). I argued that, while some of these proposals fare better than others, all leave important pieces of evidence unaccounted for. In the next chapter, I will try to fix this problem by putting forward my own proposal, the processing-time account.

IV

THE PROCESSING-TIME ACCOUNT

In this chapter, I am going to put forward my own performance account, called the processing-time account. This builds on, and incorporates, many of the ideas other early-emergentists have put forward and defended but combines them in a novel way. As we are about to see, this allows the processing-time account to explain all the evidence that, in the last chapter, was found to give trouble to other performance accounts.

The processing-time account will be introduced in §1. Then, in §2, it will be argued that the processing-time account can explain the evidence that is typically claimed to support other performance accounts. Finally, in §3, it will be argued that the processing-time account can also explain all the evidence discussed in Chapter III.

1. The Account

1.1 Processing Time

The processing-time account moves from the observation that children have a limited amount of time to carry out the cognitive processes required to pass a direct false-belief test. These processes may include (i) figuring out what the experimenter is asking, (ii) thinking about the answer and (iii) planning and executing the response. The suggestion at hand is that, by and large, children who fail direct false-belief tests do so because they have a hard time carrying out all these processes in the time available.

If this suggestion is on the right track, then mainly two factors are likely to affect children's direct false-belief performance. The first relates to how long it takes them to carry out the processing required to answer correctly. (It will be handy to have a term that refers to this variable; call it *required processing-time*, or *RPT* for short). The lower children's RPT is, the more likely will they be to complete their processing in the time available. The second variable is how much time children have at their disposal to carry out their processing (call it *time available for processing* or *TAP* for short). In some cases, the time constraints may be imposed by the experimenter who is administering the test. Thus, some experimenters, underestimating the time it takes children to complete their processing, may move on or re-prompt too soon, not leaving their participants enough time to think. More often, however, the time constraints will be self-imposed. Even when experimenters are willing to give their participants all the time they need, many children will still rush to answer, cutting their processing short. Importantly, there are likely to be both individual differences and age-related changes in the time children are either willing or able to wait before responding. Other things being equal, children who wait longer before responding will have more time to complete their processing, which will give them a better chance of answering correctly. With these two variables in place, the processing-time can be schematically represented with the following formulae:

$$\text{RPT} > \text{TAP} \rightarrow \text{Failure}$$

$$\text{RPT} \leq \text{TAP} \rightarrow \text{Success}$$

When applied to development, the processing-time account maintains that the reason older children tend to have better direct false-belief performance than younger ones is that older children tend to have either lower RPTs, or higher TAPs, or both.

Now, we know that three-year-olds tend to give reality-based answers in direct false-belief tests; thus, for example, when asked where Sally will look for the marble, they will tend to point to the box where the marble is, and when asked what someone who has not looked in the Smarties tube will think is inside it, they will tend to answer crayons. On the processing-time account, this is to be explained by appealing to the tendency people have to substitute demanding tasks for ones that are related but much easier (known as attribute-substitution). Several authors have argued that this is a pervasive phenomenon, potentially underlying many of the cognitive biases observed in reasoning and decision-making (e.g., Daniel Kahneman, 2011; D. Kahneman & Frederick, 2002; Shah & Oppenheimer, 2008). It is thus reasonable to assume that, when children realise that the task at hand requires considerable time and mental effort, they will feel the temptation to replace it with an easier one. What attribute children will substitute for the original one depends on how far they have gone with their processing, but reality-based answers, being both readily accessible and highly salient, are likely candidates. Suppose, for example, that a child is still trying to figure out what the character in the story will do when she realises it is taking too long. This task could be swapped with a related but easier one, namely that of figuring out where the object the character is looking for is. This second problem is easier because it can be solved simply by recalling what happened in the story, something our child would have to do anyway to solve the original task. Thus, the information about the current location of the object will be both readily accessible in her mind and highly related to the question she is trying to answer, making it the perfect candidate for attribute substitution. Alternatively, suppose that the child in our example is still processing the test question as she notices that the process is taking too long; in this case, she may be tempted to jump to the conclusion that, since the question contains the words “where” and “marble”, it must be about the location of the marble, without taking time to check whether this interpretation is consistent with the overall syntactic structure of the sentence or the meaning of the other words in it.

1.2 Factors Affecting RTP

1.2.1 Processing Speed

One factor that is likely to lead to a reduction in required processing-time is the well-documented, age-related increase in children's processing-speed. Reaction times across a variety of perceptual, motor and cognitive tasks are known to decrease during childhood, up until young adulthood, and to start increasing again slowly thereafter, thus following a u-shaped trajectory (Cerella & Hale, 1994; S. Hale & Jansen, 1994; Kail, 1991; Miller & Vernon, 1997). Consider for example the study by Sandra Hale (1990), which used four reaction-time tasks (a choice reaction-time task, a letter-matching task, an abstract-matching task and a mental-rotation task.) Hale found that, while there were no significant differences in error rates across different age groups, in each task ten-year-olds were slower than twelve-year-olds, who were themselves slower than fifteen-year-olds. This type of finding is typically taken to show that processing speed increases globally with age, leading to a reduction in reaction times across the board (Cerella & Hale, 1994; Kail, 1991). Based on this, Robert Kail and Timothy Salthouse (1994) argue that a domain-general processing speed parameter should be recognised as an important architectural component of the mind, comparable to the clock speed of the central processing unit (CPU) of a computer. Now, if the increase in processes speed does indeed apply to all cognitive processes across the board, then there can be little doubt that it will also apply to the cognitive processes recruited in direct false-belief tests.

Since the relation between children's processing speed and their direct false-belief performance has not been investigated, there is no direct evidence that children with higher processing speed have better direct false-belief performance. Nonetheless, two sets of findings support this hypothesis.

- The study by Tamsin German and Jessica Hehman (2006) investigated the relationships between mentalising, processing speed, working memory and inhibitory control in young adults (between eight-teen and twenty-six years of age) and elderly adults (between sixty-two and and ninety years of age). German and Hehman found that processing speed (measured using the digit-symbol task) was highly correlated with direct false-belief performance, explaining about 35% of the variance in accuracy of responses. This suggests

that the decrease in direct false-belief performance observed in elderly adults is due at least in part to their lower processing speed.

- The increase in processing-speed observed during childhood may be a result of the maturational processes affecting the brain. Indeed, many recent studies report that reaction times correlate with parameters like myelination, axonal growth, white matter volume, fibre density and white matter integrity, all factors that are likely to affect the speed at which neural impulses propagate in the brain (Chopra et al., 2018; Kerchner et al., 2012; Penke et al., 2010; Turken et al., 2008). Now remember that Wiesmann et al. (2017) found that direct false-belief performance correlated with fractional anisotropy in the white matter around the right TPJ and other key mentalising areas (this study was discussed in Chapter III, §1.7). During development, increases in the fractional anisotropy of white matter are known to depend mainly on maturational processes of axonal growth and fibre myelination, both of which are likely to enable faster propagation of electrical impulses; and indeed, several studies have found correlations between fractional anisotropy and reaction times (Borghesani et al., 2013; Burgmans et al., 2011; Karbasforoushan, Duffy, Blackford, & Woodward, 2015; Kerchner et al., 2012; Kochunov et al., 2010; Kochunov et al., 2016; Turken et al., 2008). In light of this, it is reasonable to interpret the study by Wiesmann and colleagues as showing that children who have higher processing speed are more likely to pass direct false-belief tests.

1.2.2 Efficiency of processing

Although processing speed is one of the main factors affecting processing time, it is not the only one. The time it takes one to get from A to B can be reduced either by increasing average speed or by taking a shorter route. Applying this analogy to the psychological domain, processing time can be reduced either by increasing processing speed (=the speed at which basic computational operations can be carried out), or by adopting more efficient algorithms (=algorithms that require a smaller number of computational steps to carry out the same task). Thus, so far as processing time is concerned, efficiency of processing is just as important of processing speed.

Consider, for example, the effect of practice: it has long been known that practising a task reduces one's reaction times on that task (e.g., Newell & Rosenbloom, 1981). While in theory this

could be due to an increase in processing speed, it is more likely that practice leads to the adoption of more efficient computational strategies (Kail, 1991). For example, practice may lead to the automatization of the processes recruited in the task, reducing the need for executive direction and supervision; it may also lead to the formation of task-specific mental shortcuts and heuristics that reduce the amount of processing required. Either way, the processing-time account predicts that children who have had more opportunities to practice the processes recruited in direct false-belief tasks should have a better chance of passing, on account of having a shorter required processing-time. The evidence largely corroborates this prediction. Consider, for example, the finding that children whose parents are more likely to talk about the mind tend to have better direct false-belief performance. Every time a parent discusses what someone thinks, wants or feels, she is encouraging her child to follow suit, essentially engaging in a tutored exercise of mentalising. Furthermore, given children's propensity for imitation, we can expect that children who are more exposed to mentalistic talk will be more disposed to think about the mind all around, even in the absence of parental prompt. Thus, more exposure to mental-state talk is likely to lead children to spend more time mentalising, which will in turn increase the efficiency of their mentalising processes, leading to a reduction in required processing-time. The processing-time account thus predicts that children who are more exposed to mental-state talk may have better direct false-belief performance, just like the evidence suggests.

Setting practice aside, another factor that is likely to affect efficiency of processing is working memory capacity (i.e., the amount of information that can be held in mind at any given moment.) Note that the information that is held in working memory is readily available for further processing, while the information that is stored in long-term memory would first need to be retrieved. Thus, keeping information active in working memory is a way to avoid the processing costs associated with retrieving information from long-term memory. At the same time, working memory capacity is known to be extremely limited even in adults, and to develop rapidly during childhood. It is thus possible that young children will be unable to hold all the information they need in working memory, in which case some of that information will have to be either stored in long-term memory or discarded, to be later retrieved or re-calculated, thus leading to increased processing times. For example, suppose that a child realises that Sally thinks the marble is in the basket as she listens to the story. Then, the experimenter asks the question, "Where will Sally look for her marble?", which

the child will have to process. Now, ideally, the child should be able to hold all the potentially relevant information in working memory: this will include where Sally put marble, where the marble is now, where Sally thinks the marble is, what the experimenter wants to know and so on. Let us suppose, however, that the child has very limited working memory capacity, and as such, to process the question, she is forced to either store or discard some of the information she is currently holding in working memory. Thus, for example, the information that the marble is in the box may be stored in long-term memory, while the information that Sally thinks the marble is in the basket may be discarded. Once she has figured out that the experimenter wants to know where Sally will look for the marble, the information that was stored will have to be retrieved and that that which was discarded will have to be recalculated, leading to a considerable increase in processing times. Given the close connection between working memory and efficiency of processing, the processing-time account predicts that children with better working memory capacity should have better direct false-belief performance; and indeed, several studies corroborate this prediction (Carlson et al., 2002; Davis & Pratt, 1995; Duh et al., 2016; Gordon & Olson, 1998; Keenan et al., 1998).

It is worth noting that low processing-speed can also affect children's ability to hold information in working memory, leading to re-calculation. This is because the contents of working memory tend to decay with time; thus, low processing speed increases the likelihood that, by the time a process has been completed, the outputs of earlier processes will have been lost (Salthouse, 1996). To see this, consider a case similar to the one discussed above, where the child has sufficient working memory capacity but low processing speed. Thus, even though the child can keep the information about Sally's belief in working memory as she processes the question, by the time her processing is complete that information may have been lost due to decay, leading to the same result: recalculation. This type of phenomenon may explain why measures of processing speed and working memory are highly associated (Fry & Hale, 1996; Kail, 2007).

1.3 Factors Affecting TAP

1.3.1 Inhibitory Control

Let us focus, now, on the factors that are likely to affect the second main variable in the processing-time account, namely the time available for processing (TAP). On the assumption that

the experimenters leave children sufficient time to think (an assumption that may not always be satisfied) any time constraints in the test will be self-imposed. In particular, children who are more impulsive (=who possess less inhibitory control) are likely to respond more quickly. This is because we are operating under the assumption that if a solution to the problem children are considering (e.g. what does Sally think? What does the experimenter want to know?) does not immediately come to mind, they will be tempted to substitute the current task with an easier one. Resisting this temptation and delaying the response will thus require inhibitory control. Thus, children with poor inhibitory control may immediately give in and go with whatever response is considered first, while children with good inhibitory control may be able to wait until a better response is entertained. Furthermore, since inhibitory control develops rapidly during the preschool years, we can expect older children to be better able to resist than younger children, which may then result in them having better direct false-belief performance. The evidence supports this prediction: as discussed in Chapter III, §1.2, children with better inhibitory control tend to score higher on direct false-belief tests (Carlson et al., 2015; Carlson, Mandell, & Williams, 2004; Carlson & Moses, 2001; Carlson et al., 2002; Carlson, Moses, & Claxton, 2004; Sabbagh et al., 2006).

1.3.2 Motivational factors

While inhibitory control is certainly an important factor in determining how long children wait before responding, it is not the only one. How motivated children are to answer correctly is likely to be another important factor. Some children, for example, may have the ability to delay their response but lack sufficient motivation to do so. After all, from the child's point of view, not much hangs on answering correctly (there is no promised reward.) Notably, to the best of my knowledge, no study has investigated whether introducing a reward for correct answers (e.g. a piece of candy or a sticker) could lead to improvements in children's direct false-belief performance. Thus, at the moment, this prediction of the processing-time account is largely untested.

2. Evidence for Performance Accounts

Several studies show that relatively simple changes to the procedure in direct false-belief tests can significantly improve young children's performance. These findings provide crucial evidence in support of performance accounts. In this section, I discuss the main findings and argue that the

processing-time account can explain how each of these task manipulations helps young children pass direct false-belief tests. This will also allow us to further flesh out the processing-time account.

2.1 Executive Functioning

Processing-load accounts maintain that direct false-belief tests place a heavier load on executive functioning compared to indirect false-belief tests, a load that young children cannot sustain, and which thus causes them to fail. The processing-time account agrees that poor executive skill is part of the reason young children fail. Thus, on both types of account, task manipulations that reduce the load on executive functioning should improve children's chances of passing a direct false-belief test. A few studies suggest that this is indeed the case; these include the undisclosed-location tasks, the Duplo task, and false-beliefs tasks using temporal marking. The latter will be discussed in section §2.3, while the first two will be discussed presently.

2.1.1 The Undisclosed-Location Task

The undisclosed-location task is a variation of the unexpected-transfer task, where the target object is either taken to an undisclosed location, or disappears, or was never present to begin with. Thus, for example, Anne might take the marble away with her instead of moving it from the basket to the box. Alternatively, children may be shown that both the box and the basket are empty but told that Sally thinks her marble is in the basket. Three-year-olds have been shown to perform better on direct undisclosed-location tasks compared to standard unexpected-transfer tasks (Bartsch, 1996; Kikuno, Mitchell, & Ziegler, 2007; Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983). Karen Bartsch (1996), for example, reports that the three-year-olds in her study, who were below chance on a typical unexpected-transfer task, were at-chance on a undisclosed-location task, with the difference between the two conditions being significant.

On processing-load accounts, the fact that the object is in neither of the two locations should either reduce or eliminate the load on inhibitory control: even if the child wanted to ascribe to Sally a true belief about the location of the marble, she could not do that since she herself does not know where the marble is. Thus, in undisclosed-location tasks, the strength of the true-belief bias should be significantly reduced, making it considerably easier to override, and allowing a higher portion of three-year-olds to pass.

The processing-time account suggests a different interpretation, which is also consistent with the data. Remember that on the processing-time account, inhibitory control is required to delay the response and resist the temptation to substitute an easier task for the current one (as explained in §1.4). Now, in typical unexpected-transfer tasks, one likely candidate for substitution is that of figuring out where the object the agent is looking for is. In undisclosed-location tasks, however, pointing to where the marble is, is harder than pointing to where Sally will look, since children do not know where the marble is. Other possible substitute tasks, like pointing to where Sally should look or to where she will end up looking, are also ruled out, for the same reason. Since all the main candidates for substitution are removed, either of two things may happen, both of which may result in improved performance. (1) The first possibility is that since there are no easy tasks that could be substituted for the present one, children are less tempted to substitute, and can do a better job at delaying their response, resulting in improved performance. (2) The second possibility is that children will still be tempted to respond, this time simply pointing at random. Thus, while in the standard unexpected-transfer task, a child with poor inhibitory control will point to where the marble is (the box), in the undisclosed-location task she will point at random, resulting in an overall improvement in performance.

2.1.2 The Duplo Task

Paula Rubio-Fernández and Bart Geurts (2013, 2016) argue that the reason three-year-olds children fail direct false-belief tests is that these tests tend to contain elements that are likely to draw children’s attention away from the target agent, thus disrupting their perspective-taking process. The authors suggest that two features are likely to be particularly disruptive. First, the target agent typically disappears at some point during the task. Second, either the test question or the narration that immediately precedes it (or both) mention the object, which may direct children’s attention onto the object and away from the character. To test their hypothesis, Rubio-Fernández and Geurts (2013) carried out a modified unexpected-transfer task (which they call the Duplo Task) with two important task manipulations: (1) the target agent (“Lola”), instead of disappearing during the change of location, remained in sight, although at some distance from, and giving her back to, the location of the object (“bananas”); (2) an open-ended question was used, which did not mention the object (e.g., “What is Lola going to do now?”). The three-year-olds in this study passed when

both task manipulations were in place but failed when only one of the two was implemented. In a follow-up study, Rubio-Fernández and Geurts (2016) show that even with both task manipulations in place, three-and-a-half-year-olds can still fail if, right before the test question, the experimenter either asks a control question mentioning the object (“Where are the bananas now?”) or states the protagonist’s desire for the object (“Now Lola is very hungry and wants a banana”).

The processing-time account is in line with Rubio-Fernández and Geurts’ own proposal, although (arguably) it provides more details concerning the mechanisms that disrupt children’s perspective taking. Now, remember that we are working under the assumption that children will be tempted to swap the task at hand with an easier one. It seems reasonable to assume that increasing the saliency of a potential candidate for substitution will make the temptation to substitute correspondingly stronger. This is likely to be why Rubio-Fernández and Geurts found that mentioning the object had a detrimental effect, since, as they themselves suggest, mentioning the object is likely to draw children’s attention to its location. Thus, as children interpret the question, they will be more tempted to cut their processing short, jumping to the conclusion that the experimenter wants to know where the bananas are; as they predict where Lola will go, they will be more tempted to conclude that she will reach for the bananas at their current location; and, finally, as children select their response, they will be more tempted to point to the bananas. This means, notice, that children will have to endure the temptation to substitute three times in short succession and giving in just once is sufficient to ensure failure. In contrast, if the bananas are not mentioned, the temptation to substitute may not arise until later in their processing (for example, it may only arise when they recall the events in the story as they try to predict what Lola will do.)

What about the second task manipulation? Whether Lola stays in sight or disappears is likely to affect children’s ability to hold in mind the information that relates to her. Since three-year-olds tend to have very limited working memory capacity, the information concerning what Lola believes may be deleted after she leaves the scene; after all, children do not know that she is coming back, which means that, from their point of view, what Lola believes is not so important after all. In other words, the information about what Lola thinks may be assigned “low priority,” making it a likely candidate for elimination in case working memory were close to full. Another (potentially complementary) mechanism at play may be that of refreshing. Many models of working memory assume that, to prevent the contents of working memory from decaying, one must refresh them

every few seconds by briefly focusing one's attention on them (e.g., Baddeley, 1986; Barrouillet, Bernandin, & Camos, 2004). It is possible that, by staying in sight, Lola caught children's attention more, thus leading them to refresh the information about her beliefs, preventing it from decaying. Either way, if the information about Lola's belief is lost, children will then have to recalculate it after they process the question which may well bring their processing time above threshold.

It is worth noting that whether the target agent remains in sight throughout the test should be irrelevant from the point of view of pragmatic accounts, for this is unlikely to affect how children interpret the test question. Thus, the fact that the processing-time account can explain why this task manipulation helps provides one reason to prefer it to pragmatic alternatives (more reasons will be given below).

2.2 Response-Generation Processes

On the EPD account, the demands imposed by response-generation processes play an important role in explaining why young children have trouble passing direct false-belief tests. Response-generation processes, remember, are supposed to be those involved in “interpret[ing] the question, hold[ing] it in mind, and generat[ing] a response” (Setoh et al., 2016). These processes play an important role on the processing-time account too, although they do so because they increase processing time and not because of the processing demands they impose. As a result, the processing-time account is consistent with the evidence supporting the EPD account, including the study by Setoh et al. (2016). This is worth of notice given that Setoh and colleagues have obtained success on a direct false-belief test with remarkably young children, providing some of the best evidence for performance accounts to date. Let us take a look at their study.

2.2.1 The Emma Task

To test their account, Setoh and colleagues (2016) carried out an undisclosed-location task which included response-generation practice trials. The guiding hypothesis was that the two practice trials, allowing children to practice their response-generation processes, might reduce the load imposed by these processes during the test phase. The story was narrated with the help of a large picture book depicting key events and objects in the story. The story begins with a girl (Emma) who finds an apple in one of two containers (a bowl and a yellow box.) Next, children were shown

pictures of an apple and a banana side by side, and asked, “where is the apple?” (this was the first practice trial.) Then, the story reprised: Emma puts her apple in the box and goes out to play with the ball. Children were then shown two pictures side by side, one of a Frisbee and one of a ball, and asked, “where is the ball?” Finally, children were told the last part of the story: while Emma is out playing, her brother Ethan finds the apple and takes it away. Emma comes in looking for her apple; where will Emma look for her apple?” Setoh and colleagues found that a group of two-year-olds (mean age two years and eight months) were above-chance on this task.

Can the processing-time account explain Setoh and colleagues’ remarkable finding? Yes. Allowing children to practice their response-generation processes is just as likely to reduce the time it takes children to carry out these processes as it is likely to reduce the processing demands they impose. In fact, the two effects are likely to be related; for example, practising a task is sometimes argued to result in automatisisation, and automatic processes are both faster and less cognitively demanding.

Again, it is worth noticing that this finding is not easily explained on pragmatic accounts. If the reason young children fail is that they think adults are unlikely to talk about beliefs, it is unclear why the practice trials should help them, given that they consisted in pointing at objects. Similarly, if the reason young children fail is that they cannot take a second-person perspective on the experimenter while taking a third-person perspective on the target agent, it is again unclear why the practice trials should help. Thus, the fact that the processing-time account can explain this finding provides further reason to prefer it to pragmatic accounts.

2.3 Pragmatic Interpretation

Pragmatic accounts maintain that children have difficulties interpreting the test question, difficulties that concern, specifically, the pragmatics of the question rather than its semantics or syntax. As a result, children misunderstand what the experimenter wants to know. Now, on the processing-time account, pragmatic interpretation contributes to creating children’s difficulties by increasing the overall processing time. Furthermore, being under time pressure, children may be tempted to cut their processing short, which in some cases may result in a misinterpretation of the question. Still, on the processing-time account, pragmatic interpretation is only part of the problem; in fact, in contrast with pragmatic accounts, the processing-time account is consistent with children

being perfectly able to interpret the question when not under time pressure. Still, as we are about to see, the processing-time account is similar enough to pragmatic accounts to explain the evidence that is often adduced in their support.

2.3.1 Temporally-marked tasks

Michael Siegal and Karen Beattie (1991) defend an early pragmatic account of false-belief failure. On their account, the reason most three-year-olds point to the actual location of the toy is that they think the experimenter wants to know where the character will have to look to find the object, as opposed to where she will look first; and, of course, where she will have to look is where the object actually is. To test this hypothesis, the authors carried out an explicit false-belief task adding a condition with a modified question, which included the temporal marker “first” (as in: “where will Sally look first for her marble?”). They found that the modified question was sufficient to bring the three-year-olds in their study from significantly below-chance to significantly above-chance. In a second experiment, they also included two control conditions, namely a true-belief condition with the standard question and one with the modified question. This was to rule-out the possibility that, upon hearing the words “first” and “marble”, children simply pointed to the first location of the marble, regardless of what the character believed. Just as predicted by their pragmatic account, Siegal and Beattie found that, in the true-belief condition, the modified question had no significant impact on children’s performance.

A similar result was obtained by Charlie Lewis and Amanda Osborne (1990) with an unexpected-content task. Lewis and Osborne included both a standard condition with the typical question and a condition with a modified question that included the temporal marker “before” (“What will ... think is in the box before I take the top off?”). In the standard conditions three-year-olds were below chance, whereas in the “before” condition they were above chance.

The processing-time account has an easy explanation for these results: the temporal markers, by making what the experimenter wants to know more explicit, are likely to significantly cut down the time it takes children to process the question. In other words, temporal markers contribute to ruling out some possible interpretations of the question. For example, the temporal marker “first” implies a contrast between the location where the character will look first and the location where she will end up looking or will have to look; thus, these otherwise plausible interpretations are ruled

out. Since children have a smaller number of interpretations to consider, they should be able to complete their processing in a smaller amount of time. In addition, the absence of likely candidates for substitution may reduce the temptation to cut their processing short.

2.3.2 The Elmo Task

Mikkel Hansen (2010) argues that children are likely to interpret the question “where does Sally think the marble is?” as concerning the location of the marble rather than Sally’s belief about it, thus assuming that “think” is being used indirectly. To test this hypothesis, Hansen carried out three false-belief tasks with three-year-olds, each of which included both a standard condition with a typical “think” question (e.g., “where does Elmo think the marble is?”) and an emphasized-context condition with modified question aimed at discouraging the parenthetical reading (e.g. “You and I know that the ball is in the red cup, where does Elmo think it is?”). On all the tasks, children performance was significantly better in the emphasized-context condition compared to the standard condition, although only in two tasks out of three was performance on the emphasized-context condition above-chance.

The emphasized-context question is likely to help for the same reason temporally-marked questions help: it rules out potential (but incorrect) interpretations of the question, making the process of pragmatic processing easier, and thus allowing children to more quickly settle on the correct interpretation.

2.3.3 Deceptive false-belief tasks

A few studies have found that actively involving three-year-olds in the manipulation that leads to the target agent holding a false belief significantly improves their performance. Kate Sullivan and Ellen Winner (1993) for example, carried out an unexpected-contents task with an added “trick” condition where the child helped a confederate replace the contents of a box, so as to deceive a second confederate. The authors found that three-year-olds were below-chance in the standard condition, but above-chance in the trick condition.

Helming et al. (2016) argue that the fact that the experimenter encourages children to trick the target agent helps by making a normative interpretation of the question (what should ... think is inside the box?) less likely. This interpretation is consistent with the processing-time account,

since having one less interpretation to consider will reduce the time it takes children to process the question. There is also another possible explanation, however. Since in the trick condition the child has the goal of leading the agent to believe that the tube contains Smarties, that this is what the agent is likely to believe will be a most salient piece of information, one that will be assigned the highest priority (or that will be more frequently refreshed) and will thus be unlikely to be deleted were working memory to get full (or to decay due to lack of refreshing). In the standard condition, instead, children could not calculate the target agent's belief until after the experimenter asked the question, since the target agent was not present and had not been mentioned before. Thus, the difference between the two conditions may be due to the fact that in the trick condition, children are very likely to have carried out an important part of the processing before the test phase, leading to a significant reduction of processing time.

2.4 Summary

To sum up the discussion in this section, the fact that processing-time account acknowledges the importance of executive functioning, pragmatic-interpretation and response-generation processes in determining whether children pass or fail a direct false-belief tasks allows it to explain all the main findings supporting performance accounts. Notably, we have seen that this is not true of other performance accounts, including most notably pragmatic accounts. I will now argue that the processing-time can also explain all the evidence for late-emergence accounts discussed in Chapter III.

3. Explaining Away Evidence for Late-Emergence Accounts

3.1 Brain Maturation

Among the performance accounts we discussed in Chapter III, Carruthers' (2013) triple-load account was the one that, arguably, had the best explanation for why children who perform better on direct false-belief tests should have a more mature mentalising network. The explanation in question relied on the hypothesis that a more mature mentalising network would be able to operate

more efficiently, consuming fewer executive resources, thus resulting in children having to sustain a smaller processing load. It is worth noting that this explanation is consistent with the processing-load account, since processes that are more efficient tend to be faster as well (think of how your laptop or smartphone is slower to respond when rapid access memory is running low.)

While the processing-time account is consistent with this type of explanation, it can also provide one that is arguably better supported by the evidence. As mentioned in §1.2.1, Wiesmann and colleagues found that children's direct false-belief performance correlated with the degree of fractional anisotropy in the white matter surrounding several of their mentalising areas; and many studies have found fractional anisotropy to correlate with processing speed. This also makes sense on theoretical grounds, since (as Wiesmann and colleagues themselves note) increases in fractional anisotropy during childhood may reflect processes of myelination and axonal growth, both of which can be expected to lead to faster propagation of neural impulses. In turn, faster propagation of neural impulses may lead to the neurons in a region firing in a more synchronous manner, resulting in higher regional current density, which would explain the finding by Sabbagh and colleagues as well. In brief: there is both solid evidence and theoretical grounds to expect that higher levels of fractional anisotropy in the white matter surrounding the mentalising areas should correspond to faster mentalising processes. In contrast, I know of no evidence suggesting that an increase in fractional anisotropy may reflect an increase in efficiency. This means that the triple-load account is forced to rely on a very speculative hypothesis, while the explanation provided by the processing-time account has good evidence in its support. Overall, then, the processing-time account comes out on top so far as brain maturation is concerned.

3.2 Executive Functioning

Like processing-load accounts, the processing-time account predicts that children with better executive functioning will have a better chance of passing direct false-belief. This, for mainly two reasons: first, having better working memory may decrease processing time by increasing processing efficiency (as argued in §1.2.2); second, having better inhibitory control may increase time available for processing by allowing children to delay their response for longer (as argued in §1.3.1). The fact that both working memory and inhibitory control have been found to correlate with direct false-belief performance is thus just as the processing-time predicts.

As I have argued in §2.1, the processing-time account is close enough to processing-load accounts that it can explain the evidence typically adduced in favour of the latter. On the other hand, the two types of account are not the same, and the differences between the two are important because they allow the processing-time account to explain evidence that processing-load accounts cannot explain.

3.3.1 Executive functioning is *not* the only factor at play

The processing-time account differs from both the ToMM/SP and the EPD accounts in that it maintains that executive functioning is only one of the factors determining whether young children pass or fail a direct false-belief test. Remember, for example, that the time available for processing is determined by the interaction of inhibitory control with motivation. Having more inhibitory control will make children better able to delay their response; however, a child that is able to wait may still not wait if she is not sufficiently motivated to do so. Thus, individual differences in motivation may well reduce the correlation between inhibitory control and direct false-belief performance. Similarly, having more working memory capacity will allow children to keep the information about the character's belief in working memory while processing the question; however, that information may still be lost due to decay if children, having low processing speed, take too long to process the question. Thus, individual differences in processing speed may well reduce the association between working memory and direct false-belief performance. Accordingly, while the ToMM/SP and EPD accounts predict that executive functioning and direct false-belief performance should be strongly associated, the processing-time account is consistent with there being only a weak association, and thus with the results of the meta-analysis by Devine and Hughes (2014).

The fact that, on the processing-time account, executive functioning is only one of the factors at play also allows it to account for the results of the study by Sabbagh et al. (2006), who found that Chinese children, despite performing better than American children on tests of inhibitory control, did not perform better on direct false-belief tests. As mentioned in Chapter III, §1.2, the authors take this finding to support an emergence account of the association, where the two variables correlate because inhibitory control facilitates learning that beliefs can be false which in turn results in better direct false-belief performance. Since, Sabbagh and colleagues suggest, Chinese

children do not have siblings, they have less opportunities to learn about the mind, which will then balance out the beneficial effect of having better inhibitory control. Notice, however, that the finding is also consistent with the processing-time account, since on this account, as well, inhibitory control is not the only factor at play, and may thus be balanced out by other factors. In particular, given that the processing-time account is consistent with exposure to mental-state talk having a beneficial effect on direct false-belief performance (as argued in §1.3), it is consistent with Sabbagh and colleagues' suggestion that between-group differences in exposure to mental-state talk are what cancels out between-group differences in inhibitory control.

The interaction between the effect of executive functioning and that of practice on processing time may also explain the study by Benson et al. (2013), showing that children's performance on executive functioning tests predicted the extent to which their direct false-belief performance would improve after training. Children with better inhibitory control, being able to delay their response for longer, will have a higher TAP; consequently, on average, they will require a smaller reduction in their RPT to go below threshold. Thus, it is possible that while both groups experienced a similar reduction in RPT due to the practice effect of training, only in children with good inhibitory control was this reduction sufficient to bring RPT below threshold.

3.3.2 There is No *Specific* Prepotent Response that Requires Inhibition

Another important difference between the processing-time account and processing-load accounts is that the former predicts that inhibitory control should be recruited even in the absence of a specific prepotent response that must be inhibited. This is because, on the processing-time account, what must be inhibited is not just the temptation to provide a specific response but the temptation to provide *a* response (*any* response) within a few seconds. Making an incorrect response less salient will not be sufficient to remove children's temptation to answer quickly, since children may still point randomly; this, I have argued in § 2.1.1, may be what happens in undisclosed-location tasks, where children do not know where the object is. At the same time, the presence of a very salient and "easy" response may well increase the temptation to answer quickly: this, I have argued in §2.1.2, is why mentioning the object in the test question has a detrimental effect on children's performance. Now, this difference between the processing-time account and processing-load accounts may appear subtle, but it does result in different predictions. Most notably,

differently from processing-load accounts, the processing time accounts predicts that children's performance on tasks where no specific prepotent response is available should still correlate with their inhibitory control. As it happens, this prediction holds up.

Consider, for example, false-belief explanation tasks, where children are asked to explain the behaviour of an agent with a false belief; for example, children may be asked why Sally went to look for the marble in the basket instead of the box, where the marble really was. It is not clear what the prepotent response would be in this task, and indeed, children tend to provide a wide range of responses: some mention what the character thinks ("she thinks it's there"), some mention what she wants ("she wants it"), some mention previous events ("she put it there") some mention current facts ("because the marble is there") some profess ignorance ("I don't know"), with no explanation type being predominant. Nonetheless, Josef Perner, Birgit Lang and Daniela Kloo (2002) found that children did not perform better on this type of task compared to a standard, "prediction" false-belief belief task. In addition, children's performance on both tasks was found to correlate, to a similar extent, with performance on an executive functioning task. In addition, the study by Powell and Carey (2017) (discussed in Chapter III, 1.2) found that if children were given a task aimed at depleting their inhibitory control immediately before either a prediction or an explanation false belief task, their performance on both false-belief tasks was negatively affected, thus showing that both recruited inhibitory control. The processing-time account is consistent with this result, since even in explanation tasks children will be tempted to answer quickly instead of waiting to complete their processing, thus mentioning any fact that comes to mind or simply professing ignorance.

This feature allows the processing-time account to account for another finding that is problematic for processing-load accounts. Carlson et al. (2015) found that the think-know and source-of-knowledge tasks, which, they argue, do not make demands on inhibitory control, correlated with a test of inhibitory control as strongly as a direct false-belief task. Notably, the claim that the think-know and source-of-knowledge tasks do not make demands on inhibitory control is based on the consideration that young children, instead of being consistently incorrect, are at-chance on these tests. While this does show that there was no specific prepotent response, however, it does *not* show that children were not tempted to answer prematurely. If children are tempted to respond quickly, they will do so even if they are not biased towards a specific answer; they might answer at

random or profess ignorance (“I don’t know”). On the processing-time account, there is thus no reason to expect that children’s performance on these tests should not correlate with inhibitory control. That the processing-time can explain this type of result is especially notable since, as argued in Chapter III, no other performance account can explain it.

3.2.3 Summary

In conclusion, the processing-time can explain all the evidence on the relationship between executive functioning and direct false-belief performance that was discussed in the last chapter; this sets the processing-time account apart from the all the other performance account considered in Chapter III.

3.3 Language Ability

The processing-time account gives us a few options in explaining the relationship between language ability and direct false-belief performance. These options are detailed below.

3.3.1 Better Language Ability Leads to Faster Processing of Linguistic Stimuli

Children with better language abilities should be faster at processing the linguistic stimuli in direct false-belief tests, including both the narrative elements, where present, and the question (or questions). Other things being equal, children with more language ability will have a better chance of completing their computations in the time available. Notice that this is true even though, as argued in Chapter III, §2.1.3, most children above the age of three are likely to possess sufficient language proficiency to correctly decode the literal meaning of the linguistic stimuli present in direct false-belief tests. In contrast with all other performance accounts, then, the processing-time account is consistent with evidence showing that two-year-olds pass verbal indirect false-belief tasks (Scott et al., 2012) and that language ability and direct false-belief performance continue to correlate at least up to the age of seven (Milligan et al., 2007).

3.3.2 Language Ability Mediates the Effect of Experiential Factors

Children’s language abilities may also affect the extent to which they benefit from exposure to mental-state talk. On the processing-time account, participating in conversations about the mind

with their parents helps to the extent that it encourages children to more frequently engage in mentalising; however, patently, this is only true so long as children understand what their parents are saying. Notice that this type of explanation is compatible with the finding that language correlates to similar extents with false-belief tasks that have different language demands (Milligan et al., 2007).

3.3.3 Language Ability Facilitates Mentalising

Another possibility is that natural language provides children with a medium for representing mental states, as Jill and Peter de Villiers suggest (J. G. de Villiers, 2005; J. G. de Villiers & Pyers, 2002; P. A. de Villiers, 2005; P. A. de Villiers & de Villiers, 2012). Of course, from an early-emergence perspective, we cannot agree with the contention that natural language is *necessary* for representing mental states; after all, we are committed to the claim that largely non-verbal infants can already ascribe false beliefs. Notice, however, that just because natural language is not *necessary* for representing mental states does not entail that it is never used to represent mental states. All we need to get this type of explanation going is the claim that a sizable portion of children will think in language when prompted linguistically. In other words, the hypothesis under consideration is that, when asked where Sally will look for her marble, a sizable portion of children will engage in inner speech, using natural language to reason about what Sally thinks (“uh-oh...where does Sally think the marble is? Well, she put it in that box...uhm..she wasn’t there while the other one moved it...does she still think it’s there?”). Children’s language ability will of course facilitate these episodes of inner speech, affecting the ease and speed at which they unfold. Thus, children with better language ability can be expected to have a lower RPT, and thus a better chance of passing the test.

In addition, language may also be instrumental for holding information about mental states in working memory. Carruthers (2014), for example, argues that purely amodal, non-sensory representations cannot be held in working memory, since working memory relies on the sensory circuitry in the brain. To illustrate, keeping a phone number in mind seems to always rely on some type of imagery, like imagining seeing the corresponding digits, or imagining hearing the corresponding numerals, and so on. If this hypothesis is correct, then language may provide the necessary medium in which information about other people’s beliefs can be held in working memory.

Thus, it is possible that children’s language abilities also mediate their ability to hold in mind information about mental states.

Notice that both these hypotheses are consistent both with the finding that language training improves direct false-belief performance (C. Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003) and that language correlates to the same extent with direct false-belief tasks with varying language demands (Milligan et al., 2007). This is worth mentioning given that, to the best of my knowledge, no other performance account can explain this evidence.

3.3.4 Summary

The processing-time account is the only performance account that can explain all the evidence on the relationship between language ability and direct false-belief performance that was discussed in Chapter III. This provides reason to prefer it to the alternatives.

3.4 Social Competence

The processing-time account maintains that those children who are quicker at interpreting linguistic stimuli, predicting or explaining behaviour, and planning and executing their response, and/or those children who are both better able and more motivated to delay their response, should have a better chance of passing direct false-belief tests. These same children should also have a better chance of passing many tests of social competence, and for similar reasons. Many of tests of social competence are likely to recruit similar processes: language comprehension and production, pragmatic interpretation, executive functioning and mentalising. Furthermore, tests of social competence often put children under time pressure. Consequently, it makes sense that children who are above threshold in direct false-belief tests should also tend to be above threshold in tests of social competence. To bring this point home, let us consider two examples.

3.4.1 The Peer-Persuasion task

Consider now the peer-persuasion task used by Peterson et al. (2017). The task involved convincing a puppet (“Matty”) to do something he was not inclined to do, like brushing his teeth or eating broccoli. Children received one point for each “high-level” argument they used, which the authors define as arguments that “had to supply new substantive information aimed at shifting

Matty's initial mental state of resistance". In contrast, "[e]mpty, unelaborated denials, queries, commands, and requests that included no new substantive information" (2017, p. 5) would not award them any points. Now, just as in a direct false-belief test, children are likely to feel the clock ticking. Furthermore, depending on how fast their mentalising is, coming up with a convincing argument may take several seconds and a good deal of effort. In addition, low-level alternatives (like responding with "It's not yucky!" when Matty says "It's yucky!") will be readily available and the temptation to use them will be difficult to resist. Finally, children will also have to process what Matty says and produce their own utterances. Thus, in terms of processing, the test has a very similar structure to direct false-belief tests; it only makes sense, then, that the two should correlate.

3.4.2 Playing hide and seek and keeping a secret

Peskin and Ardino (2003) tested children on a playing-hide-and-seek test and on a keeping-a-secret test. In the hide-and-seek test simply consisted in playing two rounds of hide and seek (one as the hider and one as the seeker) with the experimenter; children were scored as failing if they made any of the following mistakes: "When hiding, if the child told E where he or she was going to hide; did not attempt to hide from view; hid before E's vision was blocked; or did not remain physically hidden and quiet. When playing the role of the seeker, if the child told E where to hide; or if the child did not turn around and/or cover his or her eyes; or if the child peeked" (2003, p. 500). In contrast, the keeping-a-secret test consisted in refraining from telling a hungry confederate that there was a cake in the fridge; children were told that the cake was a surprise for the confederate's birthday.

Now notice that, with respect to both tests, how fast children can predict the effect of their actions on the other people's mental states seems to be a critical variable. Consider the hide-and-seek task. A child may be able to figure out that, if the experimenter hears her giggle, she will be able to tell where she is hidden; however, to be of any use, this thought will have to occur to her before she starts giggling. Whether this is the case will, of course, depend on how fast the child can mentalise. Consider now the keeping-a-secret task. Suppose the child hears the confederate complaining that she is hungry. Given that children are typically very eager to help other people, the child might ask herself if there is any food available. The thought that there is cake in the fridge will of course occur to her, at which point the child will be tempted to inform the confederate and,

unless a good reason against doing this comes to her mind, she will initiate the response. Thus, the crucial variable is whether the realisation that telling the confederate about the cake will ultimately make her unhappy will come quick enough to inhibit the response.

3.4.3 Summary

The processing-time account offers an intuitive explanation of why children's performance on tests of social competence tests correlate with their performance on direct false-belief tests. Notably, among the performance accounts we have considered, the processing-time account is the only one that can explain this finding.

3.5 Developmental Progression

In explaining the developmental progression revealed by the ToM Scale, the processing-time account provides a couple of options, which are detailed below.

3.5.1 The “Pragmatics” Option

The explanation of the developmental progression put forward by Westra and Carruthers (2017) (discussed in §2.5.2) is largely consistent with the processing-time account. For the diverse belief task, Westra and Carruthers' suggestion is that young children, being accustomed to indirect uses of “think”, may take “Linda thinks the cat is in the bushes” (say) to imply “the cat is in the bushes”. As they infer that the experimenter is implying that the cat is in the bushes, many children will revise their original belief (e.g. that the cat was in the garage), eliminating the disagreement between them and Linda; in which case, even if they fall prey to a referential or cooperative misunderstanding of the test question, they will still provide the correct answer (the bushes). The only way for children to fail, then, would be if they persevere in their belief that the cat is in the garage. This is what, according to Westra and Carruthers makes the diverse-belief test harder than the diverse-desire test but easier than the explicit false-belief test. There is nothing here that conflicts with the processing-time account. The processing-time account maintains, remember, that young children tend to fail direct false-belief tasks because they give in to the temptation to answer before they have completed their processing. Notice, however, that if children think that the cat is in the

bushes then there is no reason for them to answer that Linda will look in the garage, regardless of whether they give in to attribute-substitution or not.

As for the knowledge-access test, Westra and Carruthers present two complementary explanations. The first is that the knowledge-access task is easier than the explicit false-belief task because in this case there are only two salient candidate interpretations (correct and referential) instead of three (correct, referential and cooperative). At the same time, the knowledge-access task is harder than the diverse-belief task because the referential interpretation, if taken, will lead children to the wrong response. This is compatible with the processing-time account: in our case, the fact that there are fewer salient candidate interpretations is important because it means that children will need less time to decide which interpretation is correct and will be less tempted to substitute an easier task. The second explanation is that the test question in the knowledge-access task is of the “yes-no” type, which young children are biased to answer positively. Thus, children must resist this bias to answer correctly. This fits nicely with the processing-time account: children must resist the temptation to answer “yes” in order to delay their response and complete the processing required to answer correctly. Thus, the processing-time account is also consistent with Westra and Carruthers (2017) suggestion that the reason Chinese children pass the knowledge access test before the diverse-belief test is that Chinese children have better inhibitory control.

3.5.2 The “Practice” Option

The processing-time account has also the option of staying closer to Wellman and colleagues’ (2011) own account of this finding. Wellman and colleagues argue that parents in collective societies like China and Iran are more likely to emphasise the importance of knowledge, while Western parents are more likely to stress diversity of opinion and belief, thus leading children to follow slightly different progressions as they build their theories of mind. On the processing-time account, these cultural differences may affect children by leading them to rehearse certain types of mentalising tasks more than others. If their parents are more likely to comment on what other people know or do not know, children will get more practice at ascribing states of knowledge or lack thereof, and will thus get quicker at doing this, either through automatisisation or by forming mental shortcuts or schemata. If, on the other hand, their parents are more likely to comment on differences of opinion, contrasting discordant beliefs, children will get more practice at ascribing

diverse belief and will thus become able to do this faster. Either way, if given time, children would be perfectly able to ascribe all these mental states (knowledge, ignorance, false beliefs); the claim is simply that they will be quicker at ascribing certain states compared to others depending on what they get to practice more.

3.5.3 Summary

In conclusion, not only the processing-time is largely compatible with the explanation provided by the pragmatic development account; it can also provide its own explanation. This is notable given that the pragmatic development account is the only other performance account that can explain this finding.

3.6 Experiential factors

As explained in §1.2.2, having a parent that is more likely to talk about mental states is likely to improve children's direct false-belief performance by giving them more opportunities to practice their mentalising abilities, which will then decrease the time it takes them to solve mentalising tasks, reducing their RPT. Similarly, receiving training on a direct false-belief task, or even on other direct mentalising tasks, would allow children to get more mentalising practice. Finally, having a sibling with whom to interact with daily should give children more incentive and more opportunities to mentalise, and may also elicit more mental-state talk within the family; either way, children with a sibling are likely to get more mentalising practice.

This account of how experiential factors can help children pass direct false-belief tests is arguably at least as good as (if not better than) those provided by the triple-load, pragmatic perspectives and pragmatic development accounts. In fact, it is very similar to the one offered by the triple-load account: on both accounts, these types of stimuli help children by allowing them to practice their mentalising abilities.

3.7 Deaf Children

In respect to deaf children, the processing-time account is consistent with two types of explanation. First, I have argued in §3.3 that there are several ways in which better language ability can lead to lower RPTs, thus potentially resulting in better direct false-belief performance. Now,

since deaf children of hearing are exposed to impoverished linguistic stimuli during their first years of life and exhibit language delays as a result, we can expect that their performance on direct false-belief tests will be correspondingly poor. Second, we have seen in §1.2.2 and §3.4 that participating in conversations about the mind gives children an opportunity to practice their mentalising abilities, again leading to shorter RPTs. Since deaf children of hearing parents have a much harder time participating in these types of conversation, they are likely to get much less practice at predicting and explaining behaviour. The effect of practice also allows the processing-time account to explain how false-belief training allows deaf children of hearing parents to improve their direct false-belief performance.

Overall, the combined effect of these two factors (i.e., worse language ability and less opportunities to practice mentalising abilities) should put deaf children of hearing parents at a serious disadvantage in direct false-belief tests and direct mentalising tasks more generally. Arguably, this account of deaf children’s mentalising delays is at least as good as those provided by other performance accounts. In fact, since the processing-time is the only performance account that can explain all the data on the relationship between language ability and direct false-belief performance, it may be argued to offer a better account of why deaf children of hearing parents have poor direct false-belief performance.

3.8 Summary

To the best of my knowledge, the processing-time account is the only performance account that can explain all the evidence discussed in Chapter III. (See Table 2 for a visual summary.)

	ToMM/SP	EPD	TL	PP	PD	PT
Indirect False-Belief Tests	✗	✓	✓	✓	✓	✓
Developmental Progression	✗	✗	✗	✗	✓	✓
Executive Functioning	✗	✗	✗	✗	✗	✓
Language Ability	✗	✗	✗	✗	✗	✓
Experiential Factors	✗	✗	✓	✓	✓	✓
Deaf Children	✗	✗	✓	✓	✓	✓
Social Competence	✗	✗	✗	✗	✗	✓
Brain Maturation	✗	✗	✓	✓	✓	✓

Table 2. Summary of Chapters III and IV. TL=Triple-Load Account; PP=Pragmatic Perspectives Account; PD= Pragmatic Development Account; PT=Processing-Time Account.

4. Conclusions

In this chapter, I have presented a new performance account (§1) and argued that it can explain away all the evidence, seemingly supporting late-emergence accounts, that was discussed in Chapter III (§3), while also explaining most of the findings that have been claimed to support other performance accounts (§2).

Overall, all the evidence discussed so far seems to point to the early-emergence approach as the best way of dealing with the discrepancy in the results of direct and indirect false-belief tests. We have seen in Chapter III that late-emergence accounts have trouble explaining the results of indirect false-belief tests. On the other hand, I have argued in this chapter that endorsing the processing-time account would allow early-emergentists to account for the results of direct false-belief tests plus the rest of the evidence that is typically adduced in support of late-emergence accounts (which was discussed in Chapter III). There thus seems to be no reason to opt for a late-emergence account.

In the next chapter, I am going to introduce a third way of dealing with our puzzle, one that has so far not been considered in the debate. Importantly, this alternative approach is also committed to the claim that infants can reason about beliefs but provides a different explanation for the data discussed in this chapter and in Chapter III, one that is closer to the one late-emergentists defend.

V

THE HYBRID APPROACH

In this chapter, I explore an alternative solution to the developmental puzzle of belief-reasoning. I put forward and defend a new type of approach, which I call the hybrid approach. Like the late-emergence approach, the hybrid approach maintains that the mentalising system recruited in direct false-belief tests would typically not allow children under four years of age to ascribe false beliefs. Like the early-emergence approach, however, the hybrid approach also maintains that the mentalising system recruited in indirect false-belief tests would already allow infants to do so. Contradiction can be avoided, I will argue, if we reject the assumption that these two types of test must be recruiting the same system, instead of recruiting distinct belief-reasoning systems.

An important theoretical commitment of the hybrid approach is to the claim that humans possess two mentalising systems. Thus, after introducing the hybrid approach in §1, in §2 I will continue to defend it by comparing it with another two-systems proposal, which has been very influential and has generated a good deal of research; this is the account put forward by Apperly and Butterfill (2009). Apperly and Butterfill argue that their two-systems account can explain otherwise puzzling data; they also argue, however, that doing so requires embracing a late-emergence account. I will argue that they are wrong about this: the hybrid approach can also explain the evidence and, arguably, can do so better than any late-emergence account, including theirs.

1. The Hybrid Approach

1.1 A New Approach

Let us start by recapitulating the debate so far, in broad strokes. The results of direct false-belief tests suggest that children under four cannot ascribe false beliefs, while the results of indirect false-belief tests suggest that infants can already do it. The two sets of findings thus seem to point in opposite directions, with the implication that one cannot take both at face value. Consistently with this interpretation, both the approaches we have discussed so far consist in taking one set of findings at face-value while trying to explain away the other. Thus, early-emergentists accept the face-value interpretation of indirect false-belief results but defend an alternative interpretation of direct false-belief results (discussed in chapter III and IV), while late-emergentists accept the face-value interpretation of direct false-belief results but defend an alternative interpretation of indirect false-belief results (discussed in chapter II).

The reason the results of direct and indirect false-belief tests are in conflict seems to lie, however, in the tacit assumption that both types of test recruit the same mentalising system. So long as we stick with this same-system view, we are forced to choose: either the belief-reasoning abilities grounded in this system emerge early, as the results of indirect false-belief tests suggest, or they emerge later on, as the results of direct false-belief tests suggest – both cannot be true. Notice, however, that the conflict disappears if we reject the same-system view. For if direct and indirect false-belief recruit distinct belief-reasoning systems, it becomes possible to argue that each would follow its own developmental trajectory. Thus, the belief-reasoning abilities grounded in the system that is recruited in indirect false-belief tests (call it the *indirect-FB system*) would develop early on, as the results of those studies indeed suggest; the belief-reasoning abilities grounded in the system recruited in direct false-belief tests (call it the *direct-FB system*), in contrast, would only develop around the age of four, as the results of these other studies suggest. This is the gist of the *hybrid approach*.

Suppose that this type of account is on the right track; then, both the early-emergence and the late-emergence approaches would be right about some things but wrong about others. If we restrict our focus to the indirect-FB system, then the early-emergence approach appears to be correct, for the belief-reasoning abilities grounded in this system do arise early on, and it is these abilities that

allow infants to pass indirect false-belief tests. If we restrict our focus to the direct-FB system, instead, it is the late-emergence approach that appears to be correct, for the belief-reasoning abilities grounded in this system do arise later on, and the fact that these abilities are typically not present in children younger than four explains why most of these children fail direct false-belief tests. When we broaden our focus to include both systems, however, it becomes clear that neither approach is ultimately correct. Early-emergentists go wrong when they assume that direct and indirect false-belief tests must recruit the same belief-reasoning system. The case of late-emergentists is a bit more complicated, since, as we will see in the next section, many of them do accept that direct and indirect false-belief tests recruit distinct cognitive systems. Even so, all late-emergentists maintain that infants cannot ascribe false-beliefs, and thus would deny that the indirect-FB system is a genuine belief-reasoning system. As a result, neither early-emergentists nor late-emergentists recognise that direct and indirect false-belief tests recruit distinct *belief-reasoning* systems.

1.2 The Late-Emergence Approach and the Distinct-System View

As we have seen in Chapter II, late-emergentists have put forward many different proposals concerning what may allow infants to pass indirect false-belief studies. The ones we discussed appealed, for example, to such diverse cognitive phenomena as associations, affordances, behaviour rules and several types of mentalising generalisations, just to mention a few. Despite the variety in approach, none of these alternative interpretations was found to offer a satisfactory account of indirect false-belief results. There is one important question that was not addressed in Chapter II, however, and which must be addressed now. The question concerns how the abilities that, according to late-emergentists, allow infants to pass indirect false-belief tests (whatever those may be) relate to the mentalising abilities that underlie performance on direct false-belief tests. Specifically: does the system that underlies infants' performance in indirect false-belief tests gradually transform itself into the system that underlies performance in direct false-belief tests? Or are these distinct systems, coexisting in parallel, one next to the other? The first type of answer is committed to the same-system view: indirect false-belief tests recruit the same system that is recruited by direct false-belief tests, although at an earlier stage of development. In contrast, the second type of answer is committed to the

distinct-system view: direct and indirect false-belief simply recruit distinct cognitive systems. To make this more concrete, let us see some examples.

Same-system view. First, take the late-emergence account defended by Gopnik and Wellman (1992), according to which children are equipped with an innate, “starting state” theory of mind, plus a domain-general learning mechanism which they can use to revise and improve upon that theory. The starting state theory, despite being mentalistic, would not yet contemplate representational states like belief; it is only around the age of four or five that this shift would take place, according to Gopnik and Wellman. Thus, two- and three-year-olds, lacking the notion that beliefs can misrepresent reality, would be unable to pass a false-belief test. Second, take Perner’s (2010) proposal that infants pass indirect false-belief tests by relying on superficial non-mentalistic generalisations (called *behaviour rules*) such as, for example, that people look for things where they last were in their line of sight. When applied to these two proposals, the question at hand would take the following form: how do the behaviour rules that, according to Perner (2010), allow infants to pass indirect false-belief tests relate to the non-representational theory of mind that, according to Gopnik and Wellman (1992), causes two and three-year-olds to fail direct false-belief tests? The same-system view maintains that the two simply represent different developmental snapshots of the same, developing cognitive system. In other words, the same-system view requires us to make a relatively minor revision to Gopnik and Wellman’s account: the innate “theory” children start with, instead of being mentalistic, would have to be behaviouristic, comprising one or more behaviour rules; this behaviouristic theory, which allows infants to pass indirect false-belief tests, would then get replaced by a mentalistic (but not yet representational) theory shortly after the second birthday. The account just sketched is, of course, just an example, and many of the details are inessential. In particular, behaviour rules can be replaced with associations, affordances or whatever other alternative account of indirect false-belief performance one favours. The crucial idea is that the capacities that allow infants to pass indirect false-belief tests, whatever those may be, represent an earlier stage of development compared to two- and three-year-olds’ understanding of the mind, and eventually get replaced by, or transformed into, it.

Distinct-systems view. Consider again the same two accounts as above, namely Gopnik and Wellman (1992) and Perner (2010). Each of these two accounts now applies a distinct system, and each system will now follow its own developmental trajectory. The direct-FB system would be the

domain-general learning system that Gopnik and Wellman (1992) describe, containing a starting-state theory that is mentalistic but not representational, just as the authors originally suggested. On the other hand, the indirect-FB system would be the one described by Perner (2010); thus, it would contain one or more behaviour rules, plus any cognitive machinery required to deploy them in reasoning. On this picture, two and three-year-olds fail direct false-belief tests because, at that age, the direct-FB system still contains a non-representational theory of mind according to which people act based on how the world really is. On the other hand, infants pass indirect false-belief tests because the indirect-FB system contains the behaviour rule that people look for things where they last were in their line of sight (or something along those lines). Replacing Perner's behaviour rule account with another proposal will change the part of the story that concerns the indirect-FB system but will leave the part about the direct-FB system substantially unaffected.

Many late-emergentists have embraced the distinct-system view (e.g. Apperly & Butterfill, 2009; Heyes & Frith, 2014; Low, Apperly, Rakoczy, & Butterfill, 2016; Rakoczy, 2017; Sabbagh et al., 2013; Wellman, 2014). And indeed, when looking at the data, this appears to be the most plausible option for late-emergentists. Same-system late-emergence accounts predict that two and three-year-olds should fail not just direct false-belief tests but also indirect false-belief tests; this is because, on this type of view, the system that allows infants to pass indirect false-belief tests transforms into the system that causes two and three-year-olds to fail direct false-belief tests. We should thus see a u-shaped trajectory, with children younger than two passing indirect false-belief tests, children between the ages of two and four failing both direct and indirect false-belief test, and children older than four passing both direct and indirect false-belief tests. What we see, instead, is that two- and three-year-olds keep passing indirect false-belief tests while failing direct false-belief tests (e.g., Clements & Perner, 1994; Garnham & Ruffman, 2001; He et al., 2011, 2012; Ruffman, Garnham, Import, & Connolly, 2001; Southgate et al., 2007); indeed, the study by Clements and Perner (1994) shows that the very same three-year-olds who just looked at the empty location in anticipation for the agent's return will point to the actual location of the object when asked where the agent would go, displaying both responses in short succession. This seems to effectively rule out same-system late-emergent accounts.

1.3 Explaining Evidence for Late-Emergence Accounts

We have just seen that a distinct-system type of account is not only a very popular choice among late-emergentists but may also be the only viable option within the late-emergence camp. Now, notice that the disagreement between hybrid and distinct-systems late-emergence accounts concerns mainly what the indirect-FB system would do. Specifically, the hybrid approach maintains that the indirect-FB system would allow infants to ascribe false beliefs, while distinct-system late-emergence accounts deny this. The two types of approach are in substantial agreement, however, so far as the direct-FB goes: in particular, both maintain that the direct-FB system would not allow children to ascribe false beliefs until about the age of four. This is important because it is this claim that allows distinct-system late-emergentists to explain all the evidence discussed in Chapter III. Since the hybrid approach also accepts this claim, it can explain that evidence in the very same way.

Consider, for example, evidence suggesting that children who are more exposed to certain types of social stimuli (i.e. children who have siblings, and/or whose mothers are more likely to talk about mental states) tend to have better direct false-belief performance (this evidence was discussed in Chapter III, §1.4). This type of finding suggests that being exposed to certain types of social stimuli helps young children acquire belief-reasoning abilities. Those late-emergentists who embrace the distinct-system view will maintain that the system that underlies this process of acquisition is the direct-FB system. Now, since the hybrid approach maintains that the direct-FB system develops just like late-emergence approaches suggest, it can endorse essentially the same type of explanation: social stimuli help young children acquire belief-reasoning abilities. The only qualification necessary is that the belief-reasoning abilities children would acquire are those grounded in the direct-FB system, not those grounded in the indirect-FB system, since those would already be present since infancy.

1.4 An Argument for the Hybrid Approach

The main argument for the hybrid approach is that it can explain the results of indirect false-belief tests better than late-emergence accounts, and the results of direct false-belief tests better than early-emergence accounts. We saw in Chapter II that none of the proposals late-emergentists have put forward provides a compelling explanation of indirect false-belief results. To explain those results, I argued, we must grant that infants can ascribe false beliefs. The hybrid approach accepts this

claim, since it maintains that the indirect-FB system would sustain belief-reasoning abilities since infancy. On the other hand, in Chapter III, we saw that the performance accounts that early-emergentists have put forward struggle to explain the evidence that late-emergentists have gathered in support of their accounts, while in the last section I argued that the hybrid approach can explain those data as well as (indeed, in the very same way as) late-emergentists do. Overall, the hybrid approach thus seems to get the best of both worlds.

1.5 The Hybrid Approach and the Nativism-Constructivism Debate

As explained in Chapter I, one of the reasons the question of whether belief-reasoning abilities emerge early on is worth investigating is that it has important implications for the debate between nativism and constructivism. In particular, early-emergence approaches are difficult to reconcile with a constructivist position, and thus provide support for nativism. Since we now have a new player in the game, it makes sense to ponder its implications for the debate.

The hybrid approach maintains that there are two cognitive systems, each of which can sustain belief-reasoning abilities independently from the other. Furthermore, it maintains that these two systems would follow different developmental trajectories. The question we are considering must thus be split into two parts: (a) does the hybrid approach support a nativist account of the indirect-FB system, or does it support a constructivist account? (b) Does the hybrid approach support a nativist account of the direct-FB system, or does it support a constructivist account? Let us consider each of these questions in turn.

The indirect-FB system. So far as the indirect-FB system is concerned, the hybrid approach seems to straightforwardly support nativism. This is because, on the hybrid approach, the indirect-FB system would already allow infants to ascribe false beliefs. This, in turn, provides evidence that many of the cognitive structures required for that purpose must be innate and domain-specific, just as nativism demands.

The direct-FB system. The case of the direct-FB system is a bit more complicated. The hybrid approach maintains that the direct-FB system would not allow children younger than four to ascribe false-beliefs. This claim can be given a constructivist spin, in which case the direct-FB system would be domain-general and would contain only a bare minimum of innate structure. Thus, whereas the

indirect-FB system would come with many mental state concepts and folk psychological generalisations already built-in, the direct-FB system would have to acquire all those structures through learning, unaided by any domain-specific mechanisms or heuristics. Discovering that beliefs can be false would then be but one step in a continuous learning process.

It is important to stress that there are other options, however. In particular, the late emergence of a trait is not necessarily inconsistent with nativism (as mentioned in Chapter I). Indeed, nativists have at least two options in explaining late emergence: slow maturation and “nativist” learning.

Slow-Maturation Approaches. A slow-maturation type of account would argue that, while the direct-FB system comes with all the cognitive structures required for ascribing false beliefs already built-in, some of these structures would typically not reach maturation until about the age of four (Segal, 1996; Stich & Nichols, 1998; see also Scholl and Leslie, 1999 for discussion). Notably, it is possible to maintain that the direct-FB system is made up of several sub-systems that would reach maturation at different ages. Thus, for example, one can argue that the direct-FB system would comprise a sub-system for desire-reasoning and one for belief-reasoning, which would come on-line at different ages, explaining why young children pass the diverse-desire test but not the false-belief test. Furthermore, one can argue that the process of maturation is sensitive to the stimuli the organism is exposed to. For example, a module may have a set of parameters that are set based on the stimuli the organism is exposed to, and which would determine what the mature module looks like (Segal, 1996; Stich & Nichols, 1998). This would allow a slow-maturation account to accommodate evidence that children who are exposed to more social stimuli (mental-state talk, siblings and the like) tend to have better direct-FB performance. Social stimuli would impact children's mentalising abilities not by allowing them to learn about the mind but by setting parameters in their mentalising module. This type of account is thus largely consistent with the evidence discussed in chapter III.

Nativist-learning Approach. In contrast, a nativist-learning type of approach would maintain that some of the cognitive structures that make up the direct-FB system are innate while others are learned thanks to domain-specific mechanisms (this option is discussed, though not endorsed, by Westra and Carruthers, 2017 and Scholl & Leslie, 1999). As mentioned in Chapter I, section §2.3, evidence from direct mentalising tests suggests that children possess at least some understanding of desire, perception and emotion as early as they have been tested. Thus, one can argue that the direct-

FB system is a domain-specific learning mechanism, which comes with innate concepts of desire, perception and emotion already built in and would acquire the rest through learning. (This is consistent, notice, with executive functioning and language also playing a facilitating role – the claim is that a domain-specific learning mechanism would be required, not sufficient.) We can thus expect this type of account to also be consistent with evidence discussed in chapter III.

To sum up this discussion: so far as the indirect-FB system is concerned, the hybrid approach supports a nativist account, while so far as the direct-FB system is concerned, it is largely consistent with both nativist and constructivist options.

2. Two Mentalising Systems

As we have seen, the hybrid approach maintains that both the indirect-FB and the direct-FB systems sustain belief-reasoning abilities; it is thus committed to a two-systems account of mentalising. In this section, I compare the hybrid approach to another, very influential two-systems account, namely the one defended by Apperly and Butterfill (2009). Notably, in the years following the publication of Apperly and Butterfill's paper, the two-systems account has gained the support of several late-emergentists, rapidly becoming one of the main proposals in the late-emergence camp (Heyes & Frith, 2014; Low et al., 2016; Rakoczy, 2017; Wellman, 2014).

Apperly and Butterfill argue that humans possess two mentalising systems with opposite processing profiles. This hypothesis, they argue, allows us to explain an otherwise puzzling pattern of results. We start in §2.1 by taking a look at this evidence. Then, in §2.2, I discuss Apperly and Butterfill's argument that explaining the evidence requires embracing the late-emergence approach. Finally, in §2.4 and §2.5, I argue that hybrid approach can actually explain the evidence better than the late-emergence approach Apperly and Butterfill's defend.

2.1 Evidence for Two Mentalising Systems

Apperly and Butterfill argue that several sets of findings provide evidence for a two-systems account, including: (1) conflicting evidence on whether mentalising is automatic and efficient or controlled and cognitively demanding; (2) conflicting evidence on whether mentalising is dependent on language; (3) evidence that non-human animals can mentalise; (4) conflicting evidence on the

emergence of mentalising abilities in children (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). Let us take a closer look at this data.

2.1.1 Automaticity

The argument from automaticity moves from the observation that there is conflicting evidence on whether mentalising is automatic and maintains that this evidence can be explained by positing two mentalising systems, one of which is automatic and efficient while the other is controlled and cognitively demanding.

Evidence that mentalising is automatic has come from a variety of tasks (D. Buttelmann & Buttelmann, 2017; Kovács et al., 2010; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010; Schneider, Bayliss, Becker, & Dux, 2012; Schneider, Nott, & Dux, 2014; van der Wel, Sebanz, & Knoblich, 2014). In the dot-perspective task by Dana Samson and colleagues (2010), for example, adult participants were first shown a screen with a digit, and then a picture of a room with red disks (“dots”) on the walls. The dots could be placed either on the left wall or on the right wall; a human avatar was standing in the middle of the room, facing the left wall. Participants were assigned to one of two tasks: the other-perspective task consisted in judging (by pressing one of two buttons) whether the number of dots the avatar could see was consistent with the previously shown digit; the self-perspective task, instead, consisted in judging whether the number of dots they themselves could see was consistent with the digit. In *consistent* trials, all the dots were on the left wall, so that all were visible to the avatar, while in *inconsistent* trials some dots were on the right wall, out of the avatar's sight. Reaction times and errors were recorded over a vast number of trials. Samson and colleagues found that participants were slower to respond and more error prone on inconsistent versus consistent trials, and this regardless of the task they were assigned. Thus, when participants' own perspective was not consistent with the avatar's, they were not only slower to judge what the avatar could see (egocentric interference) but also slower to judge what they themselves could see (“alter-centric” interference). This suggests that participants automatically calculated what the avatar could see, which then interfered with the processes involved in deciding which button to press.

Further evidence that mentalising is automatic comes from anticipatory-looking false-belief tasks (Schneider, Bayliss, et al., 2012; Schneider, Nott, et al., 2014). Dana Schneider and colleagues (2012) submitted adult participants to repeated trials which involved looking at movies depicting

typical unexpected-transfer scenarios. In false-belief trials, the agent in the movie left the room before the object was moved, while in true-belief trials she left after the transfer. Importantly, the task participants were instructed to carry out had nothing to do with mentalising: in the first experiment, this consisted in pressing a button when one of the agents in the movie waved at the other; in the second experiment, instead, the task was to discriminate between high and low tones played at different points during the movie. At the end of the experiment, participants were asked open-ended questions designed to reveal whether they had consciously reasoned about the agent's belief (those that did were excluded from the analysis.) Despite this and the fact that what the agent believed was irrelevant to their task, participants looked significantly longer at the empty location in false-belief trials compared to true-belief trials, again suggesting that they had automatically calculated the agent's belief.

Consistent with these findings, there is evidence suggesting that neither of these tasks (dot-perspective or anticipatory-looking unexpected-transfer) is taxing on inhibitory control (Low, 2010; Qureshi, Apperly, & Samson, 2010)¹. Adam Qureshi, Ian Apperly and Dana Samson (2010) carried out a dual-task experiment with adults using the dot-perspective task (based on Samson et al., 2010) plus an executive functioning task. While participants were slower to respond in all conditions when in the dual-task condition compared to the dot-perspective task alone, the presence of the executive functioning task did not erase the altercentric effect – if anything, this increased in magnitude. This suggests that calculating the avatar's perspective was not dependent on executive resources. On the other hand, the fact that response times were slower in dual-task conditions may reflect that the presence of executive functioning task slowed the processes required to decide which button to push and made it more difficult to inhibit the irrelevant perspective, thus resulting in increased egocentric and altercentric effects.

The evidence considered so far points to mentalising processes being automatic and efficient. There is also evidence pointing in the opposite direction, however. First, in the last two chapters we have discussed at length evidence suggesting that executive functioning is required for passing direct false-belief tests. In particular, remember the finding by Powell and Carey (2017) discussed in Chapter III, §1.2, showing that depleting children's inhibitory control with a depletion task

¹ A study by Schneider et al. 2014 seems to support the opposite conclusion; this will be discussed in Chapter VI.

immediately before a direct false-belief task negatively affected their performance. In chapter IV, I suggested that this result could be explained assuming that inhibitory control was tapped to resist the temptation of responding too soon. Another interpretation is also possible, however, one that Powell and Carey themselves defend: this is that inhibitory control is involved in mentalising itself. There is further evidence from adults that is also open to this interpretation (Bull, Phillips, & Conway, 2008; Lin, Keysar, & Epley, 2010; McKinnon & Moscovitch, 2007). Shuhong Lin, Boaz Keysar and Nicholas Epley (2010), for example, carried out a dual-task experiment using Keysar, Linn and Barr's (2003) director task plus a working memory task. In the director task, a confederate (the "director") instructs participants to give her an object and participants must determine which objects the director can see to determine which object she means. Thus, for example, the director might say, "pass me the pen" when there are two pens, only one of which is visible from her point of view. The working memory task involved memorising numbers, and had two conditions, a high-load one and a low-load one. Participants were slower to react, more error prone and made a higher number of fixation on the "competitor" object (the one that was hidden from the director's view) in high-load versus low-load condition. Importantly, the effect disappeared when the competitor object was not present, suggesting that cognitive load was indeed due to mentalising rather than general task demands.

Second, other studies suggest that mentalising is not automatic (Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006; Back & Apperly, 2010; Keysar et al., 2003). Back and Apperly, for example, showed participants an unexpected-transfer scenario and told them to track the location of the ball; then, they asked them either where the ball was or what the agent believed. Participants were significantly slower to respond to belief probes compared to reality probes, suggesting that they had not calculated what the agent believed until they were prompted to.

To sum up: some studies suggest that mentalising is automatic and efficient, while others suggest that it is controlled and cognitively demanding; one way to account for this conflicting evidence is to posit two mentalising systems, one of which would be automatic and efficient while the other would be controlled and cognitively demanding.

2.1.2 Language

The argument from language moves from the observation that there is conflicting evidence on whether language is required for mentalising and maintains that this evidence can be explained by positing two mentalising systems, one of which can operate independently of language while the other cannot.

Some of the evidence suggesting that language may be required for mentalising was discussed in the last two chapters (Chapter III, §1.3, §1.5). For example, direct false-belief performance correlates with language ability and can be improved through training on sentential complements; or that deaf children of hearing parents tend to have very poor direct false-belief performance. Further evidence comes from a study with adults by Ashley Newton and Jill de Villiers (2007). Newton and De Villiers carried out a dual-task experiment using a non-verbal false-belief task plus an interference task. In the false-belief task, participants watched movies depicting unexpected-transfer scenarios, with both false-belief and true-belief conditions. At the end, participants were shown both endings side by side, one of which had a green background while the other had a red background; participants had to respond by tapping either the red or the green section of a block they were holding. The interference task was either a verbal shadowing task or a rhythmic tapping task designed to have the similar attentional demands. Participants who were given the verbal shadowing task performed significantly worse on false-belief compared to true-belief conditions, while there was no significant difference for participants who were given the rhythmic tapping task. This suggests that even a non-verbal false-belief tasks, like the one used in this study, make demands on language.

On the other hand, there is also evidence suggesting that mentalising does not require language. The fact that infants pass indirect false-belief tests, for example, strongly suggests this, as does the evidence that some non-human animals possess mentalising abilities (to be discussed presently.) In addition, there is also some neuropsychological evidence pointing to a similar conclusion. Thus, Apperly, Samson, Carroll, Hussain, and Humphreys (2006) found that PH, a patient with sever impairment in general language ability and syntax, scored almost perfectly on two non-verbal false-belief tasks.

To sum up: some findings suggest that mentalising requires language, while other data suggest the opposite conclusion. One way to explain this conflicting evidence is to posit two mentalising systems, one that is dependent on language and one that is not.

2.1.3 Infants and Non-Human Animals

The argument from infants and non-human animals moves from the observation that infants and non-human animals, both of which have poor executive functioning and language abilities, seem able to mentalise, and claims that this provides evidence for the existence of a second mentalising system that can operate independently of executive functioning and language.

The evidence from infants was discussed at length in chapters I and II and should by now be familiar. Several non-human species have also been found able to pass indirect mentalising tasks (for reviews, see Call & Tomasello, 2008; Drayton & Santos, 2016; Emery & Clayton, 2009), however. In fact, two recent studies show that great apes can pass false-belief tests (D. Buttelmann, Buttelmann, Carpenter, Call, & Tomasello, 2017; Krupenye, Kano, Hirata, Call, & Tomasello, 2016). Buttelmann and colleagues, for example, used an interactive unexpected-transfer task similar to that introduced by D. Buttelmann et al. (2009), where the experimenter comes back after the transfer and struggles to open the empty box. The task included true-belief, false-belief and ignorance conditions; the great apes in this study (chimpanzees, orangutan and bonobo) reached for the other box, the one containing the object, significantly more often in false-belief conditions compared to both true-belief and ignorance ones, suggesting that they understood the agent was trying to open the empty box because she believed the object she was looking for was inside it.

The fact that infants and non-human animals pass mentalising tests is puzzling when one considers the evidence, discussed above, that mentalising requires both language and executive functioning, since both groups have very little of both. The puzzle can be solved by arguing that there are really two systems for mentalising, only one of which requires language and executive functioning. The system that is *not* dependent on language and executive functioning may then be present both in infants and non-human animals, perhaps due to its being evolutionarily ancient and present in a common ancestor.

2.1.4 Development of Belief-Reasoning Abilities

The last piece of evidence is the discrepancy in the results of indirect and direct false-belief tests (outlined in chapter I, section §1, and discussed above). This discrepancy, Apperly and Butterfill argue, can be explained assuming that the two types of test recruit distinct mentalising systems, following different developmental trajectories.

2.1.5 Summary

Altogether, the evidence discussed in this section suggests that humans possess two mentalising systems, one of which is fast, automatic, efficient, early developing and shared with some non-human animals, while the other is slow, controlled, cognitively demanding, later-developing and unique to humans. To ease the exposition, in the following I will refer to the first as the *fast* system and to the second as the *slow* system.

2.2 The Argument from Cognitive Demands

On the face of it, the claim that humans possess two mentalising system with opposite processing profiles seems entirely consistent with the hybrid approach. Apperly and Butterfill have an argument that suggests otherwise, however. The argument in question (call it the *argument from cognitive demands*) purports to show that one cannot reason about beliefs or other propositional attitudes in a fast and efficient manner. Based on this argument, Apperly and Butterfill suggest that the fast system is likely to employ a simplified, minimal theory of mind which posits registrations instead of beliefs. Infants would thus have to rely on these minimal mentalising abilities to pass indirect false-belief tests, a proposal we discussed (and rejected) in Chapter II, §3.3. Proper belief-reasoning abilities, in contrast, would be dependent on the slow system and would only emerge around the fourth birthday. It is thus the argument from cognitive demands, together with the evidence reviewed in the last section, which leads Apperly and Butterfill to embrace the late-emergence approach. The argument from cognitive demands thus stands in the way of the hybrid approach.

2.2.1 The Argument

Apperly and Butterfill (2009) argue that reasoning about beliefs and other propositional attitudes tends to be cognitively demanding because propositional attitudes have a number of problematic features, including, most notably, the following:

- (1) They can take indefinitely complex propositions as their content;
- (2) They are intensional with an “s”, meaning that they not only represent objects and properties, but seem to always represent them under some mode of presentation or other;
- (3) They interact with each other and with both stimuli and behaviour in very complex ways;

(4) They are not just causes of action but reasons for action, and thus possess a normative dimension, being subject to the norms of rationality and truth.

This list should look familiar; these are the features that supposedly separate beliefs from registrations (as discussed in Chapter II, section §3.3.) Registrations are supposed to be simpler than beliefs precisely because they lack the features mentioned in the list above; thus, while beliefs can take any proposition as their content, are intensional with an “s”, have a complex causal role and are subjects to norms of rationality, registrations can only have an object-location pair as their content, are extensional, have a relatively simple causal role and possess no normative dimension. As a result of this, one may expect reasoning about beliefs to be a more complex endeavour than reasoning about registrations.

Take, for example, a hypothetical child called Charlie, who is trying to predict where Sally will go using a minimal theory of mind. This should involve very little processing, and of a very simple kind. First, Charlie would pay attention to Sally's line of sight to determine which objects she encounters; then, Charlie would ascribe to her a registration of the marble in the location where she last encountered it; finally, Charlie would predict that Sally will look for the marble in the location where she registers it. In contrast, suppose now that Charlie must rely on a more sophisticated theory of mind to predict where Sally will go, specifically a theory that posits beliefs instead of registrations. Charlie will now have to carry out considerably more processing to get to the same result. For example, the fact that the basket is where Sally last saw the marble does not entail that she will take it to be there; it all depends on the other beliefs Sally happens to have. If, say, Sally knows that Anne is wont to pulling pranks at the expense of her friends, then she may anticipate that Anne will move the marble to the box. Thus, keeping track of where Sally last saw the marble will not suffice to determine where she thinks the marble is; to figure out the latter, Charlie may also have to consider what other beliefs Sally is likely to have. Furthermore, even once Charlie has ascribed to Sally the belief that the marble is in the basket, this again does not guarantee that she will look for it there, since beliefs do not lead to action directly but only by interacting with other beliefs, desires and emotions in complex ways. Overall, it thus seems reasonable to suppose that, if Charlie is reasoning in terms of beliefs as opposed to registrations, she will have to take into account a considerably larger number of facts before she can make a prediction, resulting in more processing.

2.2.2 An Objection and a Counterproposal

The argument from cognitive demands seems compelling at first, but, I believe, breaks down on closer inspection. To see why, let us suppose that indirect-FB processes require the intervention of executive functioning to carry out their computations. In this case, it seems reasonable to expect that indirect-FB processes would be cognitively demanding even if they involved reasoning about registrations instead of beliefs. This is because executive processes are so limited in capacity that even seemingly simple computational operations are enough to deplete them. The average working memory span, for example, seems to be about four items (e.g. Cowan, 2000), while the focus of attention is limited to one chunk of information at a time. Thus, if we suppose that the indirect-FB processes rely on working memory as a mental blackboard to carry out their computations, keeping track of what objects a person registers and where she registers them will rapidly use up all the space available (four items). On the other hand, if we suppose that indirect-FB processes do not recruit executive functioning, then it is unclear why they should consume any executive resources to begin with, independently of their complexity. Thus, it seems that for Apperly and Butterfill's argument to work, it must be the case that indirect-FB processes would recruit executive functioning if they involve reasoning about beliefs but not if they involve reasoning about registrations. There is no reason to suppose that this should be the case, however.²

The type of consideration just aired suggests an alternative model of cognitive demands. The idea is disarmingly simple: the reason direct-FB processes are cognitively demanding while indirect-FB processes are not is that the former are hooked up to the executive system and thus consume executive resources as they run, while the latter run autonomously and independently of the executive system. This type of suggestion is in line with what other dual-process theorists have argued in recent years. Many authors have defended a distinction between intuitive (/System 1, /Type 1) and reflective (/System 2, /Type 2) processes, across many domains (for reviews of this literature, see

² At times, Apperly and Butterfill seem to suggest that the evidence that direct false-belief tests load on executive functioning provides independent support for the claim that reasoning about beliefs must be cognitively demanding. I believe this is a red herring, however; for what we are debating is not whether reasoning about beliefs is *ever* cognitively demanding but whether it must *always* be so. In particular, I am not disputing that the belief-reasoning processes based in direct-FB system would be cognitively demanding; the question at hand is whether it is possible for the indirect-FB system to sustain belief-reasoning processes that are not so. Clearly, pointing to evidence suggesting that passing a *direct* false-belief tests requires executive functioning cannot resolve the issue.

Evans, 2003; Evans, 2008). This distinction seems to map quite well onto the processing profiles that Apperly and Butterfill propose for their two mentalising systems. For example, intuitive processes are typically taken to be fast, automatic, conscious, and efficient, while reflective processes are typically taken to be slow, controlled, conscious and cognitively demanding. Now, some dual-process theorists have argued that the defining feature of reflective versus intuitive processes is that only the former make constitutive use of working memory; notably, this is taken to explain why intuitive processes tend to be efficient (Carruthers, 2015; Evans & Stanovich, 2013; Stanovich & Toplak, 2012). If this type of account is on the right track, then it should remove much of the motivation behind Apperly and Butterfill's minimal theory of mind account.

2.2.3 Abduction

Apperly and Butterfill mention another feature of belief-reasoning which, they argue, is likely to lead to increased cognitive demands. This is based on the observation that reasoning about beliefs is likely to rely on abduction, and that abduction can be expected to be cognitively demanding.

Abductive reasoning involves inferring that a given proposition (or theory, or hypothesis etc.) is true on the grounds that it explains a given phenomenon better than the alternatives. Thus, for example, suppose you are watching a movie in the cosiness of your home, at night, when suddenly you hear a noise coming from upstairs -- a noise that sounds suspiciously like someone trying to break in. It must be a burglar! What else could make such a noise? Thus, you grab a knife from the kitchen and go investigate. Now, notice that the premise that someone or something made a suspicious noise does not *entail* the conclusion that a burglar is trying to break in, since, in principle, there is an infinite number of other things that could have caused that same noise; maybe a bird flew against a window, or maybe a piece of furniture fell down, or maybe you were spooked by the movie and simply imagined it. This makes your inference a *non-sequitur* when evaluated according to the standards of deductive logic. Even so, the fact that a burglar strikes you as the most plausible explanation does seem to provide you with some reason to believe it. This is an example of abductive reasoning.

An important feature of abductive reasoning is that it is not *monotonic*: while a set of premises S may support a conclusion C, there is no guarantee that a larger set of premises S+A, which includes S plus an additional set of premises A, will continue to support C. To return to our example:

if you knew that your adolescent daughter was secretly seeing someone, the possibility of the noise being caused by him or her trying to sneak into your daughter's room behind your back may seem more a more likely explanation than a burglar trying to break in, since, after all, the lights are on and it is not that late at night. Adding one premise to our original set has now led to a very different conclusion. This has a very important implication: the only way to make sure that a given hypothesis really provides the best explanation is to consider *everything* one knows. Of course, much of the information one considers may ultimately prove to be irrelevant, but there is no way of knowing this before considering it, and thus no way of restricting the set of premises that must be taken into account. As a result, abductive reasoning has been argued to be very hard to implement computationally, since conducting an exhaustive search over a large body of information is likely to result in a combinatorial explosion. And yet, abductive reasoning seems to be incredibly common, both in everyday life and in the sciences, which raises the question: how do we manage to do it³?

Apperly (2011) suggests two ways of making abductive processes tractable, both of which involve limiting the amount of information taken into account. The first involves making part of the information stored in the mind strictly off-limits to certain cognitive systems. For example, a system that is informationally encapsulated can only consult, in the course of its processing, the information that is stored within the module itself; and this will typically be but a small portion of all the information stored in the mind. This would be an example of what Apperly calls “hard constraint”. The second method, in contrast, involves using more efficient, non-exhaustive search algorithms that rely on heuristics to preselect the information that is worth considering. This will also result in only a subset of all the information stored in the mind being taken into consideration, but what and how much information is selected will vary depending on the task at hand. Thus, in Apperly's terms, this would count as a “soft constraint”. Now, Apperly claims that systems with hard constraints can be expected to be faster and more efficient but also severely limited in what they can do, while systems with soft constraints can be expected to be slower and more cognitively demanding, but also more flexible. Indeed, he suggests that the reason the fast system is fast and efficient, while the slow

³ Perhaps the most influential exposition of this problem is due to Jerry Fodor (1983). Fodor argued that the process of belief fixation relies on a process of inference to the best explanation; since inference to the best explanation cannot be modelled in computational terms, and since there is, within cognitive science, no viable alternative to the computational theory of mind, the obvious conclusion -- Fodor alleged -- is that cognitive science cannot explain belief fixation.

system is slow and cognitively demanding, is precisely that the first is informationally encapsulated, while the second is a central process that relies on scripts, norms and other heuristics to preselect information for further processing (Apperly, 2011, pp. 119-142).

Notably, Apperly seems to think that the fact that the fast system, being encapsulated, has no access to the information stored in the rest of the mind, is the reason the system can only ascribe registrations and not beliefs. This last step in the argument is important from our point of view: after all, the hybrid approach is consistent with the fast system being modular and thus limited in important ways, so long as these limitations do not prevent it from sustaining belief-reasoning abilities. Unfortunately, however, Apperly (2011) does not elaborate on why the modularity of the fast system should result in this particular type of limitation. Furthermore, notice that it is unclear *why* abductive processes that are only “softly” constrained should necessarily impose substantial demands on executive functioning.

2.3 Explaining Evidence of Signature Limits

Apperly and Butterfill (2009) argue that the fast system, since it employs a minimal theory of mind, should exhibit several “signature limits”. In particular, since the fast system reasons about registrations, which are not intensional states, it should not provide an understanding of *how* people represent the objects they represent. Consequently, Apperly and Butterfill argue that the fast system should be unable to sustain level-2 perspective-taking or tracking of beliefs about identity, both of which seems to presuppose an understanding not just of what objects agents represent but of how they represent them. There is evidence which seems to suggest that these signature limits do indeed occur. I will, argue, however, that this evidence is better explained on the hybrid approach.

2.3.1 Level-2 Perspective-Taking

In §2.1.1, we discussed a study by Samson et al. (2010) showing that participants automatically calculate what another person can see, which then interferes with their own visual perspective in an unrelated task, creating an altercentric effect. Apperly and Butterfill (2009, p. 963) suggest that how things look to other people, in contrast, should not be automatically calculated, since it involves ascribing intensional mental states, which is something the fast system cannot do. To test this prediction, Andrew Surtees, Stephen Butterfill and Ian Apperly (2011) showed their participants (a

group of six-year-olds and a group of adults) a picture of an avatar sitting at a table across from them, with a digit placed either on a wall on the side or lying flat on the table, and instructed them to judge (pressing one of two buttons) either what number they themselves could see, or what number the avatar could see. Crucially, in “ambiguous” trials, the digit was a 6 or a 9. Thus, when the digit was on the table as opposed to the side wall, participants would see a different number from the avatar (e.g., one would see a 6 and the other would see a 9.) This counts as level-2 perspective taking: participants must judge how something looks when seen from a different point of view. The crucial finding is that there was no significant difference in the time it took participants to say what number they saw in ambiguous trials compared to unambiguous trials. Thus, in this case, no altercentric effect was found, suggesting that participants did not calculate how the number looked to the avatar.

Surtees et al. (2011) take this finding to support Apperly and Butterfill’s account. This is because Apperly and Butterfill (2009) argue that level-2 perspective-taking requires an understanding of intensionality, which should be off-limits to the fast system. However, this reasoning appears to be based on conflating two senses in which an object can be said to be “represented as” something. Consider the following sentences:

- (1) The digit I am looking at is a 9;
- (2) The digit I am looking at is a 6;
- (3) The digit on the table is a 9.

Suppose the digit on the table and the digit I am looking at are one and the same. Then, (1) and (3) could be said to present the digit in different ways: in one case as “the digit I am looking at” and in the other as “the digit on the table”. These are different ways of referring, picking out or thinking about the digit. Nonetheless, both (1) and (3) predicate of the digit the property of being a 9. Thus, while in one sense (1) and (3) represent the digit in different ways (as the digit I am looking at v. the digit on the table) in another they represent it in the same way (as being a 9). The opposite could be said of (1) and (2): they present the digit in the same way but predicate a different property of it (being a 9 v. being a 6).

Why is this important? It is true that without an understanding of intensionality, one could not understand that different people may think of the same object under different modes of presentations; that is, after all, what intensionality is all about. On the other hand, understanding intensionality is not required to understand that different people predicate different properties of the same

object and thus, in this sense, represent it differently. Now, crucially, it seems that the latter would be sufficient for level-2 perspective taking. In our example, level-2 perspective taking requires understanding whether the avatar sees the digit as a 9 or 6, but the mode of presentation under which the avatar thinks of the digit (as the digit I am looking at or as the digit on the table) is immaterial. If this is right, then there is no reason to think that level-2 perspective-taking requires an understanding of intensionality. In our example, it seems it would suffice to ascribe to the agent a registration of the digit as (having the property of) being a “9” (say). In which case, the motivation Apperly and Butterfill provide for the fast system not being to sustain level-2 perspective-taking does not stand up to scrutiny¹⁹.

On the other hand, while at first pass the finding by Surtees et al. (2011) may seem inconsistent with the hybrid approach, in fact it isn't. A study by Surtees, Apperly, and Samson (2013) provides convincing evidence that level-2 perspective-taking requires mentally rotating the object in question. We have long known that the time it takes to mentally rotate an image depends on the angle of rotation from the starting position (Shepard & Metzler, 1971). That is, the further we need to mentally rotate an image, the longer it takes. Now, the study by Surtees et al. (2013) shows that the time it takes participants to judge what number an avatar can see varies as a function of the angular disparity between avatar and participant. This suggests that participants rotated the digit to see how it would look from the avatar perspective. Thus, the greater the angular disparity between participants and the avatar, the further participants had to rotate the digit, resulting in longer reaction times. Notably, Surtees and colleagues found the very same effect when participants were asked to judge whether the digit was on the avatar's left or right, another task that has been argued to require mental rotation (Michelon & Zacks, 2006). Now, mental rotation is known to rely on working

¹⁹ Now it is true, of course, that in Apperly and Butterfill's (2009) original account registrations could only take object-locations pairs as their content; properties other than location (like, for example, *being a “9”*) could not be registered. Thus, it may be objected, even if level-2 perspective taking does not presuppose an understanding of intensionality (as I have argued) it should still be off-limits for the fast system. Two considerations. First, Apperly and Butterfill themselves suggest that their account may be extended to allow for properties other than location to be registered (Apperly & Butterfill, 2009, p. 963; Butterfill & Apperly, 2013, p. 620). Second, several indirect false-belief studies show that infants can track beliefs about properties other than location (D. Buttelmann et al., 2014; F. Buttelmann et al., 2015; He et al., 2011; Scott & Baillargeon, 2009; Scott et al., 2010; Scott et al., 2015; Song & Baillargeon, 2008). Thus, Apperly and Butterfill seem to face a dilemma: unless they allow properties other than location to be registered, they cannot explain these indirect false-belief studies; on the other hand, if they allow properties other than location to be registered, they have problems explaining why Surtees et al. (2011) found no altercentric interference from level-2 perspective taking.

memory (e.g. Hyun & Luck, 2007) and I have argued in §2.2.2 that the fast system would not be able to use working memory as a workspace for its computations. Given that level-2 perspective-taking requires mental rotation and that mental rotation relies on working memory, we have good reason to assume that the fast system would not be able to do level-2 perspective taking, even though it may well possess the conceptual sophistication required to represent how things look to other people. The finding by Surtees et al. (2011) is thus easily accommodated on the hybrid approach.

2.3.2 False Beliefs About Identity

Another prediction made by Apperly and Butterfill is that the fast system should not be able to track beliefs about identity, since this seems to presuppose an understanding of intensionality. Two studies have found evidence consistent with this prediction (Low et al., 2014; Low & Watts, 2013). In the study by Jason Low and Joseph Watts, for example, three-year-olds, four-year-olds and adults were tested with a false-belief task that combined anticipatory-looking and direct paradigms, and which involved watching a video depicting a misidentification scenario. The apparatus used in the video included a table and two specially designed boxes, each placed on a different side (left/right) of the table. The boxes had openings on the sides and on the front (i.e., the side facing participants) that were covered by fringes. During the familiarisation trials, objects that were either red or blue in colour moved from one box to the other, and the agent always reached for the blue object. In the belief induction phase, a small toy robot moved from one box to the other; the side participants could see was red. After entering the second box, the robot came out from the side facing participants, where the agent could not see it, and rotated, revealing that it was blue on the other side. Then, the robot moved back into the box again and came out from the lateral opening, returning to the first box, this time showing its red side to the agent. At this point, a sound signalled that the agent was about to reach. After a pause, during which their anticipatory-looks were recorded, participants were asked the test question, e.g., which box will the agent look in? If participant understand that the agent, not having seen the robot rotate, is not aware that “the red robot” and “the blue robot” are the same, they should expect her to reach for the second box, since the agent saw a blue robot going in and a red robot coming out. If participants do not understand this, however, they should expect her to reach for the first box, since the other box is empty. Low and Watts found that while responses to the test question improved with age, the vast majority of the participants in

each age group made first-looks towards the first box. The authors take this as evidence that the fast system does not enable tracking of beliefs about identity, just like Apperly and Butterfill's two-systems account predicts.

However, the evidence on this issue is conflicting, for other indirect false-belief studies seem to show that infants can, in fact, track beliefs about identity (F. Buttelmann et al., 2015; Scott & Baillargeon, 2009; Scott et al., 2015; Song & Baillargeon, 2008). Apperly and Butterfill's account is consistent with the findings by Low and colleagues, but (as argued in Chapter II, §3.3) it cannot explain the results of these other indirect false-belief studies. In contrast, the hybrid approach has an easy explanation for this apparently conflicting pattern of results. As Carruthers (2016) notes, the tasks used by Low and colleagues, just like the task used by Surtees et al. (2011), are likely to require mental rotation. As participants first saw the red robot moving into the second box, they would have ascribed to the agent a belief that there was a red robot in the second box. When they realised the robot had a blue side and a red side, participants would then have to retrieve a memory of the red robot and mentally rotate it to figure out what it must have looked like (i.e., blue) to the agent as it first moved into the second box. As argued above, the fast system should not be able to rely on working memory to use mental rotation. On the other hand, the fast system should have no trouble ascribing beliefs about identity when this does not involve mental rotation; and indeed, the studies that have found evidence that infants can ascribe beliefs about identity do not require mental rotation to be passed. In the experiments infants have been shown to pass the false belief about identity typically results from the fact that the agent cannot tell two objects apart (since they looked the same), not from a misleading appearance as in the experiments by Low and colleagues (Low et al., 2014; Low & Watts, 2013). Thus, the evidence lines up with the prediction made by the hybrid approach, namely that the fast system should sustain ascription of false beliefs and other mental states so long as this does not require using working memory. Even in this case, then, the hybrid account offers a best explanation of the evidence.

2.4 The Argument from Higher-Order States

One of the supposed perks of having a fast mentalising system is that it can support real-time, fast-paced social interactions. There is reason to believe, however, that the fast system could not do this effectively unless it enabled ascription of higher-order states, that is, of representational

states that take other representational states as their content (e.g., “she *thinks* that I *want* her to *believe* that I am competent”). As we will see, this ability is likely to be particularly handy in two types of social interaction: those that involve groups of people, and those that involve communicative exchanges. The fact that the fast system would be able to ascribe higher-order states on the hybrid approach but on Apperly and Butterfill's own account should thus speak in favour of the former.

2.4.1 Communication

Neo-Gricean accounts of communication hold that inferring speaker meaning involves reasoning about the intentions and beliefs of the speaker, thus entertaining higher-order beliefs about those states. Relevance theory, for example, maintains that inferring speakers meaning involves ascribing to the speaker two intentions: first, an informative intention to induce a certain belief in the audience; and, second, a communicative intention that the audience be aware of the informative intention (Sperber, 2000). Suppose Mary is eating crackers and, noticing that Peter is looking at her, she expresses her satisfaction (“mmm..so good!”) as she looks back at him. What is being communicated? According to relevance theory, Mary will have an informative intention to affect Peter's beliefs (e.g., to make him believe that she likes crackers), plus a communicative intention that Peter believe that she has that informative intention. Thus, to infer what Mary meant to communicate, Peter would have to represent that Mary intends him to believe that she intends him to believe that she likes crackers.

While not all neo-Gricean theories require the audience to entertain fifth-order representations, many still require them to entertain higher-order representations. Thus, Robert Thompson's (2014a) account, despite being presented as a more modest alternative to relevance theory, still requires the audience to represent that the speaker intends them to believe that the speaker intends to produce a response in them.

Notably, this feature of neo-Gricean accounts has often been perceived as problematic. From a young age, children engage in conversation and seem able to understand, at least in some cases if not in others, what people tell them. In fact, several authors have argued that infants are already able to understand communicative intentions, which helps them in social and language learning (Bloom, 2000; Csibra, 2010; Tomasello, 2008). If the neo-Gricean picture is correct, however, understanding a communicative intention requires entertaining and manipulating very complex,

higher-order representations. This clashes with the evidence that five and six-year-olds struggle with direct second-order false-belief tasks (Perner & Wimmer, 1985), which require them to entertain a third-order representation (e.g., “John *thinks* Mary *thinks* the van is still at the old place”). Furthermore, holding in mind a third-order representation seems taxing even for adults; indeed, the study by McKinnon and Moscovitch (2007) shows that this places a significant cognitive load. And yet, communication is often swift and effortless. Faced with this problem, some neo-Griceans have suggested that communication may be sub-served by a dedicated mentalising module, employing algorithms that allow it to manipulate higher-order representation in an efficient manner (e.g. Sperber & Wilson, 2002).

Now, Apperly and Butterfill (2009, p. 964) are clear that one of the functions of the fast system would be that of supporting real-time communicative exchanges. On their account, however, registrations do not allow embedding, and thus cannot take other mental states as their content. Peter would be able to represent that Mary registers the crackers, but there is no way he could represent that Mary intends him to “register” that she intends him to register the crackers. As a result, Apperly and Butterfill’s fast system would be unable to carry out the type of computations that neo-Gricean accounts require. Thompson (2014b) argues that this speaks against Apperly and Butterfill’s account.

The hybrid approach solves this problem, however, since it maintains that the reason the fast system is efficient has to do with its being independent of the executive system. An implication of this model of cognitive demands is that the fast system may handle relatively complex representational structures (like higher-order states) without placing heavy demands on executive resources.

2.4.2 Group Interaction

José Luiz Bermudez (2003) discusses a “computational worry” concerning the fact that attempting to mentalise in a social group may result in a combinatorial explosion. Suppose there is a group of three people (A, B and C) each trying to predict what the others will do. To predict what B will do, A must first understand what B thinks, in particular what she thinks A and C will do. What B thinks C will do, however, depends on what B thinks C thinks, in particular what B thinks C thinks A and B will do; and of course, what C thinks A will do depends on what C thinks A thinks. Thus, even in a group of just three people, relying on mentalising to predict what the others will do

may lead to a rapid escalation to increasingly higher orders of meta-representation, of the form *B thinks that C thinks that A thinks that...* and so on. Of course, at any time, one may decide to hop off the carousel and simply take a guess. Nonetheless, Bermudez argues that, even waving the peril of an infinite regress, entertaining a higher-order representation may still result in significant cognitive demands. Since many social interactions, including those that involve more than two people, seem to unfold relatively quickly, Bermudez suggests that they may be sub-served not by mentalising but by some leaner form of social understanding.

On a two-systems account of mentalising, however, this looks like the type of problem the fast system is supposed to solve; and whether the fast system can ascribe beliefs as opposed to registrations seems to be an important factor in determining how well the problem can be solved. A may be able to ascribe to B a registration of the ball (say), but not a registration of C as registering the ball. If the fast system represents beliefs, instead, there would be no in-principle restriction on the orders of meta-representation that can be entertained.

2.4.3 Summary

Apperly and Butterfill (2009) argue that the fast system would sustain mentalising in the context of fast-paced social interactions. However, there is reason to think that the fast system would not be able to do so effectively for social interactions that involve a communicative exchange or more than two people – not unless it allowed ascription of higher-order states. Since the hybrid approach is consistent with this possibility, while Apperly and Butterfill's minimal theory of mind account is not, this provides yet another reason to opt for the hybrid approach.

3. Conclusions

In this chapter, I have introduced and defended the hybrid approach, a new way of dealing with the developmental puzzle of belief-reasoning. The hybrid approach is similar to the early-emergence approach in that it agrees that infants can ascribe false beliefs; however, its account of direct false-belief results is closer to that defended by late-emergentists. Despite the fact that no one has defended this approach before, it arguably provides a better account of the evidence compared to the late-emergence approach. The evidence that the hybrid approach can explain better includes not only the results of indirect false-belief tests (which, as we have seen in Chapter II, late-

emergentists have trouble explaining) but also the evidence that Apperly and Butterfill adduce in support of their very influential two-systems account (as argued in §2).

VI

SHOWDOWN: HYBRID APPROACH V. PROCESSING- TIME ACCOUNT

In chapter IV, I defended my own early-emergence account, called the processing-time account, and argued that it can explain not just the results of indirect false-belief tests but also the results of direct false-belief tests. In Chapter V, however, I introduced and defended a new way of solving the puzzle, i.e. the hybrid approach, and argued that it can also explain both sets of findings. The question we need to address now, then, is: which of the two is the most plausible, better supported option? The processing-time account or the hybrid approach?

The disagreement between these two types of account focuses on two issues: first, how many mentalising systems do humans possess? The processing-time account answers “one”, while the hybrid approach answers “two”; second, do direct and indirect false-belief tests recruit the same mentalising system, or distinct mentalising systems? The processing-time account answers “same system”, while the hybrid approach answers “distinct systems”. Evidence that relates to the first question will be discussed in §1, while evidence that relates to the second question will be discussed in §2.

1. One or Two Mentalising Systems?

1.1 Parsimony

One argument in favour of the processing-time account is that it is the most parsimonious option. This is because, while the hybrid approach posits two mentalising systems, the processing-time account posits only one. Since it is unreasonable to needlessly multiply cognitive systems, the processing-time account should be our first choice.

I think this is right so far as it goes; though, of course, parsimony considerations only apply when two accounts can explain all the data under consideration and are equivalent in respect to all their other theoretical virtues. Still, the argument from parsimony provides the processing-time account with a tactical advantage, making it the default choice.

1.2 Evidence for Two-Systems Accounts

In Chapter V, §2.1 we discussed several findings supporting the existence of two mentalising systems. These findings included: (i) conflicting evidence on whether mentalising is automatic and efficient as opposed to controlled and cognitively demanding; (ii) conflicting evidence on whether mentalising is dependent on language; (iii) evidence of mentalising in non-human animals and infants; (iv) conflicting evidence concerning the emergence of belief-reasoning abilities. One of the main questions we need to address, then, is whether the processing-time account can explain this evidence.

1.2.1 Development of Belief-Reasoning Abilities

We saw in Chapter III, §2, that most early-emergence accounts struggle to provide a convincing account of the results of direct false-belief tests; so long as this is true, the discrepancy in the results of direct and indirect false-belief tests provides evidence for the hybrid approach. However, in Chapter IV, I have argued that the processing-time account can explain the discrepancy

(§3). If this is correct, then it removes one important piece of evidence supporting the hybrid approach.

1.2.2 Language

I have argued in Chapter IV, §3.3, that the processing-time account can account for the relationship between language ability and direct false-belief performance in children on the hypothesis that language ability (i) leads to faster processing of linguistic stimuli, (ii) mediates the beneficial effect of experiential factors and (iii) facilitates reasoning about mental states in natural language (e.g., in episodes of inner speech). If I am right, then these data do not allow us to decide between the processing-time account and the hybrid approach, since both can explain them.

In Chapter V, §2.1.2, however, we also mentioned a study by Newton and de Villiers (2007) showing that a verbal shadowing task disrupted adults' performance on a concurrent, non-verbal false-belief task. This finding is puzzling for the processing-time account because the false-belief task used by Newton and de Villiers did not include any linguistic stimuli whatsoever, not even a direct question, and was thus not just non-verbal but indirect as well; on the processing-time account, language has essentially no role to play in this type of test.

Carruthers (2011, p. 253) provides a possible explanation. The mentalising system may have been recruited to interpret the speech in the verbal shadowing task, which may have then prevented it from correctly processing the behaviour of the agent in the false-belief video; in other words, the two tasks did not interfere because they both required language, but because they both required mentalising. The rhythmic tapping task that was used as a control, in contrast, required neither language or mentalising. In effect, the study may have confounded language ability and mentalising abilities.

1.2.3 Automaticity

What about the evidence that mentalising appears to be automatic and efficient in some studies but controlled and cognitively demanding in others? Whether early-emergence accounts (including the processing-time account) can explain these data depends on whether it is possible for one mentalising system to switch between different “modes” of operation, each with its own processing profile. Several authors have argued that there is no reason to believe the contrary

(Carruthers, 2016, 2017; Thompson, 2015). Carruthers (2016, pp. 158-159), for example, embraces a distinction between intuitive and reflective mentalising, where the former relies on a mentalising module operating autonomously while the latter relies on the mentalising module operating in conjunction with working memory, language and other executive resources (see also Carruthers, 2011, pp. 236-240). If this type of position is on the right track, then early-emergence accounts (including the processing-time account) may be able to explain why mentalising is automatic and efficient in some cases and controlled and cognitively demanding in others.

1.2.4 Infants and Non-Human Animals

If it is true that one mentalising system could switch between “automatic” and “controlled” modes of operation, then early-emergentists should have no trouble explaining evidence of mentalising in infants and non-human animals, for this evidence is only problematic if one assumes that mentalising necessarily requires language and executive functioning.

1.2.5 Summary

In conclusion, the evidence for two mentalising systems discussed in Chapter V, §2.1, was found to be inconclusive, since one system may be able to switch between different modes of operation, thus removing the need to posit two mentalising systems.

1.3 Misalignment of Processing Features

Let us consider, now, some evidence that supports Some researchers have pointed out that mentalising processes exhibit a processing profile that fits neither that of the fast system nor that of the slow system. This, these authors suggest, is better explained by positing just one mentalising system, which operates in different manners depending on the context. Thus, sometimes the system will operate in a fast and efficient manner, sometimes in a slow and cognitively demanding manner, and sometimes a mix of the two. In particular, Westra (2016b) argues that level-1 perspective-taking is an unencapsulated and spontaneous process, while level-2 perspective-taking is sometimes fast and sometimes slow; Carruthers (2016), instead, argues that the mentalising processes recruited in indirect false-belief tests are spontaneous. Both authors take this to speak against two-systems accounts of mentalising. Let us consider each of these challenges in turn.

1.3.1 Level-1 Perspective-Taking Is Unencapsulated

Westra (2016b) claims that evidence coming from studies of gaze-cuing provides evidence that level-1 perspective-taking is unencapsulated. In a typical gaze-cuing study, participants are shown a face, looking either towards the left corner or towards the right corner. Immediately after, a stimulus, for example a letter or a digit, appears either on the right corner or on the left corner. Participants are told to ignore the face and simply focus on detecting the stimulus by pressing a button. Nonetheless, studies employing this type of paradigm have found that participants are faster at detecting the stimulus when it appears in the direction the face was looking at. Crucially, however, several studies have found that this effect is sensitive to the background information participants possess. For example, participants are less likely to be cued by the gaze of a robot – unless they are told that an experimenter is controlling its gaze, that is (Wiese, Wykowska, Zwickel, & Müller, 2012). Similarly, participants are cued by ambiguous eye-like shapes when they are told that the shapes are eyes but not when they are told that they are wheels of a car (Ristic & Kingstone, 2005), or by the gaze of a face wearing goggles when they have worn transparent goggles but not when they have worn opaque goggles (Teufel, Alexis, Clayton, & Davis, 2010).

Why is this problematic for two-system accounts? After all, being encapsulated is not part of the processing profile of the indirect-FB system. Remember, however, that Apperly (2011) suggests the indirect-FB system may have the processing profile it has *because* it is encapsulated (as discussed in Chapter V, §2.2.3). Thus, Westra takes evidence that level-1 perspective-taking is not encapsulated to speak against Apperly's two-systems account. Notice however, that the evidence from gaze-cuing studies mentioned by Westra is consistent with indirect-FB processes being fast, automatic and efficient. Indeed, as Westra himself notes, gaze-cuing happens very rapidly (in the order of milliseconds), outside of conscious awareness, and with no apparent cognitive demands. Thus, even though the evidence does not support Apperly's suggestion concerning what makes indirect-FB processes faster and more efficient than direct-FB processes, it is entirely consistent with the hypothesis that there are two mentalising systems with different processing profiles. This evidence simply motivates two-systems theorists to look for an alternative model of how cognitive demands are generated. I put forward one such alternative in Chapter V, §2.2.2, where I argued that indirect-FB processes tend to be fast, efficient and automatic because they can run autonomously, without

relying on working memory or other executive resources. This is consistent with the claim that indirect-FB processes are not encapsulated.

In addition, the results mentioned by Westra seem open to a different interpretation. It is possible that the background information that was found to affect the cue-gazing affect was “attached” to the visual experiences the indirect-FB system received as inputs, thus affecting whether the system was triggered into action. When I tell you that an ambiguous stimulus represents a pair of eyes, this information is likely to affect your perception of the stimulus, in such a way that you will now *perceive* the ambiguous marks as eyes. This may be thought of as conceptual content being attached to a sensory, non-conceptual representation. We can visualise this by imagining that the non-conceptual representation is a picture, while the conceptual interpretation is the notes scribbled all over it. Now, there is good reason to expect that processes that operate post-perceptually will receive as input not just the picture but the notes as well, and that their activation will be dependent on whether the scribbles include any keywords pertaining to their domain of expertise. Thus, if the picture has “these are eyes!” written over it, with an arrow connecting “these” to the ambiguous marks, this may lead the indirect-FB system into engaging in level-1 perspective-taking, resulting in a cue-gazing effect.

To sum up, not only (i) the hybrid approach is consistent with the fast system not be encapsulated, but (ii) the evidence discussed by Westra (2016b) falls short of showing that the fast system is not encapsulated.

1.3.2 Level-2 Perspective-Taking Can Be Fast

Westra (2016b) also points to two studies which found an altercentric effect for level-2 perspective taking (Elekes, Varga, & Király, 2016; Surtees, Apperly, & Samson, 2016). In the study by Fruzsina Elekes, Máté Varga and Ildikó Király (2016), for example, each participant was sitting at a table opposite to another participant, with a screen lying flat on the table. In the perspective-dependent condition, both participants carried out a number-identification task, which consisted in judging (by pressing buttons) whether the number shown on the screen was the same they heard in audio recording. In the non-perspective-dependent condition, instead, one of the participants was given the number task while the other was given a colour task which consisted in judging (again by pressing buttons) whether the digit on the screen was the same colour as the one in the previous

trial. Elekes and colleagues found that, when the digit was ambiguous (e.g. 2, 5, 6 or 9), participants in the number task were slower to press the button, thus showing an altercentric effect. Importantly, this effect was only present in the perspective-dependent condition; this suggests that participants only track how objects look to people involved in the same task, which might explain why Surtees et al. (2011) found no interference effect (their study was discussed in Chapter V, §2.4.1). Westra (2016b) takes this type of finding to show that, when participants possess sufficient motivation to do so, they can quickly and efficiently calculate what an object looks like to another person; he then argues that this speaks against two-systems accounts since these accounts maintain that level-2 perspective-taking should be slow and cognitively demanding.

As Westra himself points out, however, these results are somewhat puzzling for early-emergentists as well. This is because, to explain why Surtees et al. (2011) found no altercentric effect from level-2 perspective-taking in their study, early-emergentists argue that level-2 perspective-taking requires mentally rotating the digit, which is cognitively demanding and relatively slow cognitive process. Given this, participants will not engage in mental rotation in the absence of a motivation to either predict or explain the behaviour of the other person (Carruthers, 2017; Westra, 2016b). This means, however, that early-emergence accounts essentially agree with two-systems accounts that level-2 perspective-taking should be a relatively slow and cognitively demanding process. Thus, evidence that level-2 perspective-taking is fast and efficient really speaks against both. To obviate this problem, Westra suggests that, when the participants in the studies by Elekes et al. (2016) and Surtees et al. (2016) were told the other person was going to be carrying out the same task, they memorised mentalising schemas (e.g., if I see 6 she sees 9) which they then retrieved during the task. Notice, however, that this solves the problem for the hybrid approach as well. We can divide level-2 perspective-taking in two phases: phase 1 consists in calculating how an object would look when seen from another perspective; phase 2 consists in ascribing that perspective to another person. Thus, in the case of the digit “6”, phase 1 one would involve calculating that the digit will look like a “9” when seen from across the table; phase 2 would involve ascribing this perspective to the person sitting across the table (“she sees a 9”). Now, on Westra's interpretation, phase 1 happened before the task; what happened during the task is phase 2. Thus, on his interpretation, these results do not show that level-2 perspective-taking can be fast and efficient; they simply show that phase 2 (ascribing an already calculated perspective) can be. Notice that it is phase

1 that involves mental rotation, not phase 2. Furthermore, remember that I argued (in Chapter V, §2.4.1) that the hybrid approach should maintain, following Carruthers (2017), that level-2 perspective-taking is cognitively demanding because it involves mental rotation. It follows that the hybrid account is not committed to the claim that phase 2 should be cognitively demanding, hence it is every bit as consistent with these results as the early-emergence accounts that Carruthers' and Westra's defend.

1.3.3 Indirect-FB Processes Are Spontaneous

According to Carruthers (2016), the study by Schneider, Lam, Bayliss, and Dux (2012) shows that the mentalising processes recruited by indirect false-belief tests are spontaneous and not automatic. (Automatic processes are those that are activated independently of any goals the subject may have; spontaneous processes are those whose activation is dependent on covert, unconscious goals.) Now, Schneider, Lam, et al. (2012) carried out dual-task experiment using an anticipatory-looking unexpected-transfer task as a primary task and a working memory task as a secondary task. The working memory task had a high-load and a low-load condition. In the low-load condition, participants simply listened to the audio recording of a voice reading a random sequence of letters. In the high-load condition, participants were instructed to count the number of 2-back repetitions (e.g. “N, A, N” is a 2-back repetition of the letter “N”). Schneider and colleagues found that the secondary task (both high-load and low-load) disrupted participants' anticipatory looking to the empty box. Specifically, with the secondary task, there was no significant difference in looking behaviour in the true-belief and false-belief conditions. This seems to show that mentalising in indirect false-belief tasks makes at least some cognitive demands. Carruthers (2016) suggests that, in this type of task, participants automatically calculate the false belief of the agent and then store it in long-term memory. When they hear the sound signalling the agent's return, participants then retrieve that information from long-term memory and use it to predict her action. He argues that this operation of retrieving from memory is likely to be caused by an unconscious standing goal to predict other people's behaviour, while also being dependent on executive resources. Under cognitive load, the covert goal is deactivated to save resources, thus resulting in no anticipatory looking. Carruthers takes this to show that, at least in this type of task, mentalising is neither automatic nor controlled, but spontaneous.

It is not clear how damaging this conclusion is for two-systems accounts, however, since it seems that spontaneous processes would still be closer to automatic processes than controlled ones in terms of their processing profile: they are fast, unconscious, and make relatively minimal cognitive demands. As discussed in Chapter V, §2.1, Apperly and Butterfill (2009) argue that the evidence that mentalising is sometimes automatic and sometimes controlled supports a two-systems account of mentalising. If what the evidence shows is that mentalising is sometimes controlled and sometimes non-controlled (i.e., either automatic or spontaneous), as Carruthers suggests, Apperly and Butterfill's argument still retains its force.

In any case, the result by Schneider and colleagues is also open to a different interpretation. Selective attention is often assumed to determine which mental contents are broadcast to the rest of the mind for further processing. It is thus reasonable to assume that the contents the indirect-FB system will receive as input are those targeted by attention. Now, the working memory task used clearly required participants to focus their attention on the auditory stimuli as opposed to the visual ones (especially so in the high-load condition). This means that the indirect-FB system will have received, as inputs, the auditory stimuli and not the visual ones; in which case it can be hardly blamed for not computing what the actor in the video was going to do²⁰. (Notice that for this explanation to work, it is not required to argue that the indirect-FB system did not process any visual stimuli; missing some of them, such as, for example, the change of location, will be sufficient to impair its ability to predict the agent's behaviour.)

1.3.4 Level-1 Perspective-Taking Is Spontaneous

Westra (2016b) argues that the results of the study by Qureshi et al. (2010), which we discussed in Chapter V, §2.1.1, also show that level-1 perspective-taking is spontaneous, that is, dependent on covert goals. Qureshi et al. (2010) found, remember, that the presence of a secondary, executive-taxing task did not erase the altercentric effect due to level-1 perspective-taking. The study also found, however, that reaction times in all conditions were higher when the secondary task was present; the authors took this to show that while perspective-calculation is efficient,

²⁰ But does this not mean that, in some sense, the indirect-FB system requires attention? No; it only means that attention plays a role in selecting the inputs the indirect-FB system receives. This is consistent with saying that the indirect-FB system does not require attention (or any other executive resource) to carry out its processing.

perspective-selection is not. Now, Westra argues that since one cannot carry out level-1 perspective-taking without carrying out perspective-selection, the results do show that level-1 perspective-taking is affected by cognitive load.

There is reason to resist Westra's interpretation, however. The task participants were assigned was that of judging whether the number of dots on the wall was the same as a previously shown digit. This task will require comparing a representation kept in working memory (the previously shown digit) with the number of dots one currently perceive, and then planning the response (pushing this or that button). In doing this, participants will also have to inhibit the interference from the avatar's perspective, which will require inhibitory control. Overall, these processes clearly require executive resources. Thus, it is not exactly surprising that people took longer to press the button when given a secondary, executively-taxing task!

1.3.5 Summary

The evidence that Westra (2016b) and Carruthers (2016) discuss falls short of establishing their claim – i.e., that mentalising processes sometimes exhibit a mixed processing profile. Thus, this objection against two-systems account can be dismissed.

2. Same System or Distinct Systems?

2.1 Is the Distinct-Systems View *Ad Hoc*?

Distinct-system accounts (including both distinct-systems late-emergence accounts and hybrid accounts) maintain that direct false-belief tests recruit the direct-FB system, which, in children under four, provides the wrong answer. One can wonder why this should be the case, however. In particular, one can wonder why two and three-year-olds do not recruit their indirect-FB system, since this would allow them to answer correctly. Even worse, one can argue that the claim that the indirect-FB system would not be recruited in direct-FB tests is just an ad hoc assumption. In other words: there seems to be no reason to believe that the indirect-FB system should not be recruited in direct false-belief tests, except that the assumption is needed to explain why young children fail those tests.

Carruthers (2016) raises this type of objection to Apperly and Butterfill's two-systems account. Carruthers takes Apperly and Butterfill to be committed to the claim that the outputs of the indirect-FB system would not be accessible to higher cognition. He argues, however, that there is evidence suggesting otherwise. Consider, for example, the interactive false-belief test by D. Buttelmann et al. (2009). In this study, after E2 moved the object and locked the boxes, E1 came back and struggled to open the box where she had left the toy. Most of the infants in this study tried to help E1 by pointing or crawling to the other box, the one that contained the toy she was looking for. Now, Carruthers argues that infants' behaviour in this type of test is likely to be under the control of the executive system, which means that the outputs of the indirect-FB system cannot be able to inform infants' behaviour unless they are accessible to their executive system. If they are, then we face the problem I was sketching above: why do these outputs not inform children's pointing in direct false-belief tests, just like they inform their crawling and pointing in the study by Buttelmann and colleagues? Perhaps, Carruthers suggests, the outputs of the indirect-FB system are accessible to the executive systems that control non-communicative actions but not to those that control verbal or communicative acts; but this, he argues, is ad hoc and implausible. Even worse, it threatens the consistency of the account, for Apperly and Butterfill hold that, in adults at least, the outputs of the indirect-FB system can sometimes affect reaction-times in verbal report. Clearly this interference could not take place were the outputs of the indirect-FB system inaccessible for planning of communicative acts.

Carruthers' objection seems rather compelling if we take the "inaccessibility route", arguing that the indirect-FB system is not recruited in indirect false-belief tests because it is inaccessible to the executive system. There may be other options, however. In particular, conceding that the outputs of the indirect-FB system must be accessible to the executive system (as Carruthers convincingly argues) does not entail that participants' responses will be based on those outputs. After all, the outputs of the direct-FB system, as well, must be accessible to the executive system. We can thus imagine the executive system as a real executive (i.e., the CEO of a company) who, after asking his technical experts to solve a problem, receives two conflicting responses: one expert (representing the direct-FB system) claims that they should point to the box, while the other (representing the indirect-FB system) argues that they should point to the basket. Clearly, the executive cannot listen to both. Furthermore, how is the executive supposed to know which suggestion is correct?

Thus, even if we concede that the executive system will receive the outputs of the indirect-FB system, there is really no reason to expect that it will select a response that is consistent with them.

Now, notice that the problem is not entirely solved yet, for we still need to explain why the executive system would listen to the indirect-FB system in interactive false-belief tasks (such as D. Buttelmann et al., 2009) but not in direct false-belief tasks. In the absence of a proper justification, it seems *ad hoc* to assume that this would be the case – which brings us back to our original problem. A possible solution appeals to the distinction between intuitive and reflective processes (Chapter V, §2.2.2). The suggestion is that the indirect-FB system would sustain “intuitive” forms of mentalising, whereas the direct-FB system would sustain mentalising of a more “reflective” kind. As mentioned in Chapter V, this hypothesis lines up well with the two processing profiles sketched by Apperly and Butterfill: intuitive processes are supposed to be fast, automatic, unconscious, and efficient, while reflective processes are supposed to be slow, controlled, conscious and cognitively demanding. Now, the solution may simply be that when infants see that the agent needs help, they rush to her aid, and are thus more inclined to accept their intuitions (provided by the indirect-FB system) without questioning them. In direct false-belief tests, in contrast, children may be more likely to engage in reflective thinking to check whether their intuitions are correct, which will result in the direct-FB system taking over.

Interestingly, if this type of response is endorsed, the hybrid approach makes precisely the opposite prediction compared to the processing-time account. The processing-time accounts predicts that children who are more impulsive, being more likely to answer before they can complete the processing required to answer correctly, should have worse direct false-belief performance. The hybrid approach, in contrast, predicts that children who are more impulsive, being more likely to base their answer on the intuitions provided by the indirect-FB system, should have a better chance of passing.

A study by Wendy Garnham and Perner (2001) seems to corroborate the prediction of the hybrid approach. Garnham and Perner included several types of tasks in their study, including both an interactive and a direct false-belief task, both using an unexpected-transfer type of scenario. In the interactive task, children had to place an obstacle to block the agent (a toy mouse) from reaching its intended location; the obstacle could be placed on the way to the empty location or on the way to the location containing the object (a piece of cheese). Garnham and Perner divided their

participants (a group of children aged two to four) in spontaneous-responders, who responded quickly, and prompted-responders, who took longer and had to be re-prompted. Spontaneous-responders performed better than prompted-responders both on the interactive task and on the direct task, just as the hybrid approach predicts²¹.

In conclusion, the objection that the distinct-system view is *ad hoc* can be dismissed.

2.2 Evidence for Performance Accounts

As mentioned in section Chapter IV, §2, several studies have shown that, in direct false-belief tests, young children can be brought above chance with relatively minor changes to the procedure. For example, Siegal and Beattie (1991) showed that including the adverb “first” in the test question of an explicit false belief test (“where will Jane look first for her kitten?”) was sufficient to bring three-year-olds from below-chance to above-chance. Since it shows that children can be made to pass direct false-belief tests by removing or attenuating the factors that are likely to negatively affect their performance, this type of finding provides evidence for early-emergence accounts (including, of course, the processing-time account). On the other hand, this evidence speaks directly against the hybrid approach, which maintains that the reason most three-year-olds have trouble passing direct false-belief tests is that the system typically recruited in those tests (the direct-FB system) does not yet allow them to ascribe false beliefs. If this were the case, then three-year-olds should continue to fail even once potential performance-hindering factors are removed.

How serious is this problem? On reflection, it seems that most of this evidence can be accommodated by making a relatively small revision to the hybrid approach. This is because most of the results in question have focused on children with a mean age of about three years and a half. As a result, this evidence is consistent with the possibility that the belief-reasoning abilities grounded in the direct false-belief system may emerge a few months earlier than originally assumed, by three years and a half instead of four, while being absent in the younger children. Children

²¹ This effect may be an experimental artefact, however. It appears from the description of the procedure (Garnham & Perner, 2001, p. 424) that the authors waited only one second before repeating the question. One second is not a very long time at all; instead of giving children who had not responded more time think, it seems the authors did precisely the opposite, pressing them with a second question shortly after the first one. This may very well have a disruptive effect, pushing children to hurry and answer before they are ready, thus increasing the likelihood they would resort to attribute-substitution (see Chapter 4, §1 for details).

younger than three-and-a-half would thus fail because their direct-FB system does not yet allow them to ascribe false beliefs, while children between three-and-a-half and four would fail because of performance difficulties and would thus benefit from the removal of performance-hindering factors.

In fact, somewhat paradoxically, some late-emergentists have argued that this evidence actually speaks against early-emergence approaches (Wellman et al., 2001). This is for two reasons. (a) First, Wellman and colleagues pointed out that none of the task manipulations they included in their meta-analysis interacted with age; that is, it was still the case that younger children performed worse than older ones. This, they argued, is not what early-emergence accounts predict; on those accounts, one would expect younger children to perform as well as their older peers after the factors that were hindering their performance were removed. (b) Second, Wellman and colleagues pointed out that there was no evidence that children younger than three could be made to pass direct false-belief tests; and this, arguably, not for lack of trying. This, they argued, shows that the difficulties of younger children are not just a reflection of performance-related factors but have a deeper, conceptual root.

Both arguments can be resisted, however. First, early-emergence accounts do not necessarily predict that removing the factors that hinder performance should help younger children more than older ones (e.g., Scholl & Leslie, 2001). Specifically, on the processing-time account, whether children pass or fail depends on two variables: their required processing time (RPT) and time available for processing (TAP). Children pass if their RPT is lower or equal to their TAP and fail otherwise. RPT can be expected to decrease as children grow older, while TAP can be expected to increase; thus, older children perform better than younger ones. Now, I argued in chapter IV, §2, that the task manipulations at hand help by decreasing children's RPT. Thus, for example, temporal markers like the one used by Siegal and Beattie (1991) help because, by clarifying the question, they reduce the time it takes children to process it. Notice, however, that this cannot be expected to improve performance in all children. This is because, to improve performance, RPT must be not only decreased but brought below threshold, and this may not always be the case. Children whose RPT is too far above threshold will not be helped, and we can expect more of the younger children to be in this type of situation, compared to the older children. Thus, it would be unreasonable to expect that any task manipulation could erase age-related differences in performance.

As for the second argument, while it is true that at the time Wellman and colleagues were writing no study had managed to make children younger than three pass a direct false-belief tests, there is now an exception: this is the study by Setoh et al. (2016) discussed in Chapter IV, §2.2.1. Their experiment shows that two-and-a-half-year-olds can pass an undisclosed-location task if given two response-generation practice trials. The importance of this result cannot be overstated: it seems to effectively erase the discrepancy between the results of direct false-belief and indirect false-belief tests. As a result, the hybrid approach would lose the main piece of evidence in its support: given that two years and a half is the youngest age at which children have ever been given a direct false-belief test, the fact that they pass means that there is no evidence that children ever fail direct false-belief tests for any reason other than performance difficulties.

It is, of course, never advisable to draw strong conclusions from a single study, since it may not replicate, and/or it may be open to alternative, low-level explanations. If the finding by Setoh et al. (2016) proves to be robust, however, it may well tilt the scales in favour of the processing-time account.

2.3 Evidence from Neuroscience

Hybrid and processing-time approaches make distinctive predictions concerning what cortical areas should “light up” in neuroimaging studies using indirect-FB tasks compared to those using direct-FB tasks. Specifically, the processing-time account (as other early-emergence accounts) predicts significant overlap between the areas activated in the two types of task, while the hybrid approach predicts little to no overlap. In addition, the processing-time accounts predicts that disruption (due, for example, to lesion) in the mentalising areas that underlie performance in direct false-belief tests should also result in impaired performance on indirect false-belief tests, while the hybrid approach makes the opposite prediction. Thus, neuroscience potentially provides a very straightforward way of testing these accounts.

Now, as mentioned in Chapter III, §1.7, neuroimaging studies show that a number of areas are differentially activated during direct false-belief tasks. These areas include mainly the Temporoparietal Junction (TPJ), the medial Pre-Frontal Cortex (mPFC), the Superior Temporal Sulcus (STS), and the Pre-Cuneus (PC). The TPJ, in particular, has been argued to be a core component of the mentalising network (Saxe & Kanwisher, 2003). Patients with lesions in the left TPJ, for

example, exhibit impaired performance on direct false-belief tests (Biervoeye, Dricot, Ivanoiu, & Samson, 2016; Samson, Apperly, Chiavarino, & Humphreys, 2004). Thus, the important question for us is whether any of these areas (and the TPJ in particular) would also be activated when passively watching a false-belief scenario unfold, as in typical indirect false-belief tasks. A few studies have started investigating this question, and their results seem consistent with the processing-time account but not with the hybrid approach (Hyde, Betancourt, & Simon, 2015; Kovács, Kühn, Gergely, Csibra, & Brass, 2014; Schneider, Slaughter, Becker, & Dux, 2014). Let us take a look at the evidence.

In the study by Ágnes Kovács and colleagues (2014), which used fMRI, subjects were shown movie clips based on a previous study (Kovács et al., 2010). In these movies, a ball first rolls behind an occluder, then either (a) nothing happens, (b) the ball rolls out again and out of sight, (c) the ball rolls out but then rolls back behind the occluder again. In the test phase, the occluder is lowered revealing either the ball or no ball. In the movies there is also an agent (a Smurf) watching the events. In some conditions the Smurf is present throughout and thus is aware of all the movements of the ball, while in others he is absent during some of the movements. In the original study by Kovács et al. (2010), adult participants were instructed to press a button as soon as they detected the ball; their finding was that what the Smurf believed, though irrelevant to the task, influenced participants' reaction times as much as their own beliefs. Thus, in the condition where both participants and the Smurf believed that there was a ball behind the occluder, participants were quicker to press the button as they saw the ball compared to a condition where only them, but not the Smurf, believed the ball to be behind the occluder. The authors thus found an altercentric effect, similar to that found by Samson et al. (2010) in their level-1 perspective-taking study (discussed in Chapter V, §2.1) which can be taken as evidence that participants automatically computed what the Smurf believed. Now, Kovács et al. (2014) showed the same videos to another group of participants, while they were laying in a MRI scanner; they found that both the right TPJ and the mPFC were activated, although the right TPJ more so when only the Smurf had a false belief that the ball was behind the occluder, and the mPFC more so when the Smurf had a true belief that the ball was behind the occluder.

The study by Daniel Hyde and colleagues (2015) found a similar result. Hyde and colleagues showed adult participants movie clips depicting an unexpected-transfer scenario (along the lines of

Southgate et al., 2007) and looked at activity in the right TPJ using a technique called *functional near-infrared spectroscopy* (fNIRS). There were three conditions: false-belief, true-belief and direct-perception. In the false-belief condition a puppet moved the object from one box to the other as the actor was looking away. In the true-belief condition, instead, the actor witnessed the change of location. Finally, the direct-perception condition was the same as the false-belief condition, except that the boxes were transparent, and thus the actor could see that the object had been moved as she looked back at the boxes after the transfer. Participants were told to pay attention to the movies but were given no specific task. Hyde and colleague found a peak of activity in the right TPJ in false-belief conditions but not in true-belief or direct-perception conditions. Intriguingly, this peak took place during the change of location, and thus precisely as the actor's belief became false.

The two studies just described suggest that the TPJ is involved specifically in representing false beliefs about the location of objects, and that it is recruited in indirect false-belief tasks as well, in line with what the processing-time account predicts. The study by Schneider, Slaughter, et al. (2014), found a slightly different result. Schneider and colleagues used fMRI to record neural activity as participants watched unexpected-transfer movies (based on Schneider et al. 2012) which included both true-belief and false-belief conditions. Participants were instructed to discriminate between high and low tones that played during the videos. Schneider and colleagues found that only the PC and left STS were significantly more active in false-belief versus true-belief conditions, while no significant difference was found for either the left or the right TPJs. Still, Schneider and colleagues did find that the TPJ was highly active compared to baseline. Thus, despite the inconsistency with the two results mentioned above, this study also seems to support the processing-time account. Remember that patients with a lesion in the TPJ have been found to have impaired direct false-belief performance; on the hybrid approach, this suggests that the TPJ implements the direct-FB system (or one of its components.) Furthermore, remember that, on the hybrid approach, the direct-FB system sustains mentalising processes that are controlled and conscious. Crucially, both Hyde and colleagues and Schneider and colleagues, after showing the videos, asked their participants questions designed to reveal whether they had engaged in conscious mentalising, and found no evidence they had done so. When these considerations are taken into account, the hybrid approach predicts that the TPJ should not be highly activated, either in the false-belief condition or in the true-belief condition, while Schneider and colleagues found that it was active in both.

Is there a way to make the hybrid approach compatible with these results? It seems the only way would be to argue that both the neural network that implements the indirect-FB system and that which implements the direct-FB systems are located in the TPJ. This hypothesis is very difficult to rule out. Notice, however, that in the absence of a reason to believe that this should be the case, this move is *ad hoc*, and thus weakens the account.

3. Conclusions

Some of the main objections against two-systems accounts (§1.3) and the distinct-systems view (§2.1) were found wanting. On the other hand, the processing-time account may be able to explain the evidence adduced for two-systems accounts (§1.2). In addition, evidence for performance accounts (§2.2; see also Chapter IV, §2) and evidence from neuroscience (§2.3), together with considerations of parsimony (§1.1), were all found to support the processing-time account. Thus, while none of these arguments are conclusive, the processing-time account comes out on top.

VII

CONCLUSIONS

So, can infants reason about beliefs? Surprisingly, the answer is “yes” – or at least: this is, in my opinion, what the evidence in our possess clearly suggests. First, as argued in Chapter II, that infants can ascribe false beliefs is the best explanation for the results of indirect false-belief tests. Second, as argued in Chapters III-V, evidence that is often taken to support the opposite conclusion can be accommodated in (at least) two ways.

The first, and most conventional option is to defend a performance account. In Chapter IV, I defended my own performance account, called the processing-time account, and argued that it can explain away practically all the evidence for late-emergence accounts. Since (as seen in Chapter III) other performance accounts have trouble accommodating these data, this provides strong support for the processing-time account.

The second option, which (to the best of my knowledge) has so far gone unnoticed, consists in defending (what I have called) the hybrid approach. As explained in chapter V, the hybrid approach consists in arguing that direct and indirect false-belief tests recruit distinct cognitive systems, each of which can independently sustain the ability to reason about beliefs, but which follow different developmental trajectories. The hybrid approach allows us to explain the results of indirect

false-belief tests similarly to early-emergentists (thus maintaining that infants can ascribe false beliefs) while explaining the results of direct false-belief tests similarly to late-emergentists.

To sum up, there are two ways of explaining the discrepancy in the results of direct and indirect false-belief tests: we can take either the same-system path or the distinct-systems path. In chapter IV (plus Chapter V, §1.2) I have argued that, among same-system accounts, the processing-time account is our best option. In chapter V I have argued that, among distinct-systems accounts, the hybrid approach is our best option. Since both types of account are committed to the claim that infants can ascribe false beliefs, this claim turns out to be strongly supported by the evidence.

REFERENCES

- Adrian, J. E., Clemente, R. A., & Villanueva, L. (2007). Mother's Use of Cognitive State Verbs in Picture-Book Reading and the Development of Children's Understanding of Mind: A Longitudinal Study. *Child Development, 78*(4), 1052-1067.
- Adrian, J. E., Clemente, R. A., Villanueva, L., & Rieffe, C. (2005). Parent-Child Picture-Book Reading, Mothers' Mental State Language and Children's Theory of Mind. *Journal of Child Language, 32*(3), 673-686.
- Andrews, K. (2005). Chimpanzee Theory of Mind: Looking in All the Wrong Places? *Mind & Language, 20*(5), 521-536.
- Apperly, I. A. (2011). *Mindreaders. The Cognitive Basis of "Theory of Mind"*: Psychology Press.
- 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.
- Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is Belief Reasoning Automatic? *Psychological Science, 17*(10), 841-844.
- Apperly, I. A., Samson, D., Carroll, N., Hussain, S., & Humphreys, G. W. (2006). Intact First- and Second-Order False Belief Reasoning in a Patient with Severely Impaired Grammar. *Social neuroscience, 1*(3-4), 334-348.
- Astington, J. W., & Jenkins, J. M. (1999). A Longitudinal Study of the Relation Between Language and Theory-Of-Mind Development. *Developmental psychology, 35*(5), 1311-1320.
- 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.
- Baddeley, A. D. (1986). *Working Memory*. Oxford: Clarendon
- Baillargeon, R., He, Z., Setoh, P., Scott, R. M., Sloane, S., & Yang, D. Y.-J. (2013). False-Belief Understanding and Why It Matters. the Social-Acting Hypothesis. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the Social World. What Infants, Children and Other Species Can Tell Us* (pp. 88-95). Oxford: Oxford University Press.
- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-Belief Understanding in Infants. *Trends in Cognitive Sciences, 14*(3), 110-118.
- Baillargeon, R., Scott, R. M., He, Z., Sloane, S., Setoh, P., Jin, K.-S., . . . Bian, L. (2015). Psychological and Sociomoral Reasoning in Infancy. In M. Mikulincer & P. R. Shaver (Eds.), *Apa Handbook of Personality and Social Psychology: Vol 1. Attitudes and Social Cognition* (pp. 79-150). Washington, DC: American Psychological Association.
- Barrouillet, P., Bernandin, S., & Camos, V. (2004). Time Constraints and Resource Sharing in Adults' Working Memory Spans. *Journal of Experimental Psychology: General, 133*(1), 83-100.
- Bartsch, K. (1996). Between Desires and Beliefs: Young Children's Action Predictions. *Child Development, 67*(4), 1671-1685.

- Benson, J. E., Sabbagh, M. A., Carlson, S. M., & Zelazo, P. D. (2013). Individual Differences in Executive Functioning Predict Preschoolers' Improvement from Theory-Of-Mind Training. *Developmental psychology, 49*(9), 1615-1627.
- Bermudez, J. L. (2003). The Domain of Folk Psychology. *Royal Institute of Philosophy Supplement, 53*, 25-48.
- Biervoeye, A., Dricot, L., Ivanoiu, A., & Samson, D. (2016). Impaired Spontaneous Belief Inference Following Acquired Damage to the Left Posterior Temporoparietal Junction. *Social Cognitive and Affective Neuroscience, 11*(10), 1513–1520.
- Block, N. (1995). On a Confusion About a Function of Consciousness. *Behavioral and Brain Sciences, 18*(2), 227-247.
- Bloom, P. (2000). *How Children Learn the Meanings of Words*. Cambridge: MIT Press.
- Borghesani, P. R., Madhyastha, T. M., Aylward, E. H., Reiter, M. A., Swarny, B. R., Warner Schaie, K., & Willis, S. L. (2013). The Association Between Higher Order Abilities, Processing Speed, and Age Are Variably Mediated by White Matter Integrity During Typical Aging. *Neuropsychologia, 51*(8), 1435–1444.
- Breheny, R. (2006). Communication and Folk Psychology. *Mind & Language, 21*(1), 74-107.
- Brooks, R., & Meltzoff, A. N. (2002). The Importance of Eyes: How Infants Interpret Adult Looking. *Developmental psychology, 38*(6), 958–966.
- Brooks, R., & Meltzoff, A. N. (2005). The Development of Gaze Following and Its Relation to Language. *Developmental science, 8*(6), 535-543.
- Brown, J. R., Donelan-McCall, N., & Dunn, J. (1996). Why Talk about Mental States? The Significance of Children's Conversations with Friends, Siblings, and Mothers. *Child Development, 67*(3), 836-849.
- Buckner, C. (2014). The Semantic Problem(s) with Research on Animal Mindreading. *Mind & Language, 29*(5), 566–589.
- 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.
- Burgmans, S., Gronenschild, E. H. B. M., Fandakova, Y., Shing, Y. L., van Boxtel, M. P. J., Vuurman, E. F. P. M., . . . Raz, N. (2011). Age Differences in Speed of Processing Are Partially Mediated by Differences in Axonal Integrity. *NeuroImage, 55*(3), 1287-1297.
- Buttelmann, D., & Buttelmann, F. (2017). The Influence of a Bystander Agent's Beliefs on Children's and Adults' Decision-Making Process. *Journal of Experimental Child Psychology, 153*, 126-139.
- Buttelmann, D., Buttelmann, F., Carpenter, M., Call, J., & Tomasello, M. (2017). Great Apes Distinguish True from False Beliefs in an Interactive Helping Task. *PLOS one, 12*(4), e0173793.
- Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-Month-Old Infants Show False Belief Understanding in an Active Helping Paradigm. *Cognition, 112*(2), 337-342.
- Buttelmann, D., Over, H., Carpenter, M., & Tomasello, M. (2014). Eighteen-Month-Olds Understand False Beliefs in an Unexpected-Contents Task. *Journal of Experimental Child Psychology, 119*(1), 120-126.
- Buttelmann, F., Suhrke, J., & Buttelmann, D. (2015). What You Get Is What You Believe: Eighteen-Month-Olds Demonstrate Belief Understanding in an Unexpected-Identity Task. *Journal of Experimental Child Psychology, 131*, 94–103.
- Butterfill, S. A., & Apperly, I. A. (2013). How to Construct a Minimal Theory of Mind. *Mind & Language, 28*(5), 606-637.

- Call, J., & Tomasello, M. (2008). Does the Chimpanzee Have a Theory of Mind? 30 Years Later. *Trends in Cognitive Sciences*, 12(5), 187-192.
- Carlson, S. M., Claxton, L. J., & Moses, L. J. (2015). The Relation Between Executive Function and Theory of Mind is More Than Skin Deep. *Journal of Cognition and Development*, 16(1), 186-197.
- Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive Function and Theory of Mind: Stability and Prediction From Ages 2 to 3. *Developmental psychology*, 40(6), 1105-1122.
- Carlson, S. M., & Moses, L. J. (2001). Individual Differences in Inhibitory Control and Children's Theory of Mind. *Child Development*, 72(4), 1032-1053.
- Carlson, S. M., Moses, L. J., & Breton, C. (2002). How Specific Is the Relation Between Executive Function and Theory of Mind? Contributions of Inhibitory Control and Working Memory. *Infant and Child Development*, 11(2), 73-92.
- 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, 299-319.
- Caron, A. J., Kiel, E. J., Dayton, M., & Butler, S. C. (2002). Comprehension of the Referential Intent of Looking and Pointing Between 12 and 15 Months. *British Journal of Developmental Psychology*, 3(4), 445-464.
- Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve- and 18-Month-Olds Copy Actions in Terms of Goals. *Developmental science*, 8(1), F13-F20.
- Carruthers, P. (2000). *Phenomenal Consciousness: a Naturalistic Theory*. Cambridge: Cambridge University Press.
- Carruthers, P. (2011). *The opacity of mind: An integrative theory of self-knowledge*. Oxford: Oxford University Press.
- Carruthers, P. (2013). Mindreading in Infancy. *Mind & Language*, 28(2), 141-172.
- Carruthers, P. (2014). On Central Cognition. *Philosophical Studies*, 170(1), 143-162
- Carruthers, P. (2015). *The Centered Mind: What the Science of Working Memory Shows Us About the Nature of Human Thought*. New York, NY: Oxford University Press.
- Carruthers, P. (2016). Two Systems for Mindreading? *Review of Philosophy and Psychology*, 6.
- Carruthers, P. (2017). Mindreading in Adults: Evaluating Two-Systems Views. *Synthese*, 194(3), 673-688.
- Carruthers, P., Laurence, S., & Stich, S. P. (2007). Introduction. In P. Carruthers, S. Laurence, & S. P. Stich (Eds.), *The Innate Mind, Volume 3: Foundations and the Future*. New York; Oxford Oxford University Press.
- Cassidy, K. W. (1998). Three- and Four-Year-Old Children's Ability to Use Desire- and Belief- Based Reasoning. *Cognition*, 66(1), B1-B11.
- Cassidy, K. W., Cosetti, M., Jones, R., Kelton, E., Meier Rafal, V., Richman, L., & Stanhaus, H. (1995). Preschool Children's Understanding of Conflicting Desires. *Journal of Cognition and Development*, 6(3), 427-454.
- Cassidy, K. W., Werner, R. S., Rourke, M., Zubernis, L. S., & Balaraman, G. (2003). The Relationship Between Psychological Understanding and Positive Social Behaviors. *Social Development*, 12(2), 198-221.
- Cerella, J., & Hale, S. (1994). The Rise and Fall in Information-Processing Rates over the Life Span. *Acta Psychologica*, 86(2), 109-197.
- Chalmers, D. J. (1995). Facing up to the Problem of Consciousness. *Journal of Consciousness Studies*, 2(3), 200-219.

- Chomsky, N. (1959). Verbal Behaviour by B.F. Skinner. *Language*, 35(1), 26-58.
- Chopra, S., Shaw, M., Shaw, T., Sachdev, P. S., Anstey, K. J., & Cherbuin, N. (2018). More Highly Myelinated White Matter Tracts Are Associated with Faster Processing Speed in Healthy Adults. *NeuroImage*, 171, 332-340
- Clatterbuck, H. (2015). Chimpanzee Mindreading and the Value of Parsimonious Mental Models. *Mind & Language*, 30(4), 414-436.
- Clements, W. A., & Perner, J. (1994). Implicit Understanding of Belief. *Cognitive Development*, 9(4), 377-397.
- Clements, W. A., Rustin, C. L., & McCallum, S. (2000). Promoting the Transition from Implicit to Explicit Understanding: A Training Study of False Belief. *Developmental science*, 3(1), 81-92.
- Cowan, N. (2000). The Magical Mystery Four: How is Working Memory Capacity Limited, and Why? *Current Directions in Psychological Science*, 19(1), 51-57.
- Cowie, F. (1999). *What's Within? : Nativism Reconsidered*. New York: Oxford University Press.
- Csibra, G. (2008). Goal Attribution to Inanimate Agents by 6.5-Month-Old Infants. *Cognition*, 107(2), 705-717.
- Csibra, G. (2010). Recognizing Communicative Intentions in Infancy. *Mind & Language*, 25(2), 141-168.
- Csibra, G., Biró, S., Koós, O., & Gergely, G. (2003). One-Year-Old Infants Use Teleological Representations of Actions Productively. *Cognitive Science*, 27(1), 111-133.
- Csibra, G., & Gergely, G. (1998). The Teleological Origins of Mentalistic Action Explanations: a Developmental Hypothesis. *Developmental science*, 1(2), 255-259.
- Csibra, G., Gergely, G., Biró, S., Koós, O., & Brockbank, M. (1999). Goal Attribution Without Agency Cues: the Perception of 'Pure Reason' in Infancy. *Cognition*, 72(3), 237-267.
- D'Entremont, B., Hains, S. M. J., & Muir, D. W. (1997). A Demonstration of Gaze Following in 3- to 6-Month-Olds. *Infant Behavior and Development*, 20(4), 569-572.
- Davis, H. L., & Pratt, C. (1995). The Development of Children's Theory of Mind: the Working Memory Explanation. *Australian Journal of Psychology*, 47(1), 25-31.
- De Bruin, L., & Newen, A. (2012). An Association Account of False Belief Understanding. *Cognition*, 23(2), 240-259.
- De Bruin, L., & Newen, A. (2014). The Developmental Paradox of False Belief Understanding: a Dual-System Solution. *Synthese*, 191(3), 297-320.
- de Bruin, L., Strijbos, D., & Slors, M. (2011). Early Social Cognition: Alternatives to Implicit Mindreading. *Review of Philosophy and Psychology*, 2(3), 499-517.
- de Villiers, J. G. (2005). Can Language Acquisition Give Children a Point of View? In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 186-219). Oxford: Oxford University Press.
- de Villiers, J. G., & Pyers, J. E. (2002). Complements to Cognition: a Longitudinal Study of the Relationship Between Complex Syntax and False-Belief-Understanding. *Cognitive Development*, 17, 1037-1060.
- de Villiers, P. A. (2005). The Role of Language in Theory-Of-Mind Development: What Deaf Children Tell Us. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 266-297). Oxford: Oxford University Press.
- de Villiers, P. A., & de Villiers, J. G. (2012). Deception Dissociates from False Belief Reasoning in Deaf Children: Implications for the Implicit Versus Explicit Theory of Mind Distinction. *British Journal of Developmental Psychology*, 30(1), 188-209.

- Devine, R. T., & Hughes, C. (2014). Relations Between False Belief Understanding and Executive Function in Early Childhood: A Meta-Analysis. *Child Development, 85*(5), 1777-1994.
- Diamond, A. (2013). Executive Functions. *Annual Review of Psychology, 64*, 135–168.
- Ding, X. P., Wellman, H. M., Wang, Y., Fu, G., & Lee, K. (2015). Theory-of-Mind Training Causes Honest Young Children to Lie. *Psychological Science, 26*(11), 1812–1821.
- Doenya, C., Yavuz, H. M., & Selcuk, B. (2018). Not Just a Sum of Its Parts: How Tasks of the Theory of Mind Scale Relate to Executive Function Across Time. *Journal of Experimental Child Psychology, 166*, 485-501.
- Drayton, L. A., & Santos, L. R. (2016). A Decade of Theory of Mind Research on Cayo Santiago: Insights into Rhesus Macaque Social Cognition. *American Journal of Primatology, 78*, 106–116.
- Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA; London: MIT Press
- Duh, S., Paik, J. H., Miller, P. H., Gluck, S. C., Li, H., & Himelfarb, I. (2016). Theory of Mind and Executive Function in Chinese Preschool Children. *Developmental psychology, 52*(4), 582-591.
- Elekes, F., Varga, M., & Király, I. (2016). Evidence for Spontaneous Level-2 Perspective Taking in Adults. *Consciousness and Cognition, 41*(1), 93-103.
- Emery, N. J., & Clayton, N. S. (2009). Comparative Social Cognition. *Annual Review of Psychology, 60*, 87-113.
- Evans, J. S. B. T. (2003). In Two Minds: Dual-Process Accounts of Reasoning. *Trends in Cognitive Sciences, 7*(10), 454-459.
- Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology, 59*, 255-278.
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science, 8*(3), 223-241.
- Fink, E., Begeer, S., Peterson, C. C., Slaughter, V., & De Rosnay, M. (2015). Friendlessness and Theory of Mind: a Prospective Longitudinal Study. *British Journal of Developmental Psychology, 33*(1), 1–17.
- Fitzpatrick, S. (2009). The Primate Mindreading Controversy: A Case Study in Simplicity and Methodology in Animal Psychology. In R. W. Lurz (Ed.), *The philosophy of animal minds* (Vol. 224-246). New York: Cambridge University Press.
- 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.
- Flavell, J. H., Flavell, E. R., & Green, F. L. (1986). Development of the Appearance-Reality Distinction. *Cognitive Psychology, 15*(1), 95-120.
- Fodor, J. A. (1983). *The Modularity of Mind: an Essay on Faculty Psychology*. Cambridge, MA: MIT Press.
- Friedman, O., & Leslie, A. M. (2004). Mechanisms of Belief-Desire Reasoning. Inhibition and Bias. *Psychological Science, 15*(8), 547-552.
- Frith, U., & Frith, C. D. (2003). Development and Neurophysiology of Mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1431), 459-473.
- Fry, A. F., & Hale, S. (1996). Processing Speed, Working Memory, and Fluid Intelligence: Evidence for a Developmental Cascade. *Psychological Science, 7*(4), 237-241.
- Gallagher, S., & Povinelli, D. J. (2012). Enactive and Behavioral Abstraction Accounts of Social Understanding in Chimpanzees, Infants, and Adults. *Review of Philosophy and Psychology, 3*(1), 145-169.

- Garnham, W. A., & Perner, J. (2001). Actions Really Do Speak Louder Than Words—but Only Implicitly: Young Children’s Understanding of False Belief in Action. *British Journal of Developmental Psychology*, *19*(3), 413-432.
- Garnham, W. A., & Ruffman, T. (2001). Doesn’t See, Doesn’t Know: Is Anticipatory Looking Really Related to Understanding of Belief? *Developmental science*, *4*(1), 94-100.
- Gergely, G., Bekkering, H., & Király, I. (2002). Rational Imitation in Preverbal Infants. *Nature*, *415*(6873), 755.
- Gergely, G., Nádasdy, Z., & Csibra, G. (1995). Taking the Intentional Stance at 12 Months of Age. *Cognition*, *56*(2), 165-193.
- 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.
- Gopnik, A., & Astington, J. W. (1988). Children's Understanding of Representational Change and Its Relation to the Understanding of False Belief and the Appearance-Reality Distinction. *Child Development*, *59*(1), 26-37.
- Gopnik, A., & Wellman, H. M. (1992). Why the Child's Theory of Mind Really Is a Theory. *Mind & Language*, *7*(1-2), 145–171.
- Gopnik, A., & Wellman, H. M. (1994). The Theory Theory. In L. A. Hirschfield & S. A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 257-293). New York: Cambridge University Press.
- Gordon, A. C. L., & Olson, D. R. (1998). The Relation Between Acquisition of a Theory of Mind and the Capacity to Hold in Mind. *Journal of Experimental Child Psychology*, *68*(1), 70-83.
- Grice, P. H. (1957). Meaning. *The Philosophical Review*, *66*(3), 377-388.
- Guajardo, N. R., & Watson, A. C. (2002). Narrative Discourse and Theory of Mind Development. *The Journal of Genetic Psychology*, *163*(3), 305-325.
- Hadwin, J., & Perner, J. (1991). Pleased and Surprised: Children's Cognitive Theory of Emotion. *British Journal of Developmental Psychology*, *9*(2), 215–234.
- Hale, C., & Tager-Flusberg, H. (2003). The Influence of Language on Theory of Mind: a Training Study. *Developmental science*, *6*(3), 346-359.
- Hale, S. (1990). A Global Developmental Trend in Cognitive Processing Speed. *Child Development*, *61*(3), 653-663
- Hale, S., & Jansen, J. (1994). Global Processing-Time Coefficients Characterize Individual and Group Differences in Cognitive Speed. *Psychological Science*, *5*(6), 384-389.
- Hamlin, J. K., Hallinan, E. V., & Woodward, A. L. (2008). Do as I Do: 7-Month-Old Infants Selectively Reproduce Others’ Goals. *Developmental science*, *11*(4), 487-494.
- Hansen, M. B. (2010). If You Know Something, Say Something: Young Children’s Problem with False Beliefs. *Frontiers in Psychology*, *1*.
- 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.
- He, Z., Bolz, M., & Baillargeon, R. (2012). 2.5-Year-Olds Succeed at a Verbal Anticipatory-Looking False-Belief Task. *British Journal of Developmental Psychology*, *30*(1), 14-29.
- Hedger, J. A., & Fabricius, W. V. (2011). True Belief Belies False Belief: Recent Findings of Competence in Infants and Limitations in 5-Year-Olds, and Implications for Theory of Mind Development. *Review of Philosophy and Psychology*, *2*(3), 429-447.

- Helming, K. A., Strickland, B., & Jacob, P. (2014). Making Sense of Early False-Belief Understanding. *Trends in Cognitive Sciences*, 18(4), 167-170.
- Helming, K. A., Strickland, B., & Jacob, P. (2016). Solving the Puzzle About Early Belief-Ascription. *Mind & Language*, 31(4), 438-469.
- Heyes, C. M. (2014). False Belief in Infancy: a Fresh Look. *Developmental science*, 17(5), 647-659
- Heyes, C. M., & Frith, C. D. (2014). The Cultural Evolution of Mind Reading. *Science*, 344(6190), 1243091.
- Hofmann, S. G., Doan, S. N., Sprung, M., Wilson, A., Ebesutani, C., Andrews, L. A., . . . Harris, P. L. (2016). Training Children's Theory-Of-Mind: a Meta-Analysis of Controlled Studies. *Cognition*, 150, 200-212.
- Hughes, C. (2005). Origins of Individual Differences in Theory of Mind: From Nature to Nurture? *Child Development*, 76(2), 356-370.
- Hughes, C., & Ensor, R. (2005). Executive Function and Theory of Mind in 2 Year Olds: A Family Affair? *Developmental Neuropsychology*, 298(2), 645-668
- Hyde, D. C., Betancourt, M. A., & Simon, C. E. (2015). Human Temporal-Parietal Junction Spontaneously Tracks Others' Beliefs: A Functional Near-Infrared Spectroscopy Study. *Human Brain Mapping*, 36(12), 4831-4846.
- Hyun, J.-S., & Luck, S. (2007). Visual Working Memory as the Substrate for Mental Rotation. *Psychonomic Bulletin & Review*, 14(1), 154-158.
- Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., & Ruffman, T. (2016). Theory of Mind and Prosocial Behavior in Childhood: A Meta-Analytic Review. *Developmental psychology*, 52(8), 1192-1205.
- Jenkins, J. M., Turrell, S. L., Kogushi, Y., Lollis, S., & Ross, H. S. (2003). A Longitudinal Investigation of the Dynamics of Mental State Talk in Families. *Child Development*, 74(3), 905-920.
- Joseph, R. M., & Tager-Flusberg, H. (1999). Preschool Children's Understanding of the Desire and Knowledge Constraints on Intended Action. *British Journal of Developmental Psychology*, 17(2), 221-243.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kahneman, D., & Frederick, S. (2002). Representativeness Revisited: Attribute Substitution in Intuitive Judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: the Psychology of Intuitive Judgment* (pp. 49-81). New York: Cambridge University Press.
- Kail, R. V. (1991). Developmental Change in Speed of Processing During Childhood and Adolescence. *Psychological Bulletin*, 109(3), 490-501.
- Kail, R. V. (2007). Longitudinal Evidence That Increases in Processing Speed and Working Memory Enhance Children's Reasoning. *Psychological Science*, 18(4), 312-313.
- Kail, R. V., & Salthouse, T. A. (1994). Processing Speed as a Mental Capacity. *Acta Psychologica*, 86(2), 199-225.
- Kampis, D., Parise, E., Csibra, G., & Kovács, Á. M. (2015). Neural Signatures for Sustaining Object Representations Attributed to Others in Preverbal Human Infants. *Proceedings of the Royal Society B*, 282(20151683).
- Karbasforoushan, H., Duffy, B., Blackford, J. U., & Woodward, N. D. (2015). Processing Speed Impairment in Schizophrenia Is Mediated by White Matter Integrity. *Psychological Medicine*, 45(1), 109-120.
- Kaufman, J., Csibra, G., & Johnson, M. H. (2003). Representing Occluded Objects in the Human Infant Brain. *Proceedings of the Royal Society B: Biological Sciences*, 270, S140-S143.

- Keenan, J. P., Olson, D. R., & Marini, Z. (1998). Working Memory and Children's Developing Understanding of Mind. *Australian Journal of Psychology*, *50*(2), 76-82.
- Kerchner, G. A., Racine, C. A., Hale, S., Wilhelm, R., Laluz, V., Miller, B. L., & Kramer, J. H. (2012). Cognitive Processing Speed in Older Adults: Relationship with White Matter Integrity. *PLOS one*, *7*(11), e50425.
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on Theory of Mind Use in Adults. *Cognition*, *89*(1), 25-41.
- Kikuno, H., Mitchell, P., & Ziegler, F. (2007). How Do Young Children Process Beliefs About Beliefs? Evidence from Response Latency. *Mind & Language*, *22*(3), 297-316.
- Knobe, J. (2003). Intentional Action and Side Effects in Ordinary Language. *Analysis*, *63*(3), 190-194.
- Knudsen, B., & Liszkowski, U. (2012a). 18-Month-Olds Predict Specific Action Mistakes Through Attribution of False Belief, Not Ignorance, and Intervene Accordingly. *Infancy*, *17*(6), 672-691.
- Knudsen, B., & Liszkowski, U. (2012b). Eighteen- and 24-Month-Old Infants Correct Others in Anticipation of Action Mistakes. *Developmental science*, *15*(1), 113-122.
- Kochunov, P., Coyle, T., Lancaster, J., Robin, D. A., Hardies, J., Kochunov, V., . . . Fox, P. T. (2010). Processing Speed Is Correlated with Cerebral Health Markers in the Frontal Lobes as Quantified by Neuroimaging. *NeuroImage*, *49*(2), 1190-1199.
- Kochunov, P., Rowland, L. M., Fieremans, E., Veraart, J., Jahanshad, N., Eskandar, G., . . . Hong, L. E. (2016). Diffusion-Weighted Imaging Uncovers Likely Sources of Processing-Speed Deficits in Schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(47), 13504-13509.
- Kolodziejczyk, A. M., & Bosacki, S. L. (2015). Children's Understandings of Characters' Beliefs in Persuasive Arguments: Links with Gender and Theory of Mind. *Early Child Development and Care*, *185* (4), 562-577.
- Koster-Hale, J., & Saxe, R. (2013). Functional neuroimaging of theory of mind In S. Baron-Cohen, M. V. Lombardo, & T.-F. Helen (Eds.), *Understanding Other Minds. Perspectives from developmental social neuroscience* (3rd ed., pp. 520). Oxford: Oxford University Press.
- 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).
- Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults. *Science*, *330*(6012), 1830-1834.
- Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great Apes Anticipate That Other Individuals Will Act According to False Beliefs. *Science*, *354*(6308), 110-114.
- Lecce, S., Bianco, F., Demicheli, P., & Cavallini, E. (2014). Training Preschoolers on First-Order False Belief Understanding: Transfer on Advanced ToM Skills and Metamemory. *Child Development*, *85*(6), 2404-2418.
- Leslie, A. M. (1987). Pretense and Representation: the Origins of "Theory of Mind". *Psychological Review*, *94*(4), 412-426.
- Leslie, A. M. (2000a). How to Acquire a Representational Theory of Mind. In D. Sperber (Ed.), *Metarepresentations. A multidisciplinary perspective* (pp. 197-223). New York: Oxford University Press.
- Leslie, A. M. (2000b). "Theory of Mind" as a Mechanism of Selective Attention. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 1235-1247). Cambridge, MA: MIT Press.
- Leslie, A. M., Friedman, O., & German, T. P. (2004). Core Mechanisms in "Theory of Mind". *Trends in Cognitive Sciences*, *8*(12), 528-533.

- Leslie, A. M., Knobe, J., & Adam, C. (2006). Acting Intentionally and the Side-Effect Effect: Theory of Mind and Moral Judgment. *Psychological Science*, *17*(5), 412-427.
- Leslie, A. M., & Polizzi, P. (1998). Inhibitory Processing in the False Belief Task: Two Conjectures. *Developmental science*, *1998*(1), 247-253.
- Leslie, A. M., & Thaiss, L. (1991). Domain Specificity in Conceptual Development: Neuropsychological Evidence from Autism. *Cognition*, *43*(3), 225-251.
- Levine, J. (1983). Materialism and Qualia: the Explanatory Gap. *Pacific Philosophical Quarterly*, *64*(4), 354-361.
- Lewis, C., Freeman, N. H., Kyriakidou, C., Maridaki-Kassotaki, K., & Berridge, D. M. (1996). Social Influences on False Belief Access: Specific Sibling Influences or General Apprenticeship? *Child Development*, *67*(6), 2930-2947.
- Lewis, C., & Osborne, A. (1990). Three-Year-Olds' Problems with False Belief: Conceptual Deficit or Linguistic Artifact? *Child Development*, *61*(5), 1514-1519.
- Lillard, A. S., & Flavell, J. H. (1992). Young Children's Understanding of Different Mental States. *Developmental psychology*, *28*(4), 626-634.
- Lin, S., Keysar, B., & Epley, N. (2010). Reflexively Mindblind: Using Theory of Mind to Interpret Behavior Requires Effortful Attention. *Journal of Experimental Social Psychology*, *46*(3), 551-556.
- Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- and 18-Month-Olds Point to Provide Information for Others. *Journal of Cognition and Development*, *7*(2), 173-187.
- Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). Pointing out New News, Old News, and Absent Referents at 12 Months of Age. *Developmental science*, *10*(2), F1-F7.
- Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-Month-Olds Communicate Helpfully and Appropriately for Knowledgeable and Ignorant Partners. *International Journal of Cognitive Science*, *108*(3), 732-739
- Lohmann, H., & Tomasello, M. (2003). The Role of Language in the Development of False Belief Understanding: A Training Study. *Child Development*, *74*(4), 1130-1144.
- Low, J. (2010). Preschoolers' Implicit and Explicit False-Belief Understanding: Relations with Complex Syntactical Mastery. *Child Development*, *81*(2), 597-615.
- Low, J., Apperly, I. A., Rakoczy, H., & Butterfill, S. A. (2016). Cognitive Architecture of Belief Reasoning in Children and Adults: A Two-Systems Account Primer. *Child Development Perspectives*, *10*(3), 184-189.
- Low, J., Drummond, W., Walmsley, A., & Wang, B. (2014). Representing How Rabbits Quack and Competitors Act: Limits on Preschoolers' Efficient Ability to Track Perspective. *Child Development*, *85*(4), 1519-1534.
- Low, J., & Wang, B. (2011). On the Long Road to Mentalism in Children's Spontaneous False-Belief Understanding: Are We There Yet? *Review of Philosophy and Psychology*, *2*(3), 411-428.
- 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.
- Luo, Y. (2011a). Do 10-Month-Old Infants Understand Others' False Beliefs? *Cognition*, *121*(3), 289-298.
- Luo, Y. (2011b). Three-Month-Old Infants Attribute Goals to a Non-Human Agent. *Developmental science*, *14*(2), 453-460.
- Luo, Y., & Baillargeon, R. (2005). Can a Self-Propelled Box Have a Goal? Psychological Reasoning in 5-Month-Old Infants. *Psychological Science*, *16*(8), 601-608.

- Luo, Y., & Baillargeon, R. (2007). Do 12.5-Month-Old Infants Consider What Objects Others Can See When Interpreting Their Actions? *Cognition*, 105(3), 489-512.
- Luo, Y., & Beck, W. (2010). Do You See What I See? Infants' Reasoning About Others' Incomplete Perceptions. *Developmental science*, 13(1), 134-142.
- Luo, Y., & Johnson, S. C. (2009). Recognizing the Role of Perception in Action at 6 Months. *Developmental science*, 12(1), 142-149.
- Lurz, R. W. (2011). *Mindreading Animals: The Debate Over What Animals Know about Other Minds*. Cambridge, MA: MIT Press.
- Lycan, W. G. (1996). *Consciousness and Experience*. Cambridge, MA; London: MIT Press.
- Margolis, E., & Laurence, S. (2012). In Defense of Nativism. *Philosophical Studies*, 165(2), 693-718.
- 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.
- McAlister, A., & Peterson, C. C. (2006). Mental Playmates: Siblings, Executive Functioning and Theory of Mind. *British Journal of Developmental Psychology*, 24(4), 733-751.
- McAlister, A., & Peterson, C. C. (2007). A Longitudinal Study of Child Siblings and Theory of Mind Development. *Cognitive Development*, 22(2), 258-270.
- McAlister, A., & Peterson, C. C. (2013). Siblings, Theory of Mind, and Executive Functioning in Children Aged 3-6 Years: New Longitudinal Evidence. *Child Development*, 84(4), 1442-1458
- 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.
- Melot, A. M., & Angeard, N. (2003). Theory of Mind: Is Training Contagious? *Developmental science*, 6(3), 178-184
- Meltzoff, A. N. (1995). Understanding the Intentions of Others. *Developmental psychology*, 31(5), 838-850.
- Meltzoff, A. N., & Brooks, R. (2008). Self-Experience as a Mechanism for Learning About Others: a Training Study in Social Cognition. *Developmental psychology*, 44(5), 1257-1265.
- Meltzoff, A. N., & Moore, K. M. (1997). Explaining facial imitation: a theoretical model. *Early Development and Parenting*, 6(3-4), 179-192
- Michelon, P., & Zacks, J. (2006). Two Kinds of Visual Perspective Taking. *Perception & Psychophysics*, 68(2), 327-337.
- Miller, L. T., & Vernon, P. A. (1997). Developmental Changes in Speed of Information Processing in Young Children. *Developmental psychology*, 33(3), 549-554.
- Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and Theory of Mind: Meta-Analysis of the Relation Between Language Ability and False-belief Understanding. *Child Development*, 78(2), 622-646.
- Moll, H., & Tomasello, M. (2004). 12- and 18-Month-Old Infants Follow Gaze to Spaces Behind Barriers. *Developmental science*, 7(1), F1-F9.
- Moore, C., Pure, K., & Furrow, D. (1990). Children's Understanding of the Modal Expression of Certainty and Uncertainty and Its Relation to the Development of a Representational Theory of Mind. *Child Development*, 61(3), 722-730.
- Moriguchi, Y., Okanda, M., & Itakura, S. (2008). Young Children's Yes Bias: How Does It Relate to Verbal Ability, Inhibitory Control, and Theory of Mind? *First Language*, 28(4), 431-442.
- Moses, L. J., Coon, J. A., & Wusinich, N. (2000). Young Children's Understanding of Desire Formation. *Developmental psychology*, 36(1), 77-90.

- Muthukumaraswamy, S. D., Johnson, B. W., & McNair, N. A. (2004). Mu Rhythm Modulation During Observation of an Object-Directed Grasp. *Cognitive Brain Research*, *19*, 195–201.
- Nagel, T. (1974). What Is it Like to Be a Bat? *The Philosophical Review*, *83*(4), 435-450.
- Newell, A., & Rosenbloom, P. (1981). Mechanisms of Skill Acquisition and the Law of Practice. In J. R. Anderson (Ed.), *Cognitive Skills and Their Acquisition* (pp. 1-55). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Newton, A., M., & de Villiers, J. G. (2007). Thinking While Talking: Adults Fail Nonverbal False-Belief Reasoning. *Psychological Science*, *18*(7), 574-579.
- O'Neill, D. K. (1996). Two-Year-Old Children's Sensitivity to a Parent's Knowledge State When Making Requests. *Child Development*, *67*(2), 659-677.
- O'Neill, D. K., & Gopnik, A. (1991). Young Children's Ability to Identify the Sources of Their Beliefs. *Developmental psychology*, *27*(3), 390-397.
- Okanda, M., & Itakura, S. (2008). Children in Asian Cultures Say Yes to Yes–No Questions: Common and Cultural Differences Between Vietnamese and Japanese Children. *International Journal of Behavioral Development*, *32*(2), 131-136.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, *308*(5719), 255-258.
- Penke, L., Muñoz Maniega, S., Murray, C., Gow, A. J., Hernández, M. C. V., Clayden, J. D., . . . Deary, I. J. (2010). A General Factor of Brain White Matter Integrity Predicts Information Processing Speed in Healthy Older People. *The Journal of Neuroscience*, *30*(22), 7569-7574.
- Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.
- Perner, J. (2010). Who Took the Cog out of Cognitive Science? Mentalism in an Era of Anti-Cognitivism. In P. A. Frensch & R. Schwarzer (Eds.), *Cognition and Neuropsychology: International Perspectives on Psychological Science* (Vol. 1, pp. 241-262). New York: Psychology Press.
- Perner, J., & Kloo, D. (2003). Training Transfer Between Card Sorting and False Belief Understanding: Helping Children Apply Conflicting Descriptions. *Child Development*, *74*(6), 1823-1839.
- Perner, J., Lang, B., & Kloo, D. (2002). Theory of Mind and Self-Control: More Than a Common Problem of Inhibition. *Child Development*, *73*(3), 752-767.
- Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-Year-Olds' Difficulty with False Belief: the Case for a Conceptual Deficit. *British Journal of Developmental Psychology*, *5*(2), 125-137.
- Perner, J., & Ruffman, T. (2005). Infants' Insight into the Mind: How Deep? *Science*, *308*(5719), 214-216.
- Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of Mind Is Contagious: You Catch It from Your Sibs. *Child Development*, *65*(4), 1228-1238.
- Perner, J., & Wimmer, H. (1985). "John Thinks That Mary Thinks That..." Attribution of Second-Order Beliefs by 5- to 10-Year-Old Children. *Journal of Experimental Child Psychology*, *39*(3), 437-471.
- Peskin, J., & Ardino, V. (2003). Representing the Mental World in Children's Social Behavior: Playing Hide-and-Seek and Keeping a Secret. *Social Development*, *12*(4), 496-512.
- Peterson, C. C. (2000). Kindred Spirits: Influences of Siblings' Perspectives on Theory of Mind. *Cognitive Development*, *15*(4), 435-455.
- Peterson, C. C., & Siegal, M. (1995). Deafness, Conversation and Theory of Mind. *Journal of Child Psychology and Psychiatry*, *36*(3), 459-474.
- Peterson, C. C., & Siegal, M. (1999). Representing Inner Worlds: Theory of Mind in Autistic, Deaf and Normal Hearing Children. *Psychological Science*, *10*(2), 126-129.

- Peterson, C. C., Slaughter, V., & Wellman, H. M. (2017). Nimble Negotiators: How Theory of Mind (ToM) Interconnects With Persuasion Skills in Children With and Without ToM Delay. *Developmental psychology, Advance online publication.*
- Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory-of-mind development for children with autism and deafness. *Child Development, 76*(2), 502-517.
- Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The Mind Behind the Message: Advancing Theory-of-Mind Scales for Typically Developing Children, and Those With Deafness, Autism, or Asperger Syndrome. *Child Development, 83*(2), 469-485.
- Phillips, A. T., Wellman, H. M., & Spelke, E. (2002). Infants' Ability to Connect Gaze and Emotional Expression to Intentional Action. *Cognition, 85*(1), 53-78.
- Pinker, S. (1997). *How the Mind Works*. New York: Norton.
- Povinelli, D. J., & Vonk, J. (2003). Chimpanzee Minds: Suspiciously Human? *Trends in Cognitive Sciences, 7*(4), 157-169.
- Powell, L. J., & Carey, S. (2017). Executive Function Depletion in Children and Its Impact on Theory of Mind. *Cognition, 164*, 150–162.
- Prinz, J. J. (2002). *Furnishing the Mind : Concepts and Their Perceptual Basis*. Cambridge, MA; London: MIT Press.
- 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.
- Rakoczy, H. (2017). In Defense of a Developmental Dogma: Children Acquire Propositional Attitude Folk Psychology Around Age 4. *Synthese, 194*(3), 689-707.
- 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.
- Razza, R. A., & Blair, C. (2009). Associations Among False-Belief Understanding, Executive Function, and Social Competence: a Longitudinal Analysis. *Journal of Applied Developmental Psychology, 30*(3), 332–343.
- Repacholi, B. M., & Gopnik, A. (1997). Early Reasoning About Desires: Evidence From 14- and 18-Month-Olds. *Developmental psychology, 33*(1), 12-21.
- Repacholi, B. M., Meltzoff, A. N., & Olsen, B. (2008). Infants' Understanding of the Link Between Visual Perception and Emotion: “If She Can't See Me Doing It, She Won't Get Angry”. *Developmental psychology, 44*(2), 561–574.
- Rhodes, M., & Wellman, H. M. (2013). Constructing a New Theory From Old Ideas and New Evidence. *Cognitive Science, 37*, 592–604.
- Ristic, J., & Kingstone, A. (2005). Taking Control of Reflexive Social Attention. *Cognition, 94*(3), B55-B65.
- Rosenthal, D. (1997). A Theory of Consciousness. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness. Philosophical Debates* (pp. 729-754). Cambridge, MA: MIT Press.
- Roth, D., & Leslie, A. M. (1991). The Recognition of Attitude Conveyed by Utterance: a Study of Preschool and Autistic Children. *British Journal of Developmental Psychology, 9*(2), 315-330.
- Roth, D., & Leslie, A. M. (1998). Solving Belief Problems: Toward a Task Analysis. *Cognition, 66*(1), 1-31.
- Rubio-Fernández, P., & Geurts, B. (2013). How to Pass the False-Belief Task Before Your Fourth Birthday. *Psychological Science, 24*(1), 27-33.

- Rubio-Fernández, P., & Geurts, B. (2016). Don't Mention the Marble! The Role of Attentional Processes in False-Belief Tasks. *Review of Philosophy and Psychology*, 7(4), 835-850.
- Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review*, 34(3), 265–293.
- Ruffman, T., Garnham, W. A., Import, A., & Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology*, 80(3), 201–224.
- Ruffman, T., Perner, J., Naito, M., Parkin, L., & Clements, W. A. (1998). Older (but Not Younger) Siblings Facilitate False Belief Understanding. *Developmental psychology*, 34(1), 161-174.
- Ruffman, T., Slade, L., & Crowe, E. (2002). The relationship between children's and mothers' mental state language and theory-of-mind understanding. *Child Development*, 73(3), 734–751.
- Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C., & Garnham, A. (2003). How language relates to belief, desire, and emotion understanding. *Cognitive Development*, 18(1), 139-158.
- 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.
- Russell, P. A., Hosie, J. A., Gray, C. D., Scott, C., N., H., Banks, J. S., & C., M. M. (1998). The Development of Theory of Mind in Deaf Children. *Journal of Child Psychology and Psychiatry*, 39(6), 903-910.
- Sabbagh, M. A., Benson, J. E., & Kuhlmeier, V. A. (2013). False belief understanding in preschoolers and infants. In M. Legerstee, D. W. Haley, & M. H. Bornstein (Eds.), *The infant mind: Origins of the social brain*. New York: Guilford Press.
- Sabbagh, M. A., Bowman, L. C., Evraire, L. E., & Ito, J. M. B. (2009). Neurodevelopmental correlates of theory of mind in preschool children. *Child Development*, 80(4), 1147-1162.
- Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science*, 17(1), 74–81.
- Samet, J. (2008). The Historical Controversies Surrounding Innateness. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*.
- Samet, J., & Zaitchik, D. (2017). Innateness and Contemporary Theories of Cognition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2017 Edition)*.
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1255–1266.
- Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, 7(5), 499-500.
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporoparietal junction in “theory of mind”. *NeuroImage*, 19(4), 1835–1842.
- Schiffer, S. (1972). *Meaning*. Oxford: Clarendon Press.
- 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(3), 433-438.
- Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts implicit Theory-of-Mind processing. *Psychological Science*, 23(8), 842-847.
- Schneider, D., Nott, Z. E., & Dux, P. E. (2014). Task Instructions and Implicit Theory of Mind. *Cognition*, 133(1), 43-47

- Schneider, D., Slaughter, V., Becker, S. I., & Dux, P. E. (2014). Implicit False-Belief Processing in the Human Brain. *NeuroImage*, *101*, 268-275.
- Scholl, B. J., & Leslie, A. M. (1999). Modularity, Development and 'Theory of Mind'. *Mind & Language*, *14*(1), 131-153.
- Scholl, B. J., & Leslie, A. M. (2001). Minds, Modules, and Meta-Analysis. *Child Development*, *72*(3), 696-701.
- Schwier, C., Van Maanen, C., Carpenter, M., & Tomasello, M. (2006). Rational imitation in 12-month-old infants. *Infancy*, *10*(3), 303-311.
- Scott, R. M., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development*, *80*(4), 1172-1196.
- Scott, R. M., Baillargeon, R., Song, H.-j., & Leslie, A. M. (2010). Attributing false beliefs about non-obvious properties at 18 months. *Cognitive Psychology*, *61*(4), 366-395.
- 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.
- 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*(1), 32-56.
- Segal, G. (1996). The modularity of theory of mind. In P. Carruthers & P. K. Smith (Eds.), *Theories of Theories of Mind*. New York; Melbourne: Cambridge University Press.
- Senju, A., Southgate, V., & Snape, C. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychological Science*, *22*(7), 878-880.
- Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences of United States of America*, *113*(47), 13360-13365.
- Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics Made Easy: an Effort-Reduction Framework. *Psychological Bulletin*, *134*(2), 207-222
- Shahaeian, A., Henry, J. D., Razmjooe, M., Teymoori, A., & Wang, C. (2015). Towards a better understanding of the relationship between executive control and theory of mind: an intra-cultural comparison of three diverse samples. *Developmental science*, *18*(5), 671-685.
- Shahaeian, A., Peterson, C. C., Slaughter, V., & Wellman, H. M. (2011). Culture and the Sequence of Steps in Theory of Mind Development. *Developmental psychology*, *47*(5), 1239-1247.
- Shepard, R. N., & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects. *Science*, *171*(3972), 701-703.
- Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition*, *38*(1), 1-12.
- Simpson, T., Carruthers, P., Laurence, S., & Stich, S. P. (2005). Introduction: Nativism past and present. In P. Carruthers, S. Laurence, & S. P. Stich (Eds.), *The innate mind, Volume 1: Structure and contents* (pp. 3-22). New York; Oxford: Oxford University Press.
- Slaughter, V., & Gopnik, A. (1996). Conceptual Coherence in the Child's Theory of Mind: Training Children to Understand Belief. *Child Development*, *67*, 2967-2988.
- Slaughter, V., Imuta, K., Peterson, C. C., & Henry, J. D. (2015). Meta-Analysis of Theory of Mind and Peer Popularity in the Preschool and Early School Years. *Child Development*, *86*(4), 1159-1174.

- Slaughter, V., Peterson, C. C., & Mackintosh, E. (2007). Mind what mother says: Narrative input and theory of mind in typical children and those on the autism spectrum. *Child Development, 78*(3), 839-858.
- Slaughter, V., Peterson, C. C., & Moore, C. (2013). I Can Talk You Into It: Theory of Mind and Persuasion Behavior in Young Children. *Developmental psychology, 49*(2), 227–231.
- Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition, 96*(1), B1-B11.
- Song, H.-j., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental psychology, 44*(6), 1789–1795.
- Song, H.-j., Onishi, K. H., Baillargeon, R., & Fisher, C. (2008). Can an agent's false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition, 109*(3), 295-315.
- Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental science, 13*(6), 907-912.
- 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), 355-159.
- Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*(7), 587-592.
- Southgate, V., & Verneti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition, 130*(1), 1-10.
- Spaulding, S. (2011). A Critique of Embodied Simulation. *Review of Philosophy and Psychology, 2*(3), 579-599.
- Sperber, D. (2000). Metarepresentations in an Evolutionary Perspective. In D. Sperber (Ed.), *Metarepresentations: a Multidisciplinary Perspective* (pp. 117-137). Oxford: Oxford University Press.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford: Blackwell.
- Sperber, D., & Wilson, D. (2002). Pragmatics, Modularity and Mindreading. *Mind & Language, 17*(1-2), 3-23.
- Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind and Society, 11*(1), 3-13.
- Stein, N. L., & Levine, L. J. (1989). The Causal Organisation of Emotional Knowledge: A Developmental Study. *Cognition and Emotion, 3*(4), 343-378.
- Stich, S. P., & Nichols, S. (1998). Theory Theory to the Max. *Mind & Language, 13*(3), 421–449.
- Sullivan, K., & Winner, E. (1993). Three-Year-Olds' Understanding of Mental States: the Influence of Trickery. *Journal of Experimental Child Psychology, 56*(2), 135-148.
- Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of Beliefs by 13-Month-Old Infants. *Psychological Science, 18*(7), 580-586.
- Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British Journal of Developmental Psychology, 30*(1), 30–44.
- Surtees, A., Apperly, I. A., & Samson, D. (2013). Similarities and differences in visual and spatial perspective-taking processes. *Cognition, 129*(2), 426–438.
- Surtees, A., Apperly, I. A., & Samson, D. (2016). I've Got Your Number: Spontaneous Perspective-Taking in an Interactive Task. *Cognition, 150*(1), 43-52.
- Surtees, A., Butterfill, S. A., & Apperly, I. A. (2011). Direct and Indirect Measures of Level-2 Perspective-Taking in Children and Adults. *British Journal of Developmental Psychology, 30*(1), 75-86.

- Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child Development, 79*(4), 866-881.
- Teufel, C., Alexis, D., Clayton, N. S., & Davis, G. (2010). Mental-State Attribution Drives Rapid, Reflexive Gaze Following. *Attention, Perception, & Psychophysics, 72*(3), 695-705.
- Thompson, R. J. (2014a). Meaning and Mindreading. *Mind & Language, 29*(2), 167-200.
- Thompson, R. J. (2014b). Signature Limits in Mindreading Systems. *Cognitive Science*(38), 1432-1455.
- Thompson, R. J. (2015). Ruling Out Behavior Rules: When Theoretical Virtues and Empirical Evidence Collide. *Review of General Psychology, 19*(1), 14-29.
- Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.
- Tomasello, M., & Haberl, K. (2003). Understanding Attention: 12- and 18-Month-Olds Know What Is New for Other Persons. *Developmental psychology, 39*(5), 906-912.
- Träuble, B., Marinović, V., & Pauen, S. (2010). Early Theory of Mind Competencies: Do Infants Understand Others' Beliefs? *Infancy, 15*(4), 434-444.
- Turken, U., Whitfield-Gabrieli, S., Bammer, R., Baldo, J. V., Dronkers, N. F., & Gabrieli, J. D. E. (2008). Cognitive Processing Speed and the Structure of White Matter Pathways: Convergent Evidence from Normal Variation and Lesion Studies. *NeuroImage, 42*(2), 1032-1044.
- Tye, M. (1995). *Ten Problems of Consciousness: a Representational Theory of the Phenomenal Mind*. Cambridge, MA; London MIT Press.
- 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.
- Wang, Y., & Su, Y. (2009). False Belief Understanding: Children Catch It from Classmates of Different Age. *International Journal of Behavioral Development, 33*(4), 331-336
- Wellman, H. M. (2010). Developing a Theory of Mind. In U. Goswami (Ed.), *The Wiley-Blackwell Handbook of Childhood Cognitive Development* (pp. 258-284). Oxford: Wiley-Blackwell.
- Wellman, H. M. (2014). *Making Minds*. New York, NY: Oxford University Press.
- Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition, 30*(3), 239-277.
- 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.
- Wellman, H. M., Fang, F., & Peterson, C. C. (2011). Sequential progressions in a Theory-of-Mind scale: Longitudinal perspectives. *Child Development, 82*(3), 780-792.
- Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind tasks. *Child Development, 75*(1), 523-541.
- Wellman, H. M., & Peterson, C. C. (2013). Deafness, Thought Bubbles, and Theory-of-Mind Development. *Developmental psychology, 49*(12), 2357-2367.
- Wellman, H. M., Phillips, A. T., Dunphy-Lelii, S., & LaLonde, N. (2004). Infant social attention predicts preschool social cognition. *Developmental science, 7*(3), 283-288.
- Wellman, H. M., Phillips, A. T., & Rodriguez, T. (2000). Young children's understanding of perception, desire, and emotion. *Child Development, 71*(4), 895-912.
- Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: the early development of everyday psychology. *Cognition, 35*(3), 245-275.
- Westra, E. (2016a). Pragmatic Development and the False Belief Task. *Review of Philosophy and Psychology, 8*(2), 235-257.

- Westra, E. (2016b). Spontaneous Mindreading: a Problem for the Two-Systems Account. *Synthese*, 194(11), 4559-4581.
- Westra, E., & Carruthers, P. (2017). Pragmatic Development Explains the Theory-of-Mind Scale. *Cognition*, 158(1), 165-176
- Wiese, E., Wykowska, A., Zwickel, J., & Müller, H. J. (2012). I See What You Mean: How Attentional Selection Is Shaped by Ascribing Intentions to Others (Social Attention Is Shaped by Ascribing Intentions). *PLoS one*, 7(9), e45391.
- Wiesmann, G. C., Schreiber, J., Singer, T., Steinbeis, N., & Friederici, A. D. (2017). White Matter Maturation Is Associated with the Emergence of Theory of Mind in Early Childhood. *Nature Communications*, 8, 14692.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103-128.
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69(1), 1-34.
- Woolfe, T., Want, S. C., & Siegal, M. (2002). Signposts to Development: Theory of Mind in Deaf Children. *Child Development*, 73(3), 768-778.
- Zawidzki, T. W. (2011). How to Interpret Infant Socio-Cognitive Competence. *Review of Philosophy and Psychology*, 2(1), 483-497.
- Zmyj, N., Daum, M. M., & Aschersleben, G. (2009). The development of rational imitation in 9- and 12-month-old infants. *Infancy*, 14(1), 131-141.