Movement as a window to the mind: grasping adults' perception and understanding of others' goal-directed movements

by SCHYANA SIVANANTHAM

Thesis submitted to Te Herenga Waka–Victoria University of Wellington

in fulfilment of the requirements for the

degree of Doctor of Philosophy

Te Herenga Waka–Victoria University of Wellington 2021

Abstract

Did you know you are a mindreader? Most of us have a fundamental curiosity about people and are in fact engaging in the process of mindreading every day, as we navigate our social worlds. For instance, we often think about what other people are thinking. On a more basic level, we even predict others' goals based on their actions, such as when they reach for and grasp objects in their environment. This thesis aimed to test and validate a classic action observation paradigm. In order to do so, three experiments were conducted to investigate action observation for goal-directed movements. In Experiment 1, adults' eye gaze was tracked as they viewed reach-tograsp movements. Contrary to standard predictions, results did not replicate gaze proactivity according to a motor matching account. Rather, adults' eyes latched on to targets that were larger and/or nearer to the agent's hand. A motor matching account would have predicted that grasp type (pincer grasp or whole-hand prehension) should cue gaze proactivity to the congruent target (i.e., small object or large object, respectively). In Experiment 2, adults' reaction times were measured as they viewed the same stimuli, but presented in static rather than dynamic format. Similarly to Experiment 1, adults' response times were faster to detect a target when it appeared over an object closer to the reaching hand, rather than an object farther away from the hand. Again, this was not in line with a motor matching account. Finally, in Experiment 3, adults' explicit probability judgements were solicited for still frames taken from the original video stimuli. Yet again, a distance effect prevailed, whereby adults explicitly predicted that the agent's hand would contact the closer of two objects, even when the hand was at rest (in the absence of any motor cue). Overall, these results imply that adults are applying non-motor heuristics during action observation. As such, the stimuli tested in these classical studies may not be fit for purpose, that is, they do not reliably show a motor matching effect during action observation in adults. The implications of these findings for future research and theorising are also discussed.

Keywords: mindreading; action observation; goal-tracking; goal-directed actions; motor matching

ii

To my primary supervisor, Associate Professor Jason Low, thank you for everything, I could not have asked for more. You have always had time for me, no matter how big or small my query, and I have appreciated your guidance, patience, and understanding over the past years. You have taught me so much, and thanks to you I get what it means to stand on the shoulders of giants. I am grateful for my lab mates—past and present—Katheryn, Christopher, Claudie, Pieter, Giovanni, and Cong (as well as my newer office mates) for being my Uni family and bringing me joy during this journey.

I am ever thankful to my loving partner Benjamin, my parents Siva and Sharmini, and my sister Sarsha, for being so supportive throughout and enabling me to achieve my potential. Dearest Uni friends, you know who you are, I am indebted to you for the countless hours of catchups on campus and chats about everything under the sun. Cheers to my pub quiz team for keeping me attending at least one social event a week (up until the end phase when I dropped off!). Dear friends overseas, for all the voice messages auf Deutsch und Englisch: *dankeschön*.

Thank you to my beloved piano for literally always being there for me and being an outlet for my head and heart. Thank you to pole for taking me out of my research headspace (and comfort zone) and showing me my own strength. Thank you to Nature for keeping me grounded when my mind was pacing. I am so blessed to have spent these Ph.D. years in my favourite, beautiful, windy city that I call home. Speaking of blessings, I express much gratitude to those who have breathed light into my life these recent years.

All these things kept me sane during what has been the most difficult endeavour I have embarked upon thus far. Let us not forget the global pandemic and ensuing national lockdown that stalled my final stretch. Despite imposter syndrome never fully leaving my side, I can honestly say I am proud of myself for making it to the finish line (without even a single drop of caffeine, might I add!) as I turn over into the next quarter-century of my life.

I would like to acknowledge the institution *Te Herenga Waka*–Victoria University of Wellington, and *Te Kura Mātai Hinengaro*–the School of Psychology, for being my second home. VUW supported me to complete my doctorate, and I am honoured to have been part of the critic and conscience of society. Lastly, I extend hearty thanks to all my participants for their time and use of their data, without which this dissertation would not have been possible.

Mauri ora!

Table of Contents

Abstract	ii	
Acknowl	Acknowledgementsiii	
Table of	Contentsiv	
List of Ta	ablesviii	
List of Fi	juresix	
Preface	xiii	
Chapter	r 1	
1.1. IIIt	roduction	
1.2. MI	Or enotionalization. Deeple can used minds?	
1.2.1.	Operationalisation: People can read minds?!	
1.3. Fai	se focus on false bellef	
1.3.1.	All about Max1	
1.3.2.	Tunnel vision is problematic	
1.3.3.	Conceptual change	
1.4. Inf	ant studies spell the end for conceptual change?7	
1.4.1.	Violation of expectation	
1.4.2.	Anticipatory looking 11	
1.4.3.	Relevance of infant studies for conceptual change12	
1.5. Tw	o-systems account to the rescue?13	
1.5.1.	An answer to the paradox? Signature limits14	
1.5.2.	Evidence for two-systems account15	
1.6. A c	risis of replication18	
1.6.1.	The problem is task replicability19	
1.6.2.	Another blow to infant mindreading 22	
1.6.3.	Grim outlook calls for paring back to basics	
1.7. Co	nclusion	
Chanta		
	r 2 23	
2.1. IIII	treak on to know?	
2.2. 10	o track of to know?	
2.3. In		
2.3.1.	Habituation (looking-time) studies	
2.3.2.	Eye gaze: from looking longer to looking in anticipation	
2.4. Go	al-tracking 'au naturale'	
2.5. Te	leology: the rational way	
2.5.1.	The rational way is not the only way	
2.6. In	fants do statistics	
2.6.1.	Infants doing statistics may not spell the end for teleology	
2.7. Tin	ning is everything	

2.8. Mu	ust walk the walk to track goals	
2.9. Co	nclusion	
Chanta		
	- 3 roduation	
3.1. III	tout the own for goal treaking	
3.2. MO	Empirical avidance for a motor theory of goal tracking	
3.2.1.	Empirical evidence for a motor meory of goal-tracking	47
3.3. WI	More tempets and more acals	
3.3.1.	Desiring the meter the surface and bracking	
3.3.2.	Revising the motor theory for goal-tracking	
3.4. Du	al-process theory for goal-tracking	
3.5. Co	nclusion	
Chapter	۶ 4	55
4.1. Int	roduction	55
4.2. Th	e motor system and neurons that mirror	55
4.3. Di	rect matching hypothesis	
4.4. Ev	idence for motorically grounded mindreading	
4.4.1.	Response facilitation	
4.4.2.	Directional leaning	
4.4.3.	Object detection and Egyptian mummy: what's the link?	60
4.5. Ac	tion observation paradigm	
4.5.1.	The original	63
4.5.2.	My hands are tied	
4.5.3.	Infants' own grasping matters	
4.5.4.	Theoretical challenges	
4.5.5.	Practical challenges	
4.6. Gr	asp cueing paradigm	
4.6.1.	Static versus dynamic	
4.7. Re	plicate, replicate, replicate!	73
4.8. Co	nclusion	74
Chanta		-
5 1 Int	oduction	······/3
5.1. III.	athodology	
5.2. MA	Darticinants	
ე.∠.1. ⊑ეე	1 articipanto Matoriale	/0
J.∠.∠. Stim	materiais	·····//
Anne	ull	
Арра	11 atus Drogoduro	······································
5.2.3.	rioceuure	
5.3. ке	Data raduation	
5.3.1.		

5.3.2.	Data analysis	80
5.4. Dis	scussion	
5.5. Con	nclusion	
Chapter	• 6	
6.1. Intr	roduction	
6.2. Me	ethodology	
6.2.1.	Participants	
6.2.2.	Materials	
Stim	uli	
Appa	uratus	
6.2.3.	Procedure	
6.3. Res	sults	
6.3.1.	Data reduction	
6.3.2.	Data analysis	91
6.4. Dis	scussion	
6.5. Co	nclusion	
Chapter	• 7	
7.1. Inti	roduction	
7.2. Met	thodology	
7.2.1.	Participants	
7.2.2.	Materials	
Stim	uli	
Appa	ıratus	102
7.2.3.	Procedure	103
7.3. Res	sults	104
7.3.1.	Data reduction	104
7.3.2.	Data analysis	104
Coml	bined analysis for whole-hand prehension	104
Who	le-hand prehension when large target near to hand	106
Who	le-hand prehension when large target far from hand	
Com	bined analysis for pincer grasp	109
Pince	er grasp when small target near to hand	
Pince	er grasp when small target far from hand	112
Coml	bined analysis for closed fist when target object large	113
Close	ed fist when large target near to hand	
Close	ed fist when large target far from hand	115
Com	bined analysis for closed fist when target object small	117
Close	ed fist when small target object near to hand	
Close	ed fist when small target object far from hand	120
7.4. Dis	scussion	

7.5.	Con	clusion12	<u>23</u>
Chap	oter	8	4
8.1.	Intr	oduction12	24
8.2.	Big	picture12	24
8.2	.1.	Results refresher12	25
8.3.	Tria	angulation12	26
8.4.	Les	sons from this thesis about (non-)functionality of action observation	
stimu	ıli		27
8.4	.1.	Overlap in muscle use	28
8.4	.2.	Iconicity12	<u>29</u>
8.4	.3.	Familiarity130	;0
8.4	.4.	Spatial layout of objects 13	31
8.5.	Cur	rrent experiments: problems and solutions13	32
8.5	.1.	Engagement and motivation13	32
8.5	.2.	Assuming homogeneity in behaviour and cue processing13	33
8.5	.3.	What about from <i>your</i> perspective?13	35
8.5	.4.	On using more appropriate control trials13	37
8.5	.5.	Weirdness strikes again138	;8
8.6.	Imp	plicit versus explicit goal-tracking13	39
8.6	.1.	If it ain't broke, don't fix it14	0
8.7.	Wh	ither motor matching: what's the prognosis?14	41
8.7	.1.	Very peripersonal14	12
8.7	.2.	Conjecture: perceptual analysis and conceptual processing14	13
8.7	.3.	Support for perceptual-conceptual conjecture14	15
8.8.	Tin	ne to put the 'inter' back in inter-action14	15
8.9.	Sur	nmary and Conclusion14	17
Refe	ren	ces14	.8

List of Tables

Table 1 Mean chance (with standard deviations and standard errors of means)of hand contacting near or far object for each scene in whole-hand
grasp sequence105
Table 2 Mean chance (with standard deviations and standard errors of means)of hand contacting near or far object for each scene in pincer graspsequence.109
Table 3 Mean chance (with standard deviations and standard errors of means)of hand contacting near or far target for each scene in closed fistsequence (when target was the large object).114
Table 4 Mean chance (with standard deviations and standard errors of means)of hand contacting near or far target for each scene in closed fistsequence (when target was the small object)

List of Figures

<i>Figure 1</i> Test trial events, whereby the agent reaches into the green box or the
yellow box
<i>Figure 2</i> Test trial event frame, whereby the visored agent turns back to the scene and the two windows above the boxes illuminate to cue anticipatory looking
<i>Figure</i> 3 Dog-robot stimulus toy that has one red side and one blue side
<i>Figure 4</i> The elephant in the room: Replication Crisis
<i>Figure 5</i> Test events, whereby the rectangular block was removed and infants saw the new action or the old action
<i>Figure 6</i> Depiction of events in the hand condition
<i>Figure 7</i> Depiction of events in the claw condition
<i>Figure 8</i> Frames of the introductory stimuli, whereby a cow takes a longer path because the shorter path cannot be taken. The arrows indicate direction of movement; the oval indicates the occluder
<i>Figure 9</i> Resting posture depicting a hand (with fixation cross) before the hand grasps towards either the large or the small object
<i>Figure 10</i> 'Mummified' agent bound in linen sheets61
Figure 11 Snapshots of stimulus videos involving two tomato objects. The rows depict the following variations (from top to bottom: no shape gesture to large target, no shape gesture to small target, whole-hand prehension to large target, pincer grasp to small target)
<i>Figure 12</i> Snapshots of stimulus videos involving two green ball objects. This figure shows the target layout whereby the large object is near the hand. The row depicts the following variation: whole-hand prehension to large target. The white circle in the left-hand column depicts the area of interest superimposed on the target object 66
<i>Figure 13</i> Sample event sequence. The left grasp gesture panel depicts an incongruent trial (pincer grasp followed by target on large object), and the right grasp gesture panel depicts a congruent trial (whole-hand prehension followed by target on large object)
<i>Figure 14</i> Snapshots of stimulus videos involving two green ball objects. This figure shows the target layout whereby the large object is near the hand (object position was counterbalanced). The rows depict the following variations (from top to bottom: no-shape gesture to small target, pincer grasp to small target, no-shape gesture to large target, whole-hand prehension to large target). The white circle in the left-hand column depicts the area of interest superimposed on the target object
<i>Figure 15</i> Time of gaze proactivity (with 1 standard error mean, henceforth SEM) at the target object, relative to the end of the trial video. Gaze

proactivity is plotted as a function of target size and hand shape. The end of the trial video is represented by the horizontal line at zero seconds. Negative values represent proactive eye movements. *Figure 16* | Time of gaze proactivity (with 1 SEM) at the target object, relative to the end of the trial video. Gaze proactivity is plotted as a function of target size and distance from the agent's hand. The end of the trial video is represented by the horizontal line at zero seconds. Negative Figure 17 | Sample event sequence for Experiment 2, depicting variations of incongruent trials (top panels show incongruent pincer grasp trial; lower panels show incongruent whole-hand prehension trial). Congruent trials would have target on small object for pincer grasp Figure 18 | Mean reaction times (RTs, in ms, with 1 SEM) to detect targets in Experiment 2 with unpredictive grasps that were congruent or incongruent with the object over which the target appeared (i.e., congruent: pincer grasp and small object or whole-hand prehension and large object; incongruent: pincer grasp and large object or whole-hand prehension and small object). Target delay refers to the onset of a vellow star superimposed over target object, in ms.91 *Figure* 19 | Mean reaction times (RTs, in ms, with 1 SEM) for pincer grasp trials to detect targets in Experiment 2 that were congruent or incongruent with the object over which the target appeared, when the object near (closest to) the hand was large or small. Points A-D *Figure 20* | Visual representation of four possible outcome configurations for pincer grasp trials. Prior to each of these images, the hand displayed a pincer grasp. (A) depicts a congruent trial with the large object near to hand; (B) depicts an incongruent trial with the small object near to hand; (C) depicts a congruent trial with the small object near to hand; (D) depicts an incongruent trial with the large Figure 21 | Mean reaction times (RTs, in ms, with 1 SEM) for whole-hand prehension trials to detect targets in Experiment 2 that were congruent or incongruent with the object over which the target appeared, when the object near (closest to) the hand was large or Figure 22 | Visual representation of four possible outcome configurations for whole-hand prehension trials. Prior to each of these images, the hand displayed a whole-hand prehension. (A) depicts a congruent trial with the small object near to hand; (B) depicts an incongruent trial with the large object near to hand; (C) depicts a congruent trial with the large object near to hand; (D) depicts an incongruent trial

Figure 23 1	Stimulus images involving two green ball objects. This figure shows the target layout whereby the large object is near the hand (object location was counterbalanced). The rows depict the following variations (from top to bottom: no-shape gesture to small target, no-shape gesture to large target, pincer grasp to small target, whole-hand prehension to large target)
Figure 24 1	Snapshot of one of the 40 (five scenes × eight trial types/videos) possible images, depicting sliders used to elicit participants' probability judgements (%). It is not necessary to read the small text instructions, this figure simply depicts how participants evaluated each image
Figure 25 1	Mean percentage predictions (with 1 SEM) of hand contacting near large object for each scene in whole-hand prehension sequence. To- T4 refer to the 5 different time points/scenes corresponding to the stimulus images below the graph (i.e., To: scene 1, T1: scene 2, etc.).
Figure 26 1	Mean percentage predictions (with 1 SEM) of hand contacting far large object for each scene in whole-hand prehension sequence. To- T4 refer to the 5 different time points/scenes corresponding to the stimulus images below the graph
Figure 27	Mean percentage predictions (with 1 SEM) of hand contacting near small object for each scene in pincer grasp sequence. To-T4 refer to the 5 different time points/scenes corresponding to the stimulus images below the graph
Figure 28 1	Mean percentage predictions (with 1 SEM) of hand contacting far small object for each scene in pincer grasp sequence. To-T4 refer to the 5 different time points/scenes corresponding to the stimulus images below the graph
Figure 29 1	Mean percentage predictions (with 1 SEM) of hand contacting near large object for each scene in closed fist sequence. To-T4 refer to the 5 different time points/scenes corresponding to the stimulus images below the graph
Figure 30 1	Mean percentage predictions (with 1 SEM) of hand contacting far large object for each scene in closed fist sequence. To-T4 refer to the 5 different time points/scenes corresponding to the stimulus images below the graph
Figure 31 I	Mean percentage predictions (with 1 SEM) of hand contacting near small object for each scene in closed fist sequence. To-T4 refer to the 5 different time points/scenes corresponding to the stimulus images below the graph
Figure 32 1	Mean percentage predictions (with 1 SEM) of hand contacting far small object for each scene in closed fist sequence. To-T4 refer to the 5 different time points/scenes corresponding to the stimulus images below the graph

<i>Figure 33</i> Visual representation of the current thesis' use of triangulating multiple methods (across Experiments 1-3). A mix of implicit and explicit measures were used
<i>Figure 34</i> Visual depiction (from left to right) of closed fist, pincer grasp, and whole-hand prehension129
<i>Figure 35</i> Video excerpts, displaying a hand reaching towards two objects (a small tomato and a large tomato). The blue line in the central image indicates an example of proactive gaze moving ahead to the objects before the hand has reached its target
<i>Figure 36</i> Exemplar of diverse strategy use across development, taking inspiration from Siegler's (2016) overlapping waves model. The frequency of each strategy's use (%) varies as an action unfolds across time
<i>Figure 37</i> Visual depiction of different action observation layouts from bird's eye view, with distinctly separate target object locations and occluders to visually block motor cue of the hand before the outcome/contact phase. Four rotated perspectives are shown, which vary in respect to how closely they map on to the observer.

Preface

Your perception of me

is a reflection of you.

My reaction to you

is an awareness of me.

- Bobbi Chegwyn

Chapter 1.

1.1. Introduction

It occurs right under–or rather, above–our noses on a daily basis, helps us interact with our complex social environment, and may be implicated in human mindreading. 'It', in this case, is referring to the humble undertaking of action observation. Action observation is a staple in this social, cognitive world. Watching people around us act on different objects teaches us how to navigate through everyday life. Representation of observed actions can help with the prediction (or anticipation) of someone's actions. For instance, observing how our friend might grasp a book or a mug, we can track our friend's goal by monitoring the early part of their action, and in so doing, we can be prepared to respond in an appropriate manner. When considering others' small-scale actions (like book-grasping and mug-grasping), action observation and action anticipation often happen without conscious effort. If we were to explicitly analyse every small action we viewed, it would overload our mind's capacity. That said, how do we theoretically understand the process of tracking others' actions? And are motor representations relevant to goal-tracking in adults? Such questions will be reviewed in this thesis.

The current chapter will introduce us to the concept of mindreading, including why one particular aspect of mindreading—false belief—has been the fixation of many researchers in this field. After this, important non-verbal paradigms for infants will be discussed, which lead to a paradox that has, until recently, remained a mystery. Then, the two-systems account is offered as a potential theory to solve this paradox. Finally, the unavoidable replication crisis in psychology and the wider sciences is addressed.

1.2. Mindreading, née Theory of Mind

As a result of the global Covid-19 pandemic, lockdowns around the world have highlighted the negative effects of physical and social isolation. Humans are inherently social animals, and—in an evolutionary sense—we proliferated because we learnt to work together in groups. In order to cultivate social cohesion, action observation and action anticipation would have been crucial. These principles remain important; we monitor the actions of others every day, because we are interested in and track people's goals (i.e., actions directed to outcomes). What are they about to do? Does their upcoming action involve or affect me? As well as actions, there are a lot of other components we are interested in that originate within people (including ourselves), such as beliefs, desires, thoughts, intentions, and so on. These propositional states all fall under the concept of mindreading. So, what exactly *is* mindreading, and how is it that humans track others' mental states?

1.2.1. Operationalisation: People can read minds?!

Did you know you are a mindreader? While the lay understanding of the term mindreading may have telepathic or supernatural connotations, psychologically speaking, mindreading is a newer name for what used to be known as theory of mind. With researchers' recent move towards the term mindreading, Apperly (2011) aptly describes how mindreading is something that you do, while theory of mind is something you *have*. This thesis will predominantly use the term mindreading, since the current, most relevant body of work (i.e., relating mindreading and motor cognition) does so. Nonetheless, the concept will briefly be explored below through the lens of the 'theory of mind' label, as it was originally conceptualised. Possessing a theory of mind shapes our experience of the world (Lavelle, 2018). Once we acquire the theory, we use it whenever we engage with others; it guides our perception of why people do things, our predictions of what people might do next, and so on. "The claim is that, without the theory of mind to give meaning to our mental state terms, we would not be able to see someone's arm movement as a 'reach'" (Lavelle, 2018, p. 32). Lavelle also describes it as a tacit theory in that, although we use it so often, we do not feel as if we are explicitly drawing on a theory whenever we socially interact with a person; the theory is not conscious, as such, but it nevertheless helps to explain processes that occur in the mind.

Although mindreading is not a superpower (most mature humans have this ability to some extent), it remains impressive. This is not disputed. There is, however,

much debate over the developmental trajectory of mindreading, as well as its theoretical underpinnings. As we will see shortly, there has been research into the existence of mindreading among non-human animals. However, considering the extent to which humans mindread (i.e., the complexity of mental states we can comprehend), mature, full-blown mindreading appears to be—for now—a uniquely human capability.

Does the chimpanzee have a theory of mind? This is the question Premack and Woodruff (1978) asked in their seminal paper in the area of primatology. In their study, Sarah (a chimpanzee), watched videos of her trainer trying to solve problems. The video stopped at the last frame and Sarah received two photos: in one, the trainer solves the problem; in the other, the trainer is doing something unrelated to the task. In most of the trials, Sarah chose the photo depicting the trainer resolving the problem shown in the video. The authors argued that the likely explanation for these findings was Sarah tracking the actor's goals and beliefs. However, the opposing associationist behavioural view claims, instead, that one can recognise behavioural patterns and choose the next behaviour in the sequence (Lavelle, 2018); this explanation does not involve any psychological attribution. More broadly speaking, a central theoretical division exists between nativism and empiricism - much like that in the classic nature versus nurture debate. Nativists are of the view that mindreading is mostly innate. In contrast, empiricists posit that mindreading is learned. Additionally, constructivism holds that we also analyse our own experiences in order to construct knowledge. Indeed, there are many theoretical camps regarding the underlying explanations of empirical mindreading research, and the upcoming sections give a glimpse into these ongoing debates.

1.3. False focus on false belief

Psychologists and philosophers also use the term folk psychology to denote our everyday understanding of minds (Lavelle, 2018). Currently, the attribution of epistemic mental states (such as beliefs, desires, knowledge) is central to this folk psychological understanding. But, why does the mindreading literature focus on high-level mental state attribution in this way? Let us look back and find out. Since its inception, the false

belief task has been the most widely used tool to assess mindreading in children (beliefs are propositional attitudes we can attribute to others; beliefs can be true or false, depending on how they relate to reality). The basis of this task is such that the child participant and the target agent end up with differing perspectives about a situation.

To contextualise the focus on children, most extant work in the field has centred around the developmental time period when children obtain mindreading abilities (Apperly, 2011). Unlike other topics in the cognitive domain (e.g., number cognition, spatial cognition), there is not as much literature on mindreading in adult humans. This contrasts with the extensive cognitive psychological research on mindreading in infants, children, those with clinical conditions, and non-human animals. This accounts for why a significant portion of the ground-clearing research covered in this chapter (and in Chapters 2 and 3) has a developmental focus. Nonetheless, developmental findings are relevant to informing research on cognition in adults, as they speak to the basic kinds of processes and representations that may be (1) developing early in life, and (2) used for action observation across age groups. Additionally, there is a trope that since adults have already attained a theory of mind, it is not as pertinent to study them. It will become apparent over the next few chapters that adult mindreading remains an important inquiry of investigation.

1.3.1. All about Maxi

Let us take a closer look at the famous false belief task. Wimmer and Perner (1983) developed a false belief paradigm—commonly known as the Maxi task—which tests whether young children can attribute false beliefs to other people. A false belief is simply a belief someone has, that is not in line with reality. In a puppet show, children see Maxi hide his chocolate in cupboard-A. Maxi leaves, and in his absence his mother enters the scene. She moves the chocolate from cupboard-A to drawer-B. The mother leaves and Maxi returns. Children are asked the test question "where will Maxi look for his chocolate?" and must indicate the cupboard-A location for a correct response. If they do so it is assumed children are attributing a false belief to Maxi (about the location of the chocolate). He believes it is in one location (A) when it is, in fact, in another location

(B); since his belief opposes reality, he holds a false belief about object location. The authors found that children from age four-years and above succeed in this task, showing an ability to attribute false belief to an agent. Interestingly, children younger than four years fail the task, answering that Maxi will look in the drawer—the second location—where the chocolate actually is. A fully-fledged mindreading system would help us perceive that regardless of the chocolate's true location, Maxi did not see it moved there, so would have no reason to believe it is in the new location.

This age difference finding has been replicated numerous times. It is, however, worthwhile to note the different contexts under which this result can change. For instance, children tend to pass the test *earlier* if they have siblings, or if the test is more interactive as opposed to passively watching, but they tend to pass the test *later* if they come from a lower socioeconomic background, or if they are on the autism spectrum (e.g., see Baron-Cohen, Leslie, & Frith, 1985; Cutting & Dunn, 1999; Ruffman, Perner, & Parkin, 1999). There are also cross-cultural data suggesting that children from different countries pass this test at different ages (e.g., see Dixson, Komugabe-Dixson, Dixson, & Low, 2018 regarding children's false belief acquisition in Vanuatu; Nawaz & Lewis, 2018 regarding social understanding in Pakistan).

1.3.2. Tunnel vision is problematic

There are three main challenges to the notion that high-level mental state attribution (like belief understanding) is the central tenet of mindreading. Firstly, many social interactions occur without the need for mindreading. For instance, we often use schemas, scripts, and social norms to navigate our social world. While this thesis does not spotlight cross-cultural psychology, Lavelle (2018) points out that—from a crosscultural standpoint—the emphasis on belief/desire attribution constituting folk psychology may have been influenced by the Western conceptualisation of the mind. Secondly, aside from monitoring beliefs and desires, we also track other socially relevant cues and other psychological states to interact with the world around us. For instance, we use stereotypes and character traits to think about others' minds. Thinking back to book-grasping and mug-grasping, we even track others' goals, that is, people's actions that are directed to outcomes.

Lastly, and most importantly, is the issue of overfocus itself. Although the fixation on this particular false belief task has led to rich interpretations and robust replication, it is deeply problematic from the perspective of theoretical development. Returning to the term theory of mind, it was originally thought that a theory was necessary to understand mental states. Thus, mindreading was considered to be the same as the attribution of propositional attitudes to other people (Lavelle, 2018). Mindreading is, in fact, broader than this. Apperly (2011) notes that while false belief tasks tend to generalise to other mindreading abilities, there are still processes involving mindreading that do not require the ascription of mental states. Despite this, early theoretical work mistakenly conceptualised the two as being interchangeable terms for the same thing, thereby leading to a narrower view of what constitutes mindreading. In reality, false belief attribution and mental state understanding are only one subcomponent contained under the wider umbrella term of mindreading. Considering this history, it is valuable to look at research that claims mindreading can exist without involving propositional attitude attribution. Some such research avenues are referred to as minimal mindreading accounts, which will be elaborated upon later in this thesis.

1.3.3.Conceptual change

A meta-analysis of more than 150 studies, conducted by Wellman, Cross, and Watson (2001), found that children are more likely to pass unexpected-transfer tasks (e.g., Maxi task) at around four years of age. Children younger than this systematically respond in an egocentric way, thinking only about their own perspective and failing to compute the agent's false belief. This constitutes consistent evidence for a conceptual change account, which refers to there being a noticeable developmental shift in children's mindreading abilities (i.e., a point in development whereby their theory of mind undergoes a distinct and conceptual change). This is contrary to the early competence view that early failure on standard false belief tasks is due to task demands. Wellman and Liu's (2004) theory of mind scale addressed the developmental sequence in which children pass particular mindreading tasks. They found common patterns that quantitatively support a conceptual change account. Wellman and Liu's meta-analysis (in Study 1) showed, for instance, that children first master desire judgements, before progressing on to correctly judge beliefs. Another strong finding from the meta-analysis is that children judge diverse beliefs correctly before being able to judge false beliefs. Further to this, children appear to understand ignorance before understanding false belief. In all cases, belief—especially false belief—seems to be the most difficult of all mindreading components to fully grasp.

As can be seen, mindreading becomes more sophisticated throughout childhood. Wellman and Liu (2004) remind us that an adequate theory for mindreading must account for these sequenced acquisitions that occur throughout child development. Mindreading is also linked to executive functioning (Carlson & Moses, 2001) and language development (Apperly, 2011). Because most studies, and even theories, have emphasised this developmental period, there is a gap from childhood to adulthood. What happens to mindreading during this time? Is a theory of mind simply obtained, and does it then remain static over the remainder of the lifespan?

An important takeaway here is that false belief understanding, while complex and worthwhile to study, is not the be-all and end-all of mindreading, despite the literature focusing heavily on this one component. There are many different mental states and—aside from tracking and reasoning about others' false beliefs—one also needs to track and attribute goals, perceptions, desires, and knowledge to others at an earlier age. One can subscribe to a conceptual change account which accepts definite changes in children's mindreading abilities, and still be critical of false belief tasks being used as a litmus test, or hallmark, of mindreading abilities as a whole (Apperly, 2011).

1.4. Infant studies spell the end for conceptual change?

At this point, we must introduce a mysterious character: the developmental paradox, thus named because of paradoxical findings regarding the abilities of infants and younger children, compared to older children and adults. Assuming a typical developmental trajectory of step-wise or linear improvements in learning, would it not seem at odds if younger age groups appeared successful on tasks that older age groups find difficult? Emerging literature suggests a much younger age of acquisition for mindreading in children (Apperly, 2011), with results from infant studies interpreted as evidence for precocious mindreading ability. Keep in mind the clear and robust results of Wimmer and Perner's (1983) Maxi false belief task (threshold for success between three and four years) as we read about supposed mindreading in these much younger populations.

1.4.1. Violation of expectation

Considering the Maxi task findings, Onishi and Baillargeon's (2005) paper arguing that 15-month-old infants can understand false belief caused controversy amongst mindreading researchers. This study measured infants' looking times through what is known as the violation of expectation (VoE) paradigm. This paradigm can give us more insight into the cognitive processing of infants (Apperly, 2011), and is noticeably different to the Maxi false belief paradigm. The VoE paradigm measures a child's or infant's spontaneous, involuntary response, whereas the false belief paradigm measures a child's elicited, voluntary response. The rationale is that infants tend to look longer at unexpected or 'surprising' events. The VoE paradigm shows infants two contrasting events and determines which one they look longer at. Shorter looking time suggests an expected event, while longer looking time suggests an unexpected event: a violation of the infants' expectation.

In Onishi and Baillargeon's (2005) study, 15-month-old infants watched the following events in which an agent with a visor covering their eyes (in order to avoid infants following eye gaze as a cue) sat in front of a yellow box and a green box. In the familiarisation trials, the agent played briefly with a watermelon toy, and then hid it in one of the two boxes. Then, the agent reached into the box in which she had just hidden the toy, as if to retrieve it. Next, a belief induction trial induced either a true or a false belief in the agent (two separate conditions), regarding the location of the toy. In the true belief condition, the toy moved by its own volition in full view of the agent; thus, the

agent held a true belief about the location of the object. The false belief condition followed the same pattern, except that the agent was absent at a crucial point and did not see the toy move to its final location; hence, the agent held a false belief about the object's location. Finally, in the test trial, the actor reached into either the yellow box or the green box (counterbalanced; Figure 1).





Figure 1 | Test trial events, whereby the agent reaches into the green box or the yellow box. From Onishi & Baillargeon (2005).

According to advocates of the VoE paradigm, infants should look reliably longer at the ending which violates their expectation. Firstly, what expectations do infants have in this scenario? If infants privilege their own knowledge of the toy's actual location, they would expect the agent to always look in the box where the toy truly is (even when the agent did not see the toy move there). However, the authors predicted that if infants are computing the agent's belief, then infants would expect the agent to always look in the box where they believe the toy is located. Secondly, in which conditions would infants' expectations be violated? Based on the authors' position—that infants possess rich mindreading abilities—the following instances would be expected. When the agent believed the toy to be in the yellow box, via true belief (they saw the toy move to the yellow box) or false belief (they did not see the toy was actually moved to the green box), infants should look longer when the agent violates their belief and reaches into the green box. In contrast, when the agent believed the toy to be in the green box—via true belief (they saw the toy move to the green box) or false belief (they did not see the toy was actually moved to the yellow box)—infants should look longer when the agent violates this belief by reaching into the yellow box. To summarise, infants should look longer at endings where the agent's search location is incompatible with their belief about the location of the toy.

Onishi and Baillargeon's (2005) non-verbal task found that infants did indeed look longer in instances where the agent acted inconsistently with their belief. This finding held for both the true belief and false belief conditions, and supposedly shows that infants expect agents to search for an object based on their belief about its location. From this, the authors concluded that 15-month-olds possess a representational theory of mind. It is these controversial findings that lead to the developmental paradox, whereby three-year-olds fail certain false belief tests, yet 15-month-olds seem to possess a rudimentary understanding of false belief. How can this contradiction be reconciled? This is a question that requires us to look at theory. The findings lend themselves to multiple different interpretations, hence the debate sparked among academics about the conclusions drawn by the authors. The authors support an early competence view of children's mindreading development (i.e., rich interpretation). They claim that the nonverbal nature of this task reveals infants' ability to correctly reason about false belief understanding. Note that many similar findings originate from the same laboratory (e.g., Luo & Baillargeon, 2007; Song & Baillargeon, 2008; Song, Onishi, Baillargeon, & Fischer, 2008).

One example of opposition to the rich interpretation of Onishi and Baillargeon's (2005) watermelon study suggests that infants utilise low-level strategies from having learned about behaviour rules (Perner & Ruffman, 2005). Behaviour rules predict future behaviour based on current behaviour, without any reference to mental states. For instance, agents look for an object where they last saw it located. Due to the absence of mental state inference, behaviour rules constitute a lean interpretation of mindreading.

The behaviour rule interpretation effectively bypasses the need for mental states to be considered, making behaviour reading more parsimonious and, therefore, preferred to mindreading. However, a behaviour rules account does not explain the shift in children's mindreading around three to four years of age. Nor does it explain the fact that children talk about mental content, which indicates that mental state reasoning is available to them in some form (Apperly, 2011). At this point, there does not seem to be a theoretical solution to this puzzling finding. Let us look at one more infant study.

1.4.2. Anticipatory looking

Southgate, Senju, & Csibra's (2007) non-verbal task investigated two-year-olds' action anticipation via looking patterns. They used a slightly different paradigm to Onishi and Baillargeon (2005) to tap into false belief attribution, and they did not include a true belief condition.

In their study, 25-month-olds watched the following events in which an agent with a visor covering their eyes sat in front of two boxes, and a puppet moved the target object (a brightly coloured ball) around the scene. In the familiarisation trials, the puppet entered and put the ball in one of the boxes. A combined light and sound cue then signalled that the agent would reach for the ball. These familiarisation trials communicated to participants that the agent had a goal: to retrieve the ball. Next, the test trial showed the puppet moving the ball from its box, but critically, the agent either saw the ball in the left or the right box (depending on condition), before the agent then turned away from the scene (cued by the sound of a phone ringing). Then, the ball was removed from the scene, which the agent did not see happening. Since the participant saw the ball being taken away, they could supposedly compute that the agent now had a false belief regarding the ball's location. The object was removed to avoid participants merely looking towards the box that contained the ball (due to the salience of their own knowledge, i.e., reality bias). The last position of the ball as well as the last position of the agent's attention were counterbalanced. At the end of the test trial the agent turned back towards the boxes, and the light and sound cue happened again (in order to elicit anticipatory looking from participants; Figure 2).



Figure 2 | Test trial event frame, whereby the visored agent turns back to the scene and the two windows above the boxes illuminate to cue anticipatory looking.

From Senju et al. (2007).

Similar to Onishi and Baillargeon's (2005) results, Southgate et al.'s (2007) findings also contribute to the developmental paradox. Twenty-five-month-olds looked in anticipation towards the side congruent with the agent's belief about the ball's location, as if expecting the agent to reach to that box. Participants also spent a significantly longer amount of time looking at the belief-congruent side. The authors took this as evidence for early sensitivity to false belief reasoning in older infants. On the face of it, these results seem incompatible with a conceptual change account. However, Southgate et al.'s suggestion that 25-month-olds (1) correctly attributed a false belief to the agent, and (2) went on to anticipate the agent's behaviour based on this false belief, is only one interpretation of the findings. It is possible that infants merely employed a rule—devoid of any belief inference—that agents tend to search for an object in the last place they saw it located (Perner & Ruffman, 2005).

1.4.3. Relevance of infant studies for conceptual change

The above studies pose a challenge to the idea of conceptual change: it seems the previously held notion that false belief understanding emerges around three to four years of age is called into question by these findings with very young participants. Such studies reveal the age gap between apparent sensitivity to mindreading concepts, like false belief, on indirect behavioural measures and explicit judgement measures. The

question then becomes: what are the underlying mechanisms that enable this supposed sensitivity to mental states in infancy and early childhood?

There are certain things a strong scientific theory must explain and predict. For instance, a theory of social cognition should be able to explain the hitherto unexplained developmental paradox regarding false belief understanding (Lavelle, 2018). Ideally it should also generate testable predictions. Rather than trying to reconcile why purported innate belief concepts need executive control in children but not in infants, the parsimonious explanation is that the mindreading abilities of these age groups are not the same (Apperly, 2011). Infants may have distinct mindreading abilities, compared to older children and adults. Next, we discover a promising candidate theory.

1.5. Two-systems account to the rescue?

The early competence view has come under scrutiny for being a single-system approach. Because of this, dual-system accounts—along with other alternative accounts for false-belief reasoning (De Bruin & Newen, 2014; Fenici, 2015)—have been proposed as better alternatives. In particular, there is an increasingly popular dual-system theoretical account that might be able to accommodate both sets of findings detailed above (i.e., a theory that reconciles the developmental paradox whereby young children do not pass explicit false belief tasks, yet infants show positive evidence for mindreading through implicit measures like eye gaze). The theory in question is the two-systems account of mindreading, proposed by Apperly and Butterfill (2009).

Two-systems proponents distinguish between two processes involved in understanding others' minds: one of these is extremely fast, that is, efficient enough to guide where we look in anticipation of someone's forthcoming behaviour; the other is slower, but flexible enough to help with the particular way in which someone thinks about something. There must be a trade-off between efficiency and flexibility, because of the demand for resources. The efficient, or minimal, system is referred to as System 1, and the flexible system as System 2. Apperly (2011) characterises System 1 as relatively automatic in that it resists strategic control. Processing under this low-level efficient system may also be unconscious or implicit, occurring outside of human awareness. This

early-developing system guides anticipations and is set to track the relational state of 'registration' as a proxy for representing others' beliefs. In contrast, System 2 is much more cognitively demanding and guides direct responses like predictions. This highlevel, later-developing system is able to represent propositional states as such, and can handle complex belief and desire reasoning.

Apart from the developmental paradox, another mystery the two-systems account aims to elucidate is the difference in speed within adult mindreading; sometimes it is fast and other times slow. What else could explain the contradiction between seemingly distinct forms of mindreading—such as the rapidly updated perspective judgements made by players in a high-speed sports game versus jury members carefully deliberating over a court verdict—if not the existence of two mindreading systems (Apperly, 2011)?. Daniel Kahneman's well-known book 'Thinking, Fast and Slow' (2012) also references dual systems. Kahneman mentions how the second system may, at first glance, seem more impressive and garner more attention due to the complex feats it can manage, but the often overlooked first system is secretly the star of the show (the underdog, if you will). The same can be said of two-systems mindreading: while higher-level System 2 surely plays a critical role in mindreading in being able to flexibly represent propositional states, it is also limited, hence the importance of lower-level System 1 mindreading that develops earlier. Apperly (2011) points out the general assumption that high-level mindreading is responsible for the bulk of work in our social lives. In situations that are fast moving which require rapid decisions, it seems more likely that lower-level mindreading, along with knowledge we pick up from our social environments, underlies most of the heavy lifting regarding social cognition. This idea prefaces the focus of the current thesis on System 1 rather than System 2 mindreading.

1.5.1. An answer to the paradox? Signature limits

Importantly, Apperly and Butterfill (2009) urge mindreading scholars to not only focus on what infants *can* do, but equally on the limits of infants' abilities (e.g., consider Wellman et al., 2001). Following on from this, evidence for the two-systems account can

be acquired through investigating what are known as signature limits (much like blind spots) of the early-developing efficient system.

A system's signature limit refers to a unique and defective way the system behaves (Butterfill, 2020). It is unique in that the signature limit is peculiar to that system (e.g., for the two-systems account only System 1 is proposed to have signature limits, not System 2). Signature limits do not include conditions whereby most systems fail (as these are not unique to a particular system). Regarding defectiveness, a signature limit is something a system should plausibly be able to do, but cannot do. For instance, since the two-systems account is a theory for mindreading, a signature limit would be an aspect of mindreading in which System 1 consistently makes mistakes.

Butterfill (2020) describes how signature limits can be used to discern whether effects or results are coming from the same system. For instance, if both adults and children pass a mindreading task, there are many strategies both groups could be making use of, therefore this positive result does not inform you about the system/s they might be using. However, if both adults and children *fail* a mindreading task in a specific way, this evidences that both groups might be using the same system – and that system has a signature limit.

Since System 1 is theorised to be relatively encapsulated and more rudimentary in being set to track relational states, rather than propositional states, there are instances in which we might expect the efficient system to produce errors. For instance, since System 1 guides eye movements, we might expect erroneous eye movements on a task that is very complex, say, a task involving understanding of false beliefs over an object's identity. To illustrate, let us review Low and Watts' (2013) study.

1.5.2. Evidence for two-systems account

In this study by Low and Watts (2013), children (3- and 4-year-olds) and adults watched familiarisation trials and a belief induction trial, followed by a critical test trial – much like the previous infant studies mentioned (Onishi & Baillargeon, 2005; Southgate et al., 2007). There was an agent wearing a visor who had access to two boxes

in front of them (one to the left side and one to the right). Familiarisation trials conveyed that the agent preferred either a red toy or a blue toy. After this, the main stimulus object was introduced for the belief induction trial. This object was a toy that cleverly tapped appreciation of object identity, whereby one side was painted red and the other blue (Figure 3). In this way, if you only see one side or aspect of the toy, you assume it is an entirely blue toy (or an entirely red toy). Following on from this, the single toy can be seen to have two identities, although they exist within one physical object (i.e., it is both a red toy and a blue toy at the same time).



Figure 3 | Dog-robot stimulus toy that has one red side and one blue side. From Low and Watts (2013).

Importantly, participants saw this dual aspect toy rotating, therefore revealing both the red and the blue sides, whereas the agent was not party to this reveal. With the agent back on the scene, the toy moved between the two boxes, but turned 180° while hidden behind a box. As a viewer we see that it is merely the same toy rotated around, but to the agent it would appear as two different toys (e.g., a red toy went into the left box, but then a separate blue toy emerges from the left box and moves into the right box). From the agent's perspective, there are two distinct and differently coloured toys. Thus, the agent has a false belief regarding object identity. Using a combined light and sound cue (as did Southgate et al., 2007), participants were then cued that the agent was about to reach for a toy. This elicited anticipatory looking responses. In this identity task, Low and Watts (2013) found that, for all age groups, participants tended to look first in anticipation towards the incorrect location. The study also included a change-of-location false belief task (less demanding than the change-ofidentity false belief task described above). In both the true and false belief conditions of this location task, participants of all ages mostly looked in the correct location. So, the authors found striking differences between the location task and the identity task. The two-systems account maintains that the efficient system can deal with location-change tasks, but only the flexible system can deal with the complexity of tasks where an object changes identity (see also: Oktay-Gür, Schultz, & Rakoczy, 2018). Ultimately, Low & Watts' study showed a striking anticipation-prediction dissociation, and the twosystems account seems to provide the most viable explanation for this puzzling finding. Let us unpack this dissociation further.

When required to track how another person sees an object, the anticipatory looking of both children and adults does not seem to reveal false belief attribution. However, in the very same task, older children (four-year-olds) and adults can successfully verbally predict the agent's behaviour based on false belief attribution. This is called an anticipation-prediction dissociation, because anticipatory eye movements are in opposition to verbal answers (predictions). On the one hand, visual responses do not seem to grasp the agent's false belief (i.e., younger children look in the wrong location), but on the other hand explicit answers do, (i.e., younger children can say the correct location when asked).

This paradoxical finding can be explained by the two-systems account. The efficient system is responsible for certain anticipatory eye movements and cannot process belief as such, whereas the flexible system is responsible for verbal predictions and can process belief ascription; thus, eye movements seem to fail the task while verbal answers succeed. Three-year-olds actually fail on both types of responses and, according to the two-systems account, this makes sense because at this age young children have not yet developed a flexible mindreading system. Going back to Wimmer and Perner's (1983) original Maxi task, the two-systems account maintains that this is also the reason behind three-year-olds failing the task and four-year-olds passing.

In addition to Low and Watts' (2013) study detailed above, further research has found dissociations and reverse dissociations in the anticipations and predictions of participants across different mindreading tasks (Fizke, Butterfill, can de Loo, Reindl, & Rakoczy, 2017; Low, Drummond, Walmsley, & Wang, 2014) and across cultures (Wang, Hadi, & Low, 2015). The anticipation-prediction dissociation seems to reveal genuine signature limits on System 1 mindreading. Further, two-systems advocates maintain that infants only possess the efficient system, which gives them some early mindreading ability, but full-blown mindreading is only possible upon attaining the flexible system as children get older (Lavelle, 2018). This seems a reasonable solution to the developmental paradox. Recently, there have even been suggestions that System 1 may be partly grounded in the motor system (Low, Edwards, & Butterfill, 2020; Zani, Butterfill, & Low, 2020). This will be elaborated upon further in section 4.4.

1.6. A crisis of replication

A proverbial elephant in the room that must be acknowledged is the current replication crisis in psychology (among other disciplines; Figure 4), whereby many findings are failing to replicate in different laboratories. These recent non-replications, specifically within the mindreading field, have cast doubt on current theoretical explanations of implicit and explicit mindreading phenomena, especially in children (e.g., see Schuwerk, Priewasser, Sodian, & Perner, 2018 regarding non-verbal looking paradigms testing false belief sensitivity). Researchers have questioned the robustness and reliability of implicit mindreading findings, with Kulke, Von Duhn, Schneider, and Rakoczy (2018b) noting the small sample sizes of many studies. Kulke et al. also note that, for most explicit mindreading tasks, there are stringent control trials to rule out alternative explanations for findings; meanwhile, this is not always the case for implicit tasks. Ganglmayer, Attig, Daum, and Paulus (2019) also caution against publication biases and false positive findings.



Figure 4 | The elephant in the room: Replication Crisis.

A growing number of studies are reporting failures to replicate (directly as well as conceptually) high-profile findings of infants' and toddlers' (and in some cases adults') apparently abstract and sophisticated anticipations and expectations of others' desirebased and belief-based actions (e.g., see Burnside, Ruel, Azar, & Poulin-Dubois, 2018; Kulke, Reiß, Krist, & Rakoczy, 2018a; Powell, Hobbs, Bardis, Carey, & Saxe, 2018; Ruffman, Aitken, Wilson, Puri, & Taumoepeau, 2018 appearing in a Special Issue in Cognitive Development addressing the replication crisis facing research on early mindreading). Let us look at some studies, in particular, that have attempted to replicate mindreading research.

1.6.1. The problem is task replicability

This chapter included false belief, since the subject is central to any historical overview of the mindreading field. Despite this focus on false belief test trials with infants (which some researchers argue is evidence for precocious mindreading ability), it is rather the preceding familiarisation trials that are critically important in order for the test results to be meaningful. Familiarisation trials are essentially about action observation and goal-tracking. They check that infants track others' goal-directed actions, for instance, that the agent will produce actions directed to an outcome (grasping the toy) or, more specifically, that the agent will search in the box where the

toy is placed. Familiarisation trials have not received much attention and were assumed to be doing their job – until now. These trials, which ultimately set up the whole task for participants, have recently also come under fire. Many studies (both originals and replications) have very large exclusion rates (in the range of 30-50%) owing to low anticipatory looking in baseline familiarisation trials, even before any false belief has been introduced. This is problematic. One effect is that the number of participants included in a paper's final analyses dwindles, but a deeper question to ask ourselves as we read on is *why* do participants struggle to show anticipatory looking in these basic, baseline trials?

Kulke, Johannsen, and Rakoczy (2019) were interested in implicit mindreading tasks. Implicit tasks, such as VoE paradigms or anticipatory looking measures, do not require explicit verbal instruction or response. The authors reviewed evidence for mentalising accounts (rich interpretation) versus submentalising accounts (lean interpretation). Attempts to systematically replicate implicit tasks using anticipatory looking have returned mixed findings: failed replications, partial replications, and some successful replications. Kulke, Johannsen, and Rakoczy aimed to figure out why.

Two types of anticipatory looking false belief tasks have been employed in many replications: transfer tasks and removal tasks. Since removal tasks tend to replicate but transfer tasks yield mixed results, Kulke, Johannsen, and Rakoczy (2019) wanted to understand whether there is a conceptual difference between these types of tasks. On the one hand, transfer tasks are like the standard Maxi task in which an object changes location (or is transferred), but in this implicit version participants are not asked a question or required to give a verbal answer, rather the task simply measures participants' anticipatory looking at the end of the scene. For instance, when Maxi returns to retrieve his chocolate, do participants look to the location where it truly is, or to the other location where Maxi believes it to be? On the other hand, removal tasks involve the target object being removed from the scene instead of being transferred to another location. Participants' anticipatory looking is measured at the end of the scene when the agent is about to retrieve the object.

Kulke, Johannsen, and Rakoczy's (2019) study investigated whether mere objecttracking was at play, by creating a novel true belief control condition whereby the agent views all parts of the scene, including when the object is removed. Based on this novel condition, there are diverging predictions regarding anticipatory looking. If infants are using a low-level process such as simply looking at the last location of the object, this should be reflected in their looking patterns. However, if infants are in fact tracking the agent's belief, the true belief condition should yield chance results whereby infants look equally at both locations; this is because the agent knows the object has been removed from the scene (and as such has no reason to search either box), so there should be no systematic pattern of anticipatory looking in this case.

The first and second studies tested adults and children, and the stimuli were closely based on Southgate et al.'s (2007) action anticipation study mentioned earlier. The aim here was to test for an alternative explanation to belief-tracking, that is, do participants simply track the last location of the object in a low-level manner, rather than tracking belief? Eye-tracking software measured participants' looking patterns. Notably, the first set of familiarisation trials simply had the ball (target object) visibly sitting atop one of the two boxes, followed by the light and sound cue after which the agent reaches for the ball. It is not unreasonable to assume this basic sequence as being rather easy to follow. The novel true belief condition was implemented as a control.

Adults' and children's anticipatory looking patterns did not generally reveal preference for the belief-congruent location, thereby failing to replicate the original findings of Southgate et al. (2007). The exclusion rates for these two studies, based on inclusion criteria from the original study, were high (at 35-44% across the studies, compared to 44% in Southgate et al., 2007). Since the excluded participants did not show action anticipation in the last familiarisation, the third study altered the familiarisation trials to try and improve upon this.

The third study used animated videos instead of real-life recordings of human actors, to eliminate any confounding movement or visual differences between conditions. The design also aimed to rule out a potential confound caused by the agent directionally turning their head away when the phone rings (cueing effect). The stimuli

were again based on Southgate et al.'s (2007), but differed in a few ways including the following: (1) instead of a puppet moving the object, the object was a small animal that moved by itself, and (2) instead of being distracted by a phone ringing during the change of object location, the agent leaves the scene through a central door (to avoid directional cues). The familiarisation criteria involved correct anticipatory looking in a minimum of two of the last three familiarisation trials.

Much as before, anticipatory looking patterns did not reveal preference for the belief-congruent location. Further, despite the more tightly controlled stimulus material, this study also had high exclusion rates (52%), mostly due to failure on the familiarisation criteria (more than 51%!). Compare this again to the dropout rates of the original Southgate et al. (2007) study: 35% directly due to failed familiarisation trials, 44% overall. Kulke, Johannsen, and Rakoczy (2019) describe the results as inconclusive, complex, ambiguous, and difficult to interpret, with neither low-level object-tracking nor belief-tracking apparent in participants' anticipatory looking.

Overall, Kulke, Johannsen, and Rakoczy's (2019) recent conceptual replications of Southgate et al.'s (2007) study largely did not replicate the original findings. Taking their three studies together, anticipatory looking did not reveal consistent patterns in either familiarisation trials or test trials. Ultimately, this study—among the growing number of non-replications of false belief tasks that use anticipatory looking—begs the question: do these tasks actually tap into some such thing as implicit mindreading? And at the very least, do the familiarisation trials achieve their aim of communicating the agent's actions and goals to the participant?

1.6.2. Another blow to infant mindreading

A multi-laboratory study by Kampis, Karman, Csibra, Southgate, and Hernik (2020) aimed to directly (rather than conceptually) replicate the original Southgate et al. (2007) study (whereby infants watched a puppet move a ball while an agent viewed the scene with a goal of retrieving the ball). Note that this replication involves authors of the original paper. An important reason to replicate this study, in particular, is that it has been used as evidence for the existence of high-level belief-tracking in infants. Such controversial theoretical claims should have empirical backing, otherwise they run the risk of taking the mindreading field in an unfruitful direction – one not backed up by a rigorous scientific process. Since this was a direct replication, the method followed the original study in eliciting action anticipation, as well as following the analyses as closely as possible. One difference of note is that the sample size was sufficiently increased (about 2.5 times greater than the original sample).

Kampis et al.'s (2020) results also failed to replicate the original findings; infants did not show a looking preference to the correct location over the incorrect location. So, even using the original stimuli and involving the original researchers, Kampis et al. did not find goal-based action anticipation in two-year-olds, let alone belief-based action anticipation. Their study was also plagued with high dropout rates due to the exclusion criteria for the familiarisation trials (31%), as well as other criteria like inattention (total dropout rate: 69%). Overall, the authors suggest that Southgate et al.'s paradigm is therefore not reliable for eliciting action prediction or false belief understanding. Despite the null results, Kampis et al. adopt the view that failed replications should not necessarily be taken as evidence that infants cannot mindread at all. However, although these results do not negate the existence of implicit mindreading, they do show that the paradigm does not seem to elicit the behaviour for which it was created in the first place (i.e., action anticipation), and for which it has been widely cited (i.e., false belief attribution in infants).

1.6.3. Grim outlook calls for paring back to basics

As section 1.6. has shown, it can be very difficult getting infants and children to pass familiarisation trials as well as test trials. Not only do the recent Kulke, Johannsen, and Rakoczy (2019) and Kampis et al. (2020) studies show that the original false belief-tracking cannot be reliably replicated, but also that it is challenging to secure even the basic action- or goal-tracking response. Yet another recent study from Kulke's group found anticipatory looking behaviour—even in familiarisation trials—problematic to replicate (see Kulke & Hinrichs, 2021 study with real-life stimuli rather than videos). What are the implications of this for the many tasks and paradigms that have made use
of action- and goal-tracking? Can we rely on results from these tasks? These replication failures warrant taking an in-depth look at action observation and goal-tracking, even with adults let alone infants; hence, this will be a major focus taken up in the next chapters. Relatedly, Schuwerk et al. (2021) plan to investigate knowledge- and ignorance-ascription via anticipatory looking in a multi-laboratory study, but this thesis will focus more on the foundational areas of actions and goals.

1.7. Conclusion

The mindreading field depends on assuming that action understanding and goaltracking can be done in some spontaneous or efficient manner. With the historical overfocus on false belief understanding as constituting the basis of all mindreading, we must now probe lower-level processes that may fall under the remit of System 1 of Apperly and Butterfill's (2009) two-systems mindreading account. The next two chapters will look at actions and goals (Chapter 2), followed by a motor theory for goaltracking (Chapter 3).

Chapter 2.

2.1. Introduction

The previous chapter left us with the message that despite the focus on mental states, like false belief, mindreading may in fact depend on rapid action understanding and goal-tracking. Chapter 2 will be action-packed; this chapter will show us how infants can help us theoretically map out the importance of action understanding. The difference between tracking goals and knowing goals is discussed, and some infant goal-tracking studies are highlighted. After this, the concept of pure goal-tracking is introduced (a lean, rather than rich, type of goal-tracking). Then, teleological reasoning, a model of action interpretation which may be how humans track goals, is explained. Studies are also showcased that offer another possibility; perhaps we use statistical information to track goals, too. Finally, an interesting phenomenon is considered, whereby one's actions seem to relate to one's own ability to mindread. Chapter 3 will then segue into goal-tracking in more depth.

One of the main purposes of mindreading is the interpretation and prediction of our own actions, as well as the actions of others (Keysar, Lin, & Barr, 2003). We do not simply see others' actions as movements of a physical nature; instead, we see them as movements structured by goals (Woodward & Cannon, 2013). Further, many social interactions require the rapid generation of online predictions about others' actions and intentions (Cannon & Woodward, 2012). But how is it that we interpret actions? There are two ways in which to specify an action (Butterfill, 2020). Firstly, it can be in terms of specific bodily configuration and joint displacement, such as your fingers pre-shaping in order to grasp the handle of a mug. Secondly, it can be in terms of the broader action goal, such as to reach and grasp a mug. Remember also, that actions with different physical movements can share the same goal. This chapter will offer suggestions as to what the underlying mechanisms may be, through which infants generate predictions of others' actions.

2.2. To track or to know?

Understanding goals is essential in order to subsequently understand people's thoughts and actions. The current thesis will define the goal of an action as the outcome to which that action is directed (Butterfill, 2020). Importantly, goals are not mental states or intentions. Given this definition, it follows that to understand a goal, one understands the outcome of the relevant purposive action (because the outcome effectively *is* the goal). However, tracking the goal of an action is different to fully knowing about the goal. Goal-tracking simply requires recognition that outcomes are dependably goals of action, but knowing or representing information about the goal is not necessary. Butterfill suggests that, theoretically, one can track the goal of an action without factual knowledge of the goal itself. Tracking something, such as an object or a goal, is a leaner process that does not necessarily involve representation or knowledge of that item. With this in mind, the two-systems account would suggest that young infants can indeed track the goals of an action, but that this process may not require the infants to have knowledge of those goals per se.

There are two issues to consider: (1) whether up to 12-month-old infants track agents' intentions as well as their goals, and (2) whether it is possible to track goals without any knowledge of mental states. Regarding the first issue, Butterfill (2020) argues that research confounds intention-tracking with goal-tracking, leading to a dearth of experimental data to test the distinction between the two. Regarding the second issue, it is notable that even simple actions can be made up of many different goals, for instance, grasping to hold a mug, lifting it to reach your mouth, tilting the mug precisely to drink the contents, and so on. Even though the general goal may be to have a drink, all the embedded smaller goals help you reach this overarching goal.

If it is the case that all goal-tracking requires mental state information, this process will be complex and demanding. However, if goal-tracking can occur in the absence of mental state information, it seems more plausible that we first gain the ability to track smaller actions. In turn, this may teach us about larger action sequences, which we then go on to eventually understand in a richer way by ascribing intentions to agents. It seems goal-tracking may be a foundation for mental state ascription (Butterfill, 2020).

2.3. Infants track action goals

The inclination towards viewing others' actions in a goal-directed manner develops early on (Woodward & Cannon, 2013). The pathway through which infants begin to understand goal-directedness is by realising that actions and goals tend to belong to a hierarchy of sorts, whereby actions are the means to attaining an end goal. That is, infants understand the means-end structure of actions, and that actions are often goal-directed (Woodward & Cannon, 2013).

Returning briefly to Southgate et al.'s (2007) study (mentioned in section 1.4.2., whereby action anticipation was tested via an anticipatory looking paradigm), two-yearolds predictively looked towards the location where the agent last saw the toy. Replication problems aside, the authors interpreted their findings as showing false belief sensitivity in young children. In this instance, children's anticipatory gazes did not reflect learning of the agent's specific movements, because previously—across the familiarisation trials—the agent retrieved the target object *equally* as often from the two locations. This reveals that infants were not simply learning a basic pattern, but were rather analysing the agent's goals. Therefore, this suggests children generate their predictions (e.g., where an agent will reach towards) based on the prior goals of the agent (Woodward & Cannon, 2013), and shows that two-year-olds engage in online action prediction. Are younger infants also capable of such online predictions?

2.3.1. Habituation (looking-time) studies

Studies using looking-time paradigms show that even infants can represent the actions of others as object-directed and goal-directed, as opposed to merely being physical movements in space (Woodward & Cannon, 2013). Gergely, Nádasdy, Csibra, and Biró (1995) conducted a habituation study on infant goal-sensitivity. In one condition, infants were first habituated to a rational approach scene, in which a small circle jumps over a rectangular obstacle and contacts a large circle on the other side. If

infants interpreted this habituation event as a rational agent acting in a goal-directed way, then we might expect them to predict that agent's future behaviour. After habituation came the test events, in which the rectangular block was removed from the scene. Infants saw either (1) a new action, whereby the small circle went straight (horizontally) towards the large circle, or (2) an old action, whereby the small circle made the same jumping motion as in habituation and moved towards the large circle (Figure 5).



Figure 5 | Test events, whereby the rectangular block was removed and infants saw the new action or the old action.

From Gergely et al. (1995).

The authors hypothesised that—in line with goal-sensitivity—infants would show more interest to the old action test event, since it is not rational for the agent to still jump in the absence of an obstacle. Conversely, if infants merely represent the perceptual similarity of action movements, the opposite pattern would be expected; infants would show more interest to the new action test event, since the trajectory has changed from a jumping motion to a horizontal straight path of movement. They found that 12-month-olds were more interested in the old action test event, despite this involving no change in movement trajectory. In contrast, infants were less interested in the new action test event, even though they had previously never viewed the horizontal movement trajectory. From these results, Gergely et al. (1995) claimed that one-yearolds can interpret goal-directed behaviour of rational agents. Note that this study used circles to stand in for rational agents. Although there is precedence to use geometric shapes, as even adults attribute motives to shapes (Heider & Simmel, 1994), it is important to review another important study.

Woodward (1998) also conducted a habituation study on goal-sensitivity in infants, which has been widely replicated. Similarly to Gergely et al. (1995), Woodward's study compared changes in movement trajectory with changes in goal. Responses were then measured to determine which change interested infants, by determining which change they dishabituated to more strongly. To break down the procedure, infants were first habituated to an agent reaching for and grasping one of two toys that were situated next to each other (ball and teddy bear; Figure 6). For the test events, the toys were swapped around (i.e., so that the one previously on the left was now on the right, and vice versa). Infants observed either a trajectory change or a goal change. In the trajectory change, the agent performs a new trajectory (towards a different side than shown in habituation), but since the objects have been switched, they are reaching towards the same toy as before. In the goal change, the agent performs the same trajectory (towards the same side as that shown in habituation), but because of the objects switching places, the agent is now reaching towards a new toy, that is, a new goal. Put simply, if infants are concerned with movements not goals, they should find the new trajectory more interesting. But, if infants track goals of actions, they may be more interested in the new goal. The latter is what Woodward (1998) found; six- and nine-month-olds dishabituated more strongly (i.e., looked longer) to a new goal than to a new trajectory.



Figure 6 | Depiction of events in the hand condition.

From Woodward (1998).

When the same procedure was carried out in another condition, with an inanimate object (mechanical claw; Figure 7) instead of the human agent's hand, this pattern was not found, that is, infants did not look longer at the goal change when the observed entity was the inanimate claw (non-human agent; Woodward & Cannon, 2013). Infants do not spontaneously view a mechanical claw moving towards and contacting an object as a goal-directed action. However, when the same movement is made by a human agent's arm, infants do regard this as goal-directed. This is despite the movement trajectories being the same for both conditions. This indicates goal-tracking processes that are distinct to those used for tracking inanimate objects. It is important to be mindful of other interpretations. In the familiarisation phase of the Woodward (1998) study, infants were repeatedly shown a relationship between the agent and an object. Infants may be expecting the same learnt pairing in the test phase and will therefore look long when there is a new agent-object pairing (Ruffman, Taumoepeau, & Perkins, 2012). Similarly, infants will be more likely to generalise certain associations to familiar (biological) agents than to less familiar (mechanical) agents.



Figure 7 | Depiction of events in the claw condition.

From Woodward (1998).

It makes sense that since human beings are social creatures, tracking the goals of other humans' actions is necessary for most social cognition. From this study we can see that goal-tracking behaviour only occurs for bodily movements of a minded agent (like a human hand reaching and grasping), as opposed to the same action from something nonbodily (like a mechanical claw reaching and grasping). This would suggest that even infants engage in active information processing when predicting the actions of others. However, the point that perceived mindedness of the agent affects responses is contentious, considering that Gergely et al.'s (1995) self-propelled ball stimuli seemed to elicit goal-tracking in 12-month-olds.

2.3.2. Eye gaze: from looking longer to looking in anticipation

Woodward and Cannon (2013) raise the issue that looking-time data on its own cannot clarify whether infants' predictions can be generated rapidly enough to contribute to everyday, online social interactions (such as communication, working together, competing with others etc.). Let us now review a study that used infants' anticipatory looking (also known as predictive gaze) to remedy the timescale issue.

Cannon and Woodward (2012) have since tested an adaptation of Woodward's 1998 paradigm with 11-month-olds, that aimed to untangle the confound between goal and pattern of movement (whereby the same movement pattern is always directed to the same goal). Using predictive gaze instead of dishabituation, this study showed only an incomplete movement in the final trial (the hand or claw stopped halfway between the two objects without making contact), thereby rendering participants' gaze towards the objects as anticipatory or predictive. Cannon and Woodward found that infants predictively looked at the goal object in the hand (human agent) condition more so than they did in the claw (mechanical) condition, that is, they predicted the agent would reach towards the familiarised object that was now in a different location. However, infants did not display goal-directed action anticipation when watching the mechanical claw. From this, the authors concluded that infants recruit goal information when predicting others' small-scale actions, and, similarly to adults, infants proactively gaze to the endpoint of an agent's action.

There are also critics of the Woodward paradigm. Ganglmayer et al. (2019) conducted a multi-laboratory replication of Cannon and Woodward's (2012) study to test the idea that action anticipation relies on goal understanding. This included direct and conceptual replication attempts. By nature, the direct replication followed the original study. The conceptual replication combined Cannon and Woodward's paradigm with a paradigm by Daum, Attig, Gunawan, Prinz, and Gredebäck (2012), whose study will be elaborated on later in this chapter (in section 2.7.). Briefly, Daum et al. familiarised participants to an agent walking along a path that forked into two separate paths, after which they introduced an opaque occluder that the agent disappears behind whilst walking. This was done to trigger anticipatory eye movements from participants, towards the direction they expected the agent to reappear from.

Neither the conceptual nor the direct replication attempts were successful. The authors found that infants anticipated the agent's actions based on location information more than they did by using information about the goal object. Ganglmayer et al. (2019) suggest that rather than flexibly anticipating action goals, infants are merely anticipating patterns and paths of action. For instance, infants could encode movement paths by using visuo-spatial information and heuristics, like movement trajectories. Such an interpretation supports a low-level account of infant mindreading regarding goals, as it also indicates that infants may be using frequency information as a basis for

action processing and prediction (see also Ruffman et al., 2012). That said, action anticipation is not always reliable in younger infants. Falck-Ytter, Gredebäck, & Hofsten (2006) showed 6-month-olds an agent placing different objects in a bucket, but infants failed to generate anticipatory looks to the bucket before the ball was placed in it.

2.4. Goal-tracking 'au naturale'

Despite Ganglmayer et al.'s (2019) criticisms, infants may be sensitive to goals in some way, but how would this process of goal-tracking occur? Tracking goals does not necessarily involve ascribing mental states to the agents committing the actions.

As mentioned earlier in this chapter (section 2.2.), if goal-tracking required mental state information it would be very resource-heavy. It would be more plausible that goal-tracking can occur without mental state input. Butterfill (2020) puts forth the idea of 'pure' goal-tracking as exactly that: goal-tracking that is void of any mental state ascription. Although, if ascribing intention is necessary for goal-tracking, then pure goal-tracking cannot exist. Next, we take a look at a rational action model that may bolster the case for pure goal-tracking.

2.5. Teleology: the rational way

Teleology is a purpose-based model of action interpretation, whereby agents are expected to act rationally in order to achieve goals. A teleological framework assumes the absence of mental states, which include beliefs, desires, intentions, and emotions. A goal is not considered a mental state. According to this framework, we track goals by making inferences from principles about what we know or believe. If infants apply teleological principles, they will realise facts about the goals of actions. Teleology maintains that the way in which we gain factual knowledge about the goals of action has nothing to do with language or rich social interaction; it draws on innate abilities and experience only.

Csibra and Gergely's (1998) rational action principle (that forms the basis of teleological reasoning) suggests a possible way for goal-tracking to be achieved: in order

for an outcome to be the potential goal of an action, the way in which the action is conducted must be the most rational way for the outcome to be achieved, given any constraints. This strategy can be used to exclude possibilities, such as if they are not the most efficient way for something to be achieved. To illustrate, let us consider two locations A and B. If someone is walking a path that seems a shortcut to B, using this principle we can assume the goal of the action is reaching location B, not location A. Walking the path to location A would have taken longer and there did not appear to be any constraints on the agent. Applying this principle is known as the teleological stance.

This principle of rational action does not always hold true, but in certain situations it can be useful to track the goal of an action without the need for mental state ascription; therefore, this is a case in which pure goal ascription is theoretically possible. However, a limitation of the teleological stance is that it can only lead to accurate goaltracking when the action is successful, as opposed to an action that fails. To illustrate, if you intend to take a step but accidentally stub your toe, your goal was never to stub your toe, but this is technically what the teleological stance would imply. Following the same reasoning, a better way to take a step was available and you did not take it, so this is excluded as a potential goal. In addition, no better way to stub your toe was available, so this remains a candidate goal of the action. Hence, the teleological stance is limited to successful actions, and cannot be informative for unintended or accidental actions. This limit could be overcome by adding to the rationality principle, for instance, by adding that potential outcomes that are candidate goals of an action are to be *desirable* to the agent. Such an addition would discount hurting oneself (by accident) as a goal.

2.5.1. The rational way is not the only way

Empirically speaking, teleology does not always come out on top as the type of reasoning used when predicting others' actions, in particular with younger age groups (conflicting with a strong nativism view). A recent study by Gönül and Paulus (2021) investigated the age at which children begin to predict others' actions based on efficiency. Specifically, they focused on young children's explicit predictions, rather than implicit measures such as action anticipation. The authors refer to this explicit form of prediction as *rational action prediction*, which falls in line with teleology. One of the main reasons as to this study's theoretical importance is this: if a discrepancy is found between implicit and explicit measures, this may have implications for theories of early social cognition. For instance, it might be that Apperly and Butterfill's (2009) two-systems account, which has predominantly focused on representational differences between implicit and explicit false belief ascriptions, could also extend to the highlighting of representational differences between implicit and explicit for the predictions.

Gönül and Paulus (2021) raise the issue that research on efficient action prediction has mainly focused on implicit measures, leading to a dearth in relevant research using explicit measures. Hence, the authors' use of explicit reasoning to investigate the teleological stance is a useful approach. It is interesting to note that this is the opposite to the historical development of the mindreading field, which we have seen (in section 1.3.) began with a heavy focus on explicit theory of mind reasoning, at the expense of investigating implicit reasoning. In fact, infant mindreading scholars are currently debating whether infants have precocious mindreading abilities, whether they possess the first system of a two-systems account, or whether mindreading does not even exist in infancy, and can rather be explained away by leaner processes such as statistical learning (see Ruffman et al., 2012). With such arguments regarding implicit measures in the related field of mindreading, it does the teleology field well to branch out to include explicit measures, in order to avoid drawing conclusions or developing theories from implicit measures alone.

While implicit measures and explicit measures of rational action are clearly distinct from one another, what is not clear is how these measures actually relate to the underlying representations of the observed actions. Does rational action understanding develop early? Do implicit representations develop first, followed by explicit reasoning? The authors point out how in early development, younger children possess the precocious ability to understand separate components of action prediction (e.g., the goal state, the situational constraint, the target etc.), but what changes with age is that older children come to grasp part-whole hierarchical relationships (i.e., they appreciate how

each component, when combined, will account for someone's reasons for action). According to this framework, as children go through development, their representations explain action in progressively more detail. Gönül and Paulus (2021) prefer this theoretical framework because of the gradual progression in explicit detail, as opposed to the more typical binary distinction between implicit and explicit. On this view, lowlevel action abilities can gradually transform to higher-level conceptual understanding. Rather than being mutually exclusive, these processes may in fact interact along the developmental trajectory.

In Gönül and Paulus' (2021) study, participants (3- to 6-year-olds) saw brief vignettes depicting an agent on a path at a crossroad. There were two possible directions the agent could take; one was a shorter and therefore more efficient path, the other was longer and thus less efficient. Children's explicit predictions (i.e., verbal answers) were measured, in response to being asked which path they reasoned the agent would take. Gönül and Paulus found that—when anticipating whether an agent would take a shorter efficient path, or a longer inefficient path—five- and six-year-olds reasoned about action efficiency, but three-year-olds did not. Four years appeared to be a transitional age. The older children predicted that the agent would take the shorter path, while the youngest children did not show a systematic prediction towards this more rational path. The authors even manipulated whether or not a goal was presented, and found that having a clear goal target only facilitated older children's post-hoc reasoning (more specifically, only when the agent was shown to take the shortest path to the goal item and participants had to explain why).

This points to rational action prediction, as part of the teleological stance, being something that emerges throughout the development of the preschool years and that does not emerge before four years of age. Similarly to mindreading, the preschool years appear to be an important developmental period for teleology. Gönül and Paulus even suggest that the parallel between their findings and those of explicit mindreading tasks (e.g., false belief tasks) could mean that the two-systems account (Apperly & Butterfill, 2009) may also hold for the way in which teleological reasoning develops. In sum, this

study highlights that efficiency and teleological reasoning are not the primary way in which young children understand and predict others' actions.

2.6. Infants do statistics

As we have seen, Csibra and Gergeley's (1998) teleological stance demonstrates the possibility for pure goal-tracking, but there remain questions about the descriptive and explanatory adequacy of this stance. Focusing in on descriptive adequacy, there are some situations in which the teleological stance generates correct predictions about others' actions, but other situations when it does not. For instance, routines and statistical regularities also help us predict actions without a deeper understanding of the goals of the action. To illustrate how infants can come to learn predictable patterns of movement, let us review a study by Paulus et al. (2011).

Paulus et al. (2011) investigated how statistical learning (via frequency information) and teleological reasoning contribute to action prediction in nine-montholds and adults. Participants viewed a cow (the agent) travelling to a goal location via a longer, less efficient path (a shorter path was obstructed so could not be taken; Figure 8). After repeatedly watching this, the obstruction was removed so that both the long and short path could be taken. Which path would participants anticipate that the agent will take? Now, if statistical occurrence (like the frequency of prior observations) is more privileged than action prediction, we would expect anticipations to the longer path, as that had been observed multiple times before. However, if teleological reasoning dominates over frequency learning, then we would expect anticipations to the shorter path, as it is the most efficient way for the agent to reach their goal location.



Figure 8 | Frames of the introductory stimuli, whereby a cow takes a longer path because the shorter path cannot be taken. The arrows indicate direction of movement; the oval indicates the occluder.

From Paulus et al. (2011).

Paulus et al. (2011) found that initially (i.e., in the first test trial), when both the short and long path became available, adults and infants alike still anticipated the agent would continue to move to the longer path (despite the availability of a more efficient path). Only after an instance of actually seeing the agent take the shorter path (i.e., only in the following test trials), did adults then adjust to anticipating the agent would move to the shorter path. Unlike adults, infants kept anticipating towards the longer path across all the test trials.

Another study with pre-schoolers, adolescents, and adults, used a similar procedure to investigate statistical learning. Schuwerk and Paulus (2016) measured participants' anticipatory eye movements as they watched an agent: a turtle with a preference for lettuce. The agent stated that they intended to reach their goal of the lettuce as quickly as possible, then took a shorter (efficient) path, rather than a longer (inefficient) path, several times. When the occluder was introduced in the first test trial, the authors found that participants' predictions of the agent taking the shorter path were only at chance levels. Overall, only the older age groups predicted the shorter path after they were repeatedly shown the subsequent test trials. From this, Schuwerk and Paulus suggest that visual action anticipation is guided by frequency information, rather than efficiency (which would fit with teleology).

2.6.1. Infants doing statistics may not spell the end for teleology

It seems infants' and children's (and initially adults') predictions in both these studies were based on statistical regularity, rather than the teleological stance of considering the most rational outcome; they kept anticipating the longer path used in familiarisation and not the shorter, more efficient path. Therefore, the teleological stance does not seem descriptively adequate, since participants do not always track the goal of a purposive action according to this stance. However, just because statistical regularity information is sometimes prioritised over goal information, that does not mean infants are unable to track goals according to the teleological stance. This is explained next.

Rather than treating the exploitation of statistical regularities as being an alternative process to goal-tracking, exploitation of statistical regularities could in fact build on goal-tracking and may sometimes depend on it. When using some type of routine or regularity to help predict action, it is often in reference to a goal. For instance, in Woodward's (1998) study, there is a tacit assumption that infants are using statistical regularities—regarding what people have done in the past—to help them predict what people will do in the future (in relation to a goal, such as reaching or grasping). With that said, we require a theoretical framework that explains when and why infants respond differently (i.e., according to statistical regularities versus according to the teleological stance). For now, this remains a puzzle regarding the development of goal-tracking.

Another point regarding statistical learning for actions is this: not only does statistical learning come into play during a single experiment (i.e., participants view an agent performing the same actions in repeated trials); one's life experience of action observation can also be taken as accumulated data for input into statistical learning. That is, prior observational experience, broadly construed, contributes to learning the regularity and frequency with which agents tend to perform certain actions. Such experience can then aid in action prediction. Infants learn to anticipate the regularities and patterns they generally observe in others' movements (Woodward & Cannon, 2013). Take, for instance, a finding from Kochukhova and Gredebäck (2010), whereby when an

agent grasps a spoon, six-month-olds look in expectation that the agent will subsequently bring the spoon to their mouth. They have learnt that hands grasping food (or grasping a food-related tool) tend to be brought to one's mouth. In addition, infants' own motor repertoire may interact with frequency information accrued through daily life, to contribute to their understanding of goal-directed actions (Monroy, Gerson, & Hunnuis, 2017).

2.7. Timing is everything

A further adaptation of Woodward's study, by Daum et al. (2012), measured anticipatory looking as well as dishabituation. They discovered a dissociation between these two types of eye gaze measures. The dishabituation findings were replicated, in that nine-month-olds seemed to use goal information to predict action. However, when it came to anticipatory looking, their findings indicated infants used statistical regularity information to predict action. The authors suggest that due to the speed of anticipatory looks versus the requirements for goal computation, perhaps the infant is simply unable to compute the goal in such a short space of time (a few hundred milliseconds). Some researchers argue against this idea (that infants cannot compute a goal in the time available for anticipatory looking), claiming instead that infants are capable of goalbased action prediction. If this claim is correct, it remains unclear why infants in Paulus et al.'s (2011) study (agent comes to a fork in the road with the option of taking a shorter or a longer path) used statistical information to anticipate behaviour.

2.8. Must walk the walk to track goals

It turns out that during the first nine months, infants' goal-tracking ability for others' actions is correlated with their own motor repertoires (i.e., their own ability to perform those actions; Woodward & Cannon, 2013). One way this is evidenced is by enhancing an infant's ability to act and observe the consequences. Sommerville, Woodward, and Needham (2005) conducted one such study, whereby they gave threemonth-olds *sticky mittens* (based on Needham, Barrett, & Peterman, 2002) that would stick to objects covered in Velcro® strips. Cleverly, the purpose of these mittens was to allow infants to manipulate and grasp objects with their hands, as they typically cannot yet do so at this age. The question is: does extending infants' capability for action also improve their goal-tracking during action observation?

The researchers used Woodward's (1998) paradigm (see previous Figures 6 and 7), and in addition gave infants toys to play with while wearing the mittens. One group played with the toys before the Woodward paradigm (*play first*) and the other group played with the toys afterwards (*play after*). The reasoning was that since playing with toys while wearing the mittens improves action abilities, the infants who played before doing the task should improve in their goal-tracking. This is indeed what they found. Three-month-olds, who were yet to play with toys while wearing the sticky mittens, did not show goal-tracking. In contrast, those infants that played with toys first did in fact show goal-tracking. In this way, it is apparent that infants' individual ability to perform particular actions facilitates their ability to track the goals of that action in others.

Simply put, after these young infants acquired experience in grasping, they began to track others' goal-related grasping actions. Such findings indicate that the very representations underlying one's own actions also play a part in our understanding of others' actions. This idea can be referred to as embodied cognition, in which cognitions (like thoughts and knowledge) co-opt mental structures that originally evolved for action perception (Woodward & Cannon, 2013). With evidence suggesting correspondence between infants' action performance and their ability to track goals of actions, this leads to yet another puzzle: in what way is infants' goal-tracking connected to their action ability, and why is there a relation?

A criticism of Sommerville et al.'s (2005) study is that the *play first* group spent longer observing actions before they completed the goal-tracking task, compared to the *play after* group who did not observe any actions prior to the task. Plausibly, the mere observation of goal-directed actions by the *play first* infants could be the reason for their performance, rather than their actual active experience in manipulating and grasping the toys. Sommerville, Hildebrand, & Crane (2008) addressed this issue. They had a slightly different task, in which infants watched an agent use a cane to pull an outof-reach toy towards themselves. Importantly, they controlled for the observation issue by having both groups observe action before taking part in the habituation and test events. One group of infants actively trained and practiced using the novel tool; the other group of infants received observational experience that matched the training and practice. Infants only tracked the goals of action involving a novel tool if they themselves had learnt to use the tool.

2.9. Conclusion

Infants may represent the actions of others as having goals, rather than just being movements in space (Woodward & Cannon, 2013) from at least three months old, which is also when they develop the ability to track physical object behaviour. Since objects are quite different from agents with goals acting on objects, how do infants manage this? One possibility is that from early in life, humans may engage in pure goal-tracking, that is, tracking of outcomes of actions that do not involve mental state ascription. Pure goal-tracking could be the foundation for later mental-state-tracking in infants and adults. The teleological stance shows how pure goal-tracking is theoretically plausible, in that relating action to goals can enable someone to track the goals of that action, and infants' anticipatory responses seem attuned to goal-directed actions.

As such, there are two puzzles to consider when thinking about the descriptive adequacy of the teleological stance. Firstly, sometimes infants' responses are in line with a teleological explanation, but other times they seem to be based on statistical regularity. That is, the same responses could come about through patterns infants learn (e.g., a hand grasping a mug tends to be brought towards the agent's mouth). This discrepancy may not be problematic in and of itself, since sometimes sensitivity to statistical information is useful for goal-tracking. However, a good theory needs to appreciate when and why these differential responses occur. Are the anticipatory responses infants generate guided by teleological goal analysis or statistical regularities? How can we predict when infants will use one process versus another? With this discrepancy in mind, the findings covered in this chapter regarding infant goal analysis are best taken with a grain of salt. That is to say, the findings do not whole-heartedly prove that infants' anticipatory responses are necessarily guided by analysis of others' goals (Woodward & Cannon, 2013).

Secondly, infants' goal-tracking abilities seem to be linked to their physical abilities to perform actions, but it is as yet unclear how or why this link exists. A strong theory regarding goal-tracking within the first nine months needs to solve both these puzzles; only then can we acquire an understanding of goal-tracking within the context of development.

Chapter 3.

3.1. Introduction

The previous chapter gave us insight into action understanding and goaltracking, and highlighted how the teleological stance can support the idea of pure goaltracking (i.e., that which occurs in the absence of mental state input). However, while pure goal-tracking is possible, it also runs into problems because several studies suggest different underpinning mechanisms, such as statistical learning, instead of teleology. In addition, further studies suggest that some goal-tracking is determined by one's own ability to act. This all raises the pressing question of how else to explain the ways in which goal-tracking is performed.

In the current chapter, a motor theory for goal-tracking is thus offered, as a viable theoretical candidate to counter the above issues. After this, studies evidencing such a motor theory are reviewed. Then, a counterargument to the theory is introduced, but subsequently put to rest by explaining the important distinction between targets and goals. Finally, a revised dual process theory for goal-tracking is revealed, along with theoretical conjecture that answers the two developmental puzzles mentioned in Chapter 2 (i.e., discrepancy whereby infants' responses are sometimes according to teleology, and other times according to statistical regularity; mystery as to how or why infants' goal-tracking abilities seem linked to their own physical ability to perform actions).

3.2. Motor theory for goal-tracking

We know action understanding is essential for social animals like humans; therefore, it makes sense that tracking the goals of others' actions is an ability that emerges early in life, and that this ability continues to operate into adulthood. When adults reach for an object, their eyes are usually looking at the object they will grasp (the target), not at their hand doing the reaching. Adults also do the same when watching someone else engage in an action (i.e., their eyes quickly fixate on the target object of the grasp, rather than simply following the grasping hand itself). Such eye movements are anticipatory—or proactive—and are an example of rapid goal-tracking. The eyes are anticipating an action in advance of it occurring (i.e., well in advance of the agent's hand contacting the target object). The connection between goal-tracking and proactive eye movements is evidenced by the fact that changes in the action observed will change eye movements. To illustrate, take the case that your hand pre-shapes ahead of time depending on the size of the object to be grasped; for instance, you would use your whole hand to grab an apple, but a precise pinching motion of the fingers to pick up a berry.

An expansion of this is when there are two differently sized objects that a hand is approaching (Figure 9). We pre-shape our hands depending on the size of the object we intend to grasp, so adults can ascertain which of the two objects is the target object by the way in which the hand is shaped. Ambrosini, Costantini, and Sinigaglia (2011) conducted a famous study confirming that observing the pre-shaping of an agent's hand triggered which of two objects adults would proactively gaze towards. In other words, Ambrosini et al. found that adults proactively gazed at the large object when they observed the agent making a whole-hand prehension, and adults proactively gazed at the small object when they observed the agent making a pincer grasp. Adults' eye gaze latched onto the appropriately sized target object well before the agent herself actually grasped the particular object.



Figure 9 | Resting posture depicting a hand (with fixation cross) before the hand grasps towards either the large or the small object.

From Ambrosini et al. (2011).

The reason this goal-tracking can happen with such speed is partly due to motor representations (representations that help us prepare, perform, and monitor action sequences). Motor representations can be described as living a double life, because they are involved not only in the performance of actions, but also when observing another person perform actions (Butterfill, 2020). Why might this be? It is possible that action representations are embodied (Woodward & Cannon, 2013). That is, the very representations that guide action production could also aid in action understanding.

In order to explain how motor representations facilitate goal-tracking, Butterfill (2020) proposes a motor theory for goal-tracking, which can be summarised as follows. Firstly, motor processes compute sensory and behavioural outcomes by using representations of outcomes. Secondly, goal-tracking using motor processes can start off with multiple candidate outcomes, because many means-ends computations can occur quickly. It is not the case that merely one action outcome is identified; action preparation processes can select one of many possible outcomes. Thirdly, not all goal-tracking necessarily involves motor processes under this theory. The theory acknowledges that some instances of goal-tracking can occur through deliberation for instance, but claims that pure goal-tracking can only be enabled by motor representations and motor experiences.

3.2.1. Empirical evidence for a motor theory of goal-tracking

This section briefly reviews empirical evidence that supports a motor theory for goal-tracking. If motor representations are important for goal-tracking, then interfering with such representations would disrupt one's ability to identify outcomes of action. Ambrosini, Sinigaglia, and Costantini (2012) conducted a study whereby they tied participants' hands, which the authors conjectured would interfere with adults' action representation abilities. In a study by Costantini, Ambrosini, Cardellicchio, and Sinigaglia (2014), they instead used TMS (transcranial magnetic stimulation) to interfere with participants' action representation abilities. The TMS was targeted towards brain areas linked to motor representation. In both research studies, participants were observing an agent reach for an object, and their proactive eye movements were found to be greatly reduced. So, although motor representations were originally posited in the context of action performance, it seems they also contribute to goal-tracking. Further, from a separate laboratory to the above studies, confirmation that motor processes do indeed play a functional role in action anticipation also comes from Krogh-Jespersen and Woodward (2018). Like the sticky Velcro® mittens study described in section 2.8. (Needham et al., 2002; Sommerville et al., 2005), Krogh-Jespersen and Woodward found that 8-month-olds' active experience facilitated their goal analysis of subsequently observed actions.

As mentioned earlier, proactive gaze is when one looks ahead to the target of an action. For instance, when a hand is approaching an object to grasp it, the proactive gaze lands on the target object before the hand reaches it. Falck-Ytter et al. (2006) found that infants also do this. Crucially, however, infants only engage in proactive gazing when they can perform such reaching actions themselves. If infants are not yet capable of reaching, their gaze does not land on an object before the hand grasps it (Kanakogi & Itakura, 2011). Thus, goal-tracking ability in infants appears related to the infant's ability to act, although why this is the case had so far remained unclear. The motor theory for goal-tracking can potentially explain the reason for this constraint. This is because the ability to represent actions motorically is connected to the ability to perform

those same actions. The theory then implies that infants' goal-tracking ability should, therefore, be constrained by their ability to motorically represent actions.

Infants in Green, Li, Lockman, and Gredebäck's (2016) cross-cultural study predicted action goals regarding eating, based on their cultural background. When shown an agent eating with either chopsticks or a spoon, Chinese infants only recognised the goal of eating when the action used chopsticks, while Swedish infants only recognised the goal of eating when the action used a spoon. Further, Ambrosini et al. (2013) found that infants' proactive gaze also differentiated between the two grasp types of whole-hand prehension and pincer grasp (dependent on the infants' own grasping ability). What one's cultural background and personal grasping ability have in common is that they both equate to having more experience with certain actions. In the case of culture, Chinese infants are more used to seeing chopsticks as tools to aid eating, while Swedish infants are more used to seeing spoons serve this purpose. In the case of physical ability to perform particular actions, younger infants who cannot yet grasp precisely therefore have no experience doing so, while older infants who have mastered precise grasping inherently have more experience making such actions. Thus, in both these studies (Ambrosini et al., 2013; Green et al., 2016), infants' prior physical or motor experiences modulated their gaze behaviour.

3.3. What about shapes?

As shown by the previous section, there is support for the idea that infants' goaltracking ability is constrained by their ability to motorically represent actions. However, this explanation encounters a problem; namely, how is it that infants track goals when actions are performed by geometric shapes and cartoon animations (e.g., Gergely et al. 1995; Paulus et al., 2011)? The motor theory for goal-tracking implies that infants' ability to track goals should depend on their ability to motorically represent events. But, presumably moving shapes and cartoon animals are not likely to be motorically represented as they do not have physical effectors similar to the observer (Csibra, 2008). Although it seems that such an argument proves the theory incorrect, it may not need to be rejected. To understand why, we must tease apart the difference between goals and targets.

3.3.1. Mere targets and proper goals

Arguably, the traditional view from the literature to date has conflated targets and goals, when they are in fact critically distinct from each other. The target of an action is that towards which the action is directed. For instance, if the goal of an action is to pick up an apple, then the apple is the target of this action (because the action of picking up is directed towards the apple). Specifying a target of an action means partially specifying one of the goals of that action. However, to fully specify a goal, one requires more than just knowledge of the target. Goals usually involve a few components regarding the action, for instance, *type* of action and *manner* of action: type of action refers to what the action is (such as grasping or kicking); manner of action refers to how the action is done (such as gently or firmly). Distinguishing between targets and goals is important. Next, let us review an example in which this distinction was not made.

Adults possess an understanding of the concept of perceptual animacy, which allows them to track targets exclusively, that is, to track targets and not goals. Perceptual animacy detects animate objects and their targets using perceptual processes. Based on the way animated shapes move, observers can perceive a rich scenario. For instance, Heider and Simmel's (1944) classic display of moving shapes led participants to perceive that the two small shapes were fleeing from the large shape, which was trying to catch them. Although the instructions were general, most participants richly perceived the shapes as animated beings. Similarly, Gao, Newman, & Scholl (2009) reported on perceptual animacy with a study that involved participants visually searching a scene in which—likened to a wolf chasing a sheep—one shape was 'chasing' another. The study only involved target-tracking; as such, the instructions merely required participants to track a target. While we can at least be sure that adults realise the geometric shapes are not alive, often we cannot help but see them as such when they move in particular patterns (Gao et al., 2009).

Perceptual animacy is often claimed to involve tracking goals, when it in fact only involves target-tracking (not goal-tracking). The type of action, the manner of action, and outcome features are not specified, and also irrelevant to the task. Perceptual animacy is underpinned by broadly perceptual abilities which are separate to those used for proper goal-tracking (Butterfill, 2020). To summarise, perceptual animacy involves tracking targets only, whereas proper goal-tracking tracks targets as well as goals.

Having differentiated mere target-tracking from proper goal-tracking, the next question is whether infants are able to track goals proper during the first nine months. Importantly, many studies do not make this distinction between targets and goals. If infants can properly track goals, not just targets, they should be able to distinguish between the target of an action and the type of action, since these are components of specifying a goal. Can infants identify action type, such as whether an action is a push or a reach, and can infants differentiate between actions that have the same target but are different in type?

Behne, Carpenter, & Tomasello (2005) showed that infants inferred an agent's gestures as having communicative signal relevant to the context. In a hiding game, the agent's gesture revealed the location of a hidden toy, and infants recognised that rather than simply pointing or looking to that location for no reason, the agent intended this gesture specifically to reveal a secret hiding place. In Kochukhova and Gredebäck's (2010) study, infants anticipated that an agent feeding themselves with a spoon would bring that spoon to their mouth, but did not anticipate that a self-propelled spoon would move towards the agent's mouth. In this way, infants identified actions that shared a target but were nevertheless different.

Evidently, these studies suggest that infants can indeed differentiate the target and the type of goal-directed actions. On the one hand, infants properly tracking goals suggests they can engage in more than just perceptual animacy. On the other hand, it is possible that previous studies purporting to evidence goal-tracking in infants may actually only involve perceptual animacy, that is, target-tracking not goal-tracking. Butterfill (2020) uses this as a starting point to develop a theory which solves the two puzzles about infant goal-tracking. This is detailed in the following section.

3.3.2. Revising the motor theory for goal-tracking

Let us return to the motor theory for goal-tracking, which purports to explain all goal-tracking during the first nine months. This theory can potentially explain the link between infants' ability to track goals and their ability to perform actions. However, there was an issue in that infants appear to show goal-tracking behaviour when the 'agents' are cartoon animals and self-propelled geometric shapes (Csibra & Gergely, 1998; Daum et al., 2012). Such situations may involve mere target-tracking as a result of perceptual animacy, not proper goal-tracking. Following on from this, Butterfill (2020) considers a revision to the motor theory for goal-tracking: in addition to all proper goaltracking during the first nine months being explained by the motor theory, these infants also possess mere target-tracking processes which are the same as adults' perceptual animacy processes. In cases where it seems infants' pure goal-tracking is not constrained by what they can motorically represent, this is due to the erroneous conflation between mere target-tracking and proper goal-tracking. Such cases actually involve target-tracking only. This theoretical revision leads to a dual process theory of goal-tracking, whereby different types of tracking involve different processes; namely, proper goal-tracking calls on motor processes, while mere target-tracking calls on perceptual processes.

Returning once again to the motor theory for goal-tracking, the issue was that infants at nine months appear to track goals of cartoons and shapes, but according to this theory they should not be able to do so. The revision—resulting in a dual-process theory for goal-tracking—deals with this issue through the distinct perceptual and motoric processes required for different kinds of goal-tracking. This dual-process theory generates testable predictions, but because of its recency, it has not yet been tested empirically. There may be separate signature limits—as well as separate processes between goal-tracking and perceptual animacy, which lead to empirical predictions. On the one hand, goal-tracking can be impaired by manipulating a person's ability to motorically represent actions (this notion will be elaborated upon soon in Chapter 4). On the other hand, perceptual animacy can be impaired through manipulation of factors like subtlety (how directly the target is approached by the chaser) and directionality

(whether the chaser seems to be facing the target; Gao et al., 2009). While the dualprocess theory for goal-tracking makes bold claims, it does make sense theoretically.

3.4. Dual-process theory for goal-tracking

The dual-process theory for goal-tracking can potentially explain both the puzzles (from Chapter 2) about infant goal-tracking within the first nine months. Let us remind ourselves of these two puzzles and see how the dual-process theory could account for them. The first puzzle regarded the discrepancy between infants' responses sometimes being based on statistical regularity, and other times being in line with the teleological stance. The dual-process theory gives us a possible answer, based on the following three assumptions. Firstly, in scenarios involving human agents, infants' responses tend to be governed by motor processes, not perceptual animacy processes. Secondly, perceptual animacy processes in infants cannot influence anticipatory looking; such processes can only influence pupil dilation and looking time. Thirdly, in cases where goal-tracking of any type is not possible, infants' responses tend to be governed by statistical regularity. These assumptions are plausible. Taking these three assumptions into account, the dual-process theory indicates that when infants observe actions they cannot perform, their responses should rely on statistical regularity and help them anticipate the way in which the action will unfold. An empirical example of this is Daum et al.'s (2012) paper that found a dissociation between infants' anticipatory looking and dishabituation (see section 2.7.).

The second puzzle regarded how and why infants' goal-tracking abilities are connected to their own abilities to perform actions. The dual-process theory for goaltracking suggests that the type of connection depends on the type of tracking, that is, mere target-tracking or proper goal-tracking. Infants' goal-tracking within the first nine months should be limited by their ability to motorically represent observed actions, since proper goal-tracking during this time relies on motor processes. This is similar to the way in which adults' motor-based goal-tracking is constrained. However, mere target-tracking should not have the same constraints, since it relies on perceptual animacy, not motor processes. Going back to some aforementioned studies (Csibra &

Gergely, 1998; Daum et al., 2012), if we assume that they involve target-tracking only, then it follows that in those cases infants' tracking abilities are not limited by what they can motorically represent.

3.5. Conclusion

The teleological stance shows how pure goal-tracking is possible, at least in theory. Humans can engage in fast, pure goal-tracking from a young age, which can be shown through anticipatory looking. This chapter considered the possibility that some infant goal-tracking is underpinned by motor processes, since there is evidence that adult goal-tracking can sometimes result from motor representations and processes alone.

The motor theory for goal-tracking explains all pure goal-tracking during the first nine months. However, a potential complication comes from research suggesting that infants engage in goal-tracking for objects such as shapes and cartoons. In order to better understand goal-tracking, we must differentiate between targets and goals, which are often conflated with one another. Doing so also allows us to distinguish between *types* of tracking, that is, mere target-tracking and proper goal-tracking. Furthermore, we can distinguish between the processes behind each type of tracking, namely, perceptual animacy and motor processes respectively.

The revised dual-process theory for goal-tracking builds on the motor theory for goal-tracking. Regarding how knowledge about action is developed, the dual-process theory posits that goal-tracking does not necessarily lead to knowledge of action goals. Further, it posits that goal-tracking during the first nine months involves perceptual representations and motor representations, which are separate from knowledge states. Knowledge of action goals develops later, possibly due to language communication and social interaction becoming richer. Accepting the dual-process theory means that goal-tracking can be seen as a starting point for mindreading abilities such as mental-state-tracking. The pure goal-tracking that underpins these abilities is important, as it links mere body movements to representations of action outcomes, which helps us make sense of our social world. In sum, a dual-process motor theory for goal-tracking seems

to be a viable candidate explanation for current findings, and the connection between movement kinematics and goals may very well be foundational for full blown, mature mindreading. The next chapter will detail a central paradigm used for action observation in the motor context, to forecast the experiments undertaken for this thesis.

Chapter 4.

4.1. Introduction

The previous chapter gave us an overview of a motor theory of goal-tracking. The current chapter will review the motor system in more depth by introducing mirror neurons. Then, a novel and exciting proposal is discussed, whereby mindreading is theorised to have a motoric basis. After this, an action observation paradigm central to the current thesis is introduced, followed by a grasp cueing procedure.

4.2. The motor system and neurons that mirror

Gesturally, we tend to mirror conversational partners without conscious effort (take note of this next time you have a one-on-one conversation). On a finer scale, it turns out there are neurons in the brain that activate when we perform an action, as well as when we observe someone else performing that same action (Goswami, 2020); these are called mirror neurons. Mirror neurons are specialised cells in the cerebral cortex that have the potential to shed light on mindreading and social cognition, since they behave in the same way regardless of whether an action is executed by oneself or another person.

Mirror neurons were discovered by Rizzolatti et al. (1988); they found that certain neurons in area F5 in the monkey premotor cortex that discharged during execution of specific goal-directed actions (e.g., grasping towards a peanut) also discharged in response to visual observation of the same action performed by another monkey or by an experimenter. Importantly, the mirror neurons did not discharge if the observed movement was performed in the absence of the goal-directed object, or if the object was presented alone. Rizzolatti et al. thus emphasised that mirror neurons encoded goal-directed actions.

Other research suggested that a similar neural mechanism also existed in humans. For instance, there was overlapping activation in the premotor cortex when adults observed actions performed on objects and when the adult observer themselves

actively performed the same actions (e.g., Buccino et al., 2001). It has also been shown that the activations in the human premotor cortex extended to the inferior frontal gyrus which, in the human brain, includes Brodmann areas 44, 45 and 47 (these are considered analogous to the monkey premotor cortex F5 where mirror neurons were first found; e.g., Rizzolatti, Fadiga, Gallese, & Fogassi, 1996).

Mirror neurons and the mirror system (both in humans and animals) are starting to be considered as a potential mechanism for the cognitively efficient low-level tracking of others' actions and their mental states (Apperly, 2011). That said, the neuronal pathways relevant for action understanding and action prediction are still debated and under investigation (Csibra, 2008). The next section looks at a hypothesis that integrates mirror neurons and action understanding.

4.3. Direct matching hypothesis

The direct matching hypothesis (also known as the motor simulation account) is an account—based on the discovery of these mirror neurons—which posits that mirror neurons are the neural substrate of mindreading (Lavelle, 2018). The basic function of action mirroring is thought to be action understanding in relation to goals. According to the direct matching hypothesis, the way this occurs is through simulation of the perceived action in the motor system of the observer; this motor resonance allows the observer to understand the action outcome, and therefore the goal (Csibra, 2008). The main tenet of this hypothesis is that mirror neurons are functional to mindreading because they directly help us understand the actions of others from the inside. In doing so, mirror neuron activity (and by extension, our motor system) does *not* involve the attribution of any propositional attitudes to others. Hence, this minimal, motor-based mindreading view contrasts with the traditional view (reviewed in section 1.3.), that mindreading *does* involve attributing propositional attitudes to other people.

The direct matching hypothesis maintains that because of mirror neurons, movements are given meaning, which then leads to understanding. Actions by agents are seen as a means to an end goal (grasping an object), rather than isolated physical movements (fingers moving horizontally in space). On this view, the neuronal network takes visual information during action observation, and turns it into knowledge about action outcomes. While the visual system receives information about physical properties, this hypothesis maintains it is the mirror neuron system that adds information about the goals of behaviour.

However, if it is the case that observing an action or anticipating someone's potential action causes one's own motor system to activate, then motor representations should influence our own behaviour or responses in socially relevant situations. There is little research documenting whether motor representations are accessed or functionally relevant for tracking social action (e.g., for processing others' actions that are based on false belief). A few recent studies, however, have made strides on those fronts (e.g., Edwards & Low, 2017; Low et al., 2020; Zani, et al., 2020), going so far as to suggest that certain forms of mindreading may even be motorically grounded; let us turn to look at those findings more closely.

4.4. Evidence for motorically grounded mindreading

Research shows that when we imagine, observe, or predict others' actions, there is motor-cortex activation that translates to a preparatory state in which cognitive processing is optimised (e.g., Southgate & Begus, 2013). For instance, evidence for the link between the motor experience of observers and activation of the mirror neuron system during action observation comes from Haslinger et al. (2005). In their study, professional pianists showed stronger motor activation, compared to musically naïve controls, when observing piano playing actions. Pianists and control participants displayed similar levels of motor activation when viewing control stimuli that consisted of non-musical finger movements.

4.4.1. **Response facilitation**

Taking on board such ideas, Edwards and Low (2017) tested the extent to which adults exhibited behavioural indicators of motor-cortex activation (which would be marked by optimised or speeded response times) when anticipating an actor's reach-tograsp action; the actor's reaching actions were determined by combinations of her goals,

desires, and beliefs. To maximise activation of an equivalent internal motor representation in the observer, participants' own button-reaching responses had to correspond to the right-handed reaching movements of the agent.

Edwards and Low (2017) found motor-cortex response facilitation when adults anticipated an executable reaching action (the agent falsely believed there was a desirable object inside the box). Participants did not show motor-cortex response facilitation when a non-executable action was anticipated (the agent falsely believed there was an undesirable object inside the box). It seems that fast-tracking of someone else's mental states may interact with our motor processes, thereby enabling us to consider some of that person's mistaken beliefs when anticipating their actions motorically.

4.4.2. Directional leaning

If motor processes are successfully used in action observation, will observers motorically represent action outcomes that are the goals of that action? Zani et al. (2020) argue that motor representations go beyond goal ascription and can even accommodate belief-tracking. Their recent study used an interactive task design to investigate the connection between motor and mind processes. The authors found that participants' motor activity regarding directional leaning (leftwards or rightwards on a balance board), as well as their anticipatory looking, forecasted how an agent prepared to act based on their false belief. Such results provide evidence that rapid tracking of belief can affect the motor representations one generates when interacting with an agent.

Zani et al. (2020) examined motor representations of action anticipation in the context of a false belief task. Adults stood on a balance board and observed an agent moments before she was about to execute a search action (the agent had a false belief that her ring was inside one box when it was really inside another box; there were no overt cues to suggest which box the agent would move towards). Analyses of adults' mediolateral (left-right) balance shifts showed spontaneous motor representation of belief-based actions by the observers, in response to the agent's situation.

In the false belief condition, adults leaned towards the empty box, suggesting that they were motorically embodying the agent's erroneous action. Similar findings were reported in the true belief control condition. The results document—for the first time adults' motorically grounded expectations of the agent's action being modulated by the workings of a false-belief-tracking system. Zani et al. discussed that because some motor processes can be planning-like (e.g., Zhang & Rosenbaum, 2008) and can be activated when we imagine actions (e.g., Jeannerod, & Frak, 1999), the mediolateral leaning data suggested that motor processes are not tied to the actual state of the world, but rather can generate expectations based on others' belief-like states of the environment.

However, we should keep in mind that in Zani et al.'s (2020) study, the participant observed as they stood facing the agent. Therefore, the results showed that observers leaned, for instance, to their own right side when proactively anticipating that the agent from her own perspective would search in the empty left box. In other words, Zani et al.'s participants leaned right in anticipating that the agent would be leaning left to go to the empty box. At first glance, the results would appear to contradict classical ideas of point-for-point effector correspondence that are often assumed in direct matching or mirroring accounts. Yet, there is some coherence in Zani et al.'s findings, because there are also studies suggesting that motor representations do not have to rely on there being strict direct matching between self and other; there can be selective discharge in motor activity according to the goal an action is directed towards, regardless of the specific effector used (e.g., see Cattaneo, Sandrini, & Schwarzbach 2010). Nonetheless, more work needs to be done to unpack whether, and to what extent, our motor system is even able to remap another person's allocentric frame of reference into our own egocentric frame of reference.

Finally, Zani et al. (2020) discussed their findings as highlighting that motor and belief-tracking processes are integrated (and that the efficient mindreading system might end up being a motor mindreading system). However, it may be that the belieftracking process generates predictions, and these predictions then guide observers' mediolateral activity (see also Csibra, 2008). If motor processing and belief-tracking are
integrated, then temporarily impairing motor processing should impact belief-tracking; Low et al. (2020) carried out such an experiment.

4.4.3. Object detection and Egyptian mummy: what's the link?

Before introducing Low et al.'s (2020) study, it is necessary to briefly touch on a ball-detection task (from Kovács, Téglás, & Endress, 2010) that was used to measure automatic belief-tracking. In this task, participants are simply asked to press a button on a computer keyboard when a screen falls away to reveal a ball. In some trials (P-A+ condition), participants and the agent observe a ball roll behind a screen on a table; then the agent leaves, and, in the agent's absence, participants watch the ball roll from behind the screen to leave the scene altogether. In other trials (P-A- condition), both the participants and the agent see the ball roll from behind the screen to leave the scene, whereupon the agent then leaves. In both trials, the agent returns, and the screen falls away to reveal a ball.

Kovács et al., (2010) show that participants respond faster to indicate the presence of the ball in the P-A+ condition than compared to the P-A- condition. It seems like our automatic tracking of the agent's belief about the ball's whereabouts (even though incorrect) boosts our object detection performance. Low et al.'s (2020) version of the task was such that when the agent returned at the end (just before the screen fell away to reveal a ball), the agent's appearance changed – he was bound up tightly in linen sheets (somewhat like an Egyptian mummy; Figure 10), and he was consequently unable to potentially interact with the ball. Low et al. found that the automatic belief-tracking effect (P-A+ < P-A-) was obliterated when the agent was physically constrained, compared to when the agent was free (i.e., dressed in a way that permitted interaction with the object). It is possible that the stalling of the belief-tracking effect was simply perceptually driven (due to the change in clothing); Low et al. found that when the agent returned in changed appearance but remained able to potentially grasp the ball (the agent wore the linen sheet loosely), the belief-tracking effect returned.



Figure 10 | 'Mummified' agent bound in linen sheets. From Low et al. (2020).

In the ball-detection task, participants never see the agent interacting with the ball when the screen falls away; Low et al. (2020) emphasise that "even observing an agent who has the potential to act can trigger observers' readiness to map the motor potentialities of someone else's belief-based situation onto their own representations of the environment" (p. 3). Essentially, Low et al. suggest that certain kinds of mindreading may be a bodily social sense, that is, we automatically track others' situated bodies and not just others' minds. Low et al.'s findings are intriguing and (similar to Zani et al.'s, 2020, findings) suggest a close connection between motor processes and belief-tracking processes.

However, it is possible that the belief-tracking effect in the ball-detection task remained present, but was merely unexpressed in the mummified agent version. We do not know whether 'mummifying' the agent early on in the event sequence (and not just when the agent returns at the end of the sequence) would yield similar interference effects. Moreover, we do not know whether belief-tracking effects would be detectable via other modalities (e.g., via pupil dilation). We also do not know whether temporarily interfering with participants own motor affordances (e.g., binding participants hands, or applying direct intervention on the motor cortex using TMS) would also eliminate the P-A+ < P-A- effect.

Most problematically, in positing that motor representations can weigh in on belief-relevant situations, Low et al. (2020) take for granted that motor representations reliably accommodate goal ascription during action observation. The conjecture that motor representations "underpin the primary ways in which human beings engage in social cognition to play a role in the tracking of beliefs which specify someone's reasons for action" (p. 8) may be built on a house of cards if it turns out that researchers cannot take for granted evidence of motor representations enabling tracking of others' goals during action observation. The view that certain goal-tracking processes involve motor processes is often motivated by the touchstone work of Ambrosini and colleagues on motor processes occurring in action observation. This chapter will now turn to explain the action observation paradigm in relation to Ambrosini and colleagues' findings.

4.5. Action observation paradigm

A fundamental question in research on human mindreading is to understand how we comprehend that people's outward actions are motivated by goals. When watching a simple action, such as someone reaching towards a teacup, there tend to be three levels of description: muscle level, kinematic level, and goal level (Hamilton & Grafton, 2008). The *muscle* level involves the specific muscles involved in the action; the *kinematic* level involves general hand-shaping and arm movement; the *goal* level involves the action outcome. These levels of description are independent, as illustrated by the following examples. Firstly, the same goal of drinking tea can be achieved in different kinematic and muscular ways; one could (1) precisely grip the cup handle between thumb and forefinger, or (2) use both hands to bring the cup upwards as if cupping a bowl. Secondly, the same sequence of specific muscle activation or kinematic movement can lead to different goals depending on the context; one could use the same initial pincer grasp of the cup handle to (1) drink the tea, or (2) tip the cup's contents down the sink because they over-steeped the teabag. Lastly, the same general kinematic motion of reaching towards a cup handle might not use identical muscle activation in separate instances; this can be true within the same individual over time, as well as between individuals.

Research shows that when young children and adults observe another individual make, for instance, goal-directed movements like grasping one of two objects or

reaching for an object hidden in one of two locations, participants display proactive gazing well in advance of the action outcome (e.g., Ambrosini et al., 2011; Cannon & Woodward, 2012; Kanakogi & Itakura, 2011). What kinds of representations underlie such proactive or anticipatory eye movements? Ambrosini et al. (2011), coming from their broader research background in mirror neuron activity, suggest that deriving others' goals relies on direct matching of the observed action to an action in our own motor repertoire. Different grip apertures, such as pincer grasps and whole-hand prehensions, can provide motor representations of potential actions, so let us now take a closer look at somewhat of a lynchpin of the field, that is, the action observation paradigm used to test this idea.

4.5.1.The original

If our motor representations can drive our eye movements proactively towards the target of a goal-directed action, even when the agent's mental state is not known in advance, it will be important to specify what kinds of information the motor system might capitalise upon. In Ambrosini et al.'s (2011) seminal study, adult participants (N =15) watched videos in which an agent reached out to grasp either a small tomato (with a pincer grasp) or a large tomato (with a whole-hand prehension; see Figure 11 for an exemplar of each experimental condition's movement kinematics). The control videos showed the agent simply touching one of the two objects without a pre-shaped grasp (i.e., with a closed fist). The two objects were deliberately placed close together so that participants would be unable to extrapolate the movement trajectory and target object during the early portion of the action. The dependent measure was the proactivity of participants' eye movements, that is, how fast participants' eyes latched onto the appropriate object before the agent contacted the target. Gaze proactivity was recorded throughout the trials.



Figure 11 | Snapshots of stimulus videos involving two tomato objects. The rows depict the following variations (from top to bottom: no shape gesture to large target, no shape gesture to small target, whole-hand prehension to large target, pincer grasp to small target).

From Ambrosini et al. (2011).

Eye movements showed more proactivity (i.e., landed on the relevant object earlier) when participants watched either the small or large pre-shaped hand grasp, compared to when watching an agent simply touch an object with a closed fist. It is worthwhile to note that these findings regarding action observation complement other work on gaze proactivity during action performance: studies show that when we act, our eyes move just ahead of our action (e.g., Flanagan & Johansson, 2003). Ambrosini et al. (2011) interpret their findings as showing that adults benefitted from tracking motor cues (such as the pre-shaping of someone's hand) to anticipate other individuals' actions towards objects, thereby speeding up target object identification.

4.5.2. My hands are tied

In a follow-up study, Ambrosini et al. (2012) sought to replicate their original findings and provide converging evidence of direct matching as a process for deriving

task goals. If adults represent motor cues to proactively anticipate others' goal-directed actions, then interfering with adults' abilities to track motor cues should also disrupt gaze proactivity. The 2012 study used a within-subjects design (N = 15) with two conditions: hands-free and hands-tied. In the hands-free condition, participants watched the various stimulus videos with their hands resting freely on a table. In the hands-tied condition, participants watched the videos with their hands constrained behind their backs (subjects kept their hands crossed at the wrists, and their fingers closed into a fist shape). A hypothesis based on direct matching would predict that when participants' hands are constrained behind their backs, their saccadic eye movements would be slower to latch onto the correct target than when participants' hands are resting freely. The 2012 study replicated the 2011 study and, further, displayed that gaze proactivity of the hands-tied participants (as compared to the hands-free participants) was dramatically impaired in the pre-shape trials. The authors suggest that efficient action observation depends not only on motor representations, but also on our physical ability to exercise them.

4.5.3. Infants' own grasping matters

While these studies looked at adults, infants are slowly developing their grasping ability; at first, they cannot grasp at all, then they use a whole-hand prehension, later they begin to use fingers in a more precise way. Ambrosini et al. (2013) conducted a study in which infants with different grasping abilities observed a hand make whole-hand prehension and pincer grasp movements towards a target object. The stimulus videos were essentially the same as that used in Ambrosini et al. (2011; Figure 12), except that the 2013 videos were visually clearer to permit testing different age groups. For younger infants, variation in their own motor ability (via an independent task measuring participants' precision of grasping) predicted proactivity of performance during observation of the pincer grasp trials. In other words, infants' grasping ability matched their proactive gaze; (1) infants capable of a whole-hand prehension would gaze proactively at a target object when the hand was using this large grip, and (2) infants capable of a pincer grasp would gaze proactively when the hand was using this

small grip. Adults in this study showed similar proactivity when anticipating the target of the pincer and whole-hand grasps.



Figure 12 | Snapshots of stimulus videos involving two green ball objects. This figure shows the target layout whereby the large object is near the hand. The row depicts the following variation: whole-hand prehension to large target. The white circle in the left-hand column depicts the area of interest superimposed on the target object.

From Ambrosini et al. (2013).

Ambrosini et al. (2013) maintain that the overall findings confirm that adults as well as infants can distinguish both the type and target of a goal-directed action, and that human beings' tracking of the goals of others' actions can be achieved motorically. Yet, there are issues standing in the way of a motor theory of goal-tracking. The next section details some theoretical, as well as practical issues facing the action observation paradigm.

4.5.4. Theoretical challenges

A problem for Ambrosini et al.'s action observation paradigm exists in the form of pushback to the theorised involvement of the motor system and its processes for representing goals. Csibra (2008) has long maintained that a motor theory of social cognition faces significant hurdles for the following reasons. Firstly, its claim that, via motor resonance, visual information is directly transformed to goal information by mirror neurons begs the question: how do neurons *know* which specificity level of an action to encode? It is unclear at which level direct matching requires the mirror system to match the observed motor action (e.g., does this involve detail of motion speed, precise joint configuration, exact angle of movement direction etc.).

Secondly, if it is in fact the case that the motor system receives input containing goal information, then it cannot be claimed that motor representations are the primary ways for action understanding. Further unpacking this point, Csibra (2008) argues that the recruitment of motor representations in action observation does not necessarily translate to understanding the goals of an action. His reasoning is that while one's motor representations may emulate a suitable motor program for achieving the goal of the agent, the motor system itself does not generate a representation of the goal. It is just as likely that the meaning of an action is generated outside of the motor system first; only then is the motor system able to simulate the action that was observed, thereby producing motor representations. In other words, motor activations may be mere by-products of top-down cognitive processing as a result of abstract inferences and judgements about the intentionality of others' actions (Carruthers, 2017; Pomiechowska & Csibra, 2017). While both processes are simulative in nature, Csibra prefers this topdown inferential process of action reconstruction whereby motor system activation and the efficient emulation of action is a result of core mindreading systems. This is as opposed to direct motor matching-which is more of a bottom-up process-whereby motor system activation and the simulation of action is a cause of action understanding.

Butterfill, Low and their colleagues' (e.g., Butterfill & Apperly, 2016; Butterfill & Sinigaglia, 2012; Low, et al., 2020) view that it is only efficient processing of others' mental states—tracked in terms of agent-object relations—that could interface with motor representations. For instance, (1) tracking the outcome to which an action is directed rather than representing intention as such, (2) tracking preference rather than representing desire as such, or (3) tracking registration rather than representing belief as such. Nonetheless, a theoretical hurdle is how the interfacing could operate when there are no clear motor characteristics in common between motor plans and minimalist mental states, like preference or registration. The focus of this thesis, however, will not be to tackle the theoretical challenges. This thesis will look more closely at whether Ambrosini et al.'s action observation paradigm is robust; replication problems could

67

mean that even the more fundamental view, that the goals of others' actions are represented motorically, may be practically challenged.

4.5.5. Practical challenges

As stated in the previous section, the action observation paradigm was originally created by Ambrosini et al. (2011), and was validated by the same authors by testing for interference effects (Ambrosini et al., 2012), as well as by testing both infants and adults (Ambrosini et al., 2013). Outside of these three studies from the same laboratory, a recent study by Vannuscorps and Caramazza (2017) has provided a significant challenge to Ambrosini et al., due to a failed replication of one of their studies.

Using Ambrosini et al.'s (2011) stimuli, Vannuscorps and Caramazza (2017) attempted to conceptually replicate the original findings with a case study of an individual born without arms (D.C.). Unusually, D.C. exhibited *proactive* gazes like those in Ambrosini et al.'s study. Conceptually, assuming direct matching processes, D.C. should have shown *reactive* gaze behaviour since it was physically impossible for him to perform the observed action (i.e., hand reaching and grasping an object). Even more pertinent was the direct replication data from Vannuscorps and Caramazza's typically-developing adult participants (N = 10) acting as a control group. The researchers found strong evidence in favour of the null hypothesis, suggesting no relationship between the proactive gaze of the controls and D.C. (the alternative hypothesis was that D.C.'s eye movements would be slower than controls).

More worryingly, in terms of the direct replication attempt, the control participants' gaze proactivity in the no-shape trials turned out to be just as fast as their gaze proactivity in the pre-shape trials, contradicting Ambrosini et al.'s (2011) original findings. This cannot be reconciled by the direct matching hypothesis. A telling comparison is that Vannuscorps and Caramazza's (2017) control participants displayed only a 15 ms difference in gaze proactivity between the shape and no-shape conditions, whereas Ambrosini et al.'s (2011) control participants showed a 153 ms difference between the conditions.

The failure to directly replicate findings of adults' proactive eye movements in relation to goal anticipation is worrying, as it echoes the wider replication crisis flagged in section 1.6. Vannuscorps and Caramazza's (2017) findings could suggest that difficulties associated with non-verbal measures of mental state understanding may be the tip of an iceberg that extends from belief-tracking scenarios to goal-tracking scenarios. More specifically, their findings with D.C., who was born without upper limbs, are a challenge to the idea that motor simulation is vital for the emergence of biomechanical constraint effects (Ostarek & Bottini, 2021). What processes were underlying D.C.'s predictive eye movements, considering his physical inability to simulate the observed action? Do those with congenital motor deficits form compensatory motor representations for actions that use limbs they are born without? There are more questions than answers (some answers will be offered in Chapter 8). Of course, it is important to acknowledge that there has been only one attempt to independently replicate Ambrosini et al.'s general findings (i.e., by Vannuscorps & Caramazza).

Given the wider replication challenges currently facing the psychological science field, more replication studies are needed. To forecast, for its first experiment this thesis will seek to replicate Ambrosini et al.'s (2013) findings of adults' ability to rapidly anticipate a likely goal, using the action observation paradigm. This will act as one point of reference for triangulation (whereby multiple measures and methods are used to gain an increased understanding of a concept or construct). Of course, even if the action observation paradigm turns out to be less robust or reliable than previously assumed, there may still be converging evidence—from other related paradigms—of motor representations having good explanatory power for the tracking of others' goals.

4.6. Grasp cueing paradigm

Related to Ambrosini et al.'s (2011; 2013) work, Fischer, Prinz, and Lotz (2008) developed a grasp cuing paradigm to demonstrate that when we see someone's specific hand gestures, certain motor cues automatically direct our attention to corresponding object outcomes. The authors took inspiration from motor theorists such as Craighero,

Bello, Fadiga, and Rizzolatti (2002), who highlighted the importance of the human hand as a cue for directing attention. Fischer et al. presented still frames consecutively, rather than moving stimulus videos (like the Ambrosini et al. studies in the previous section). This method of display allowed for control over visual stimulation across different conditions.

Participants carried out a detection task in which they observed a resting hand (with closed fist) positioned at the centre of a screen. There was one small and one large object at the bottom corners of the screen. The hand then displayed either a pincer grasp or a whole-hand prehension, and the next image showed the hand in its resting (closed fist) state. Subsequently, a target shape (yellow star) appeared on either the small or the large object, and participants had to press the space bar of the keyboard with their right hand when they detected the yellow star. Reaction times were measured by way of the space bar press after the onset of the target (superimposed on one of the objects). The appearance of the target star had a variable time delay and was unpredictable, that is, it would randomly appear over either of the two objects. Grasp congruency refers to the pincer grasp being congruent with the small object and the whole-hand prehension being congruent with the large object. See Figure 13 for a visual display of the procedural stimuli.



Figure 13 | Sample event sequence. The left grasp gesture panel depicts an incongruent trial (pincer grasp followed by target on large object), and the right grasp gesture panel depicts a congruent trial (whole-hand prehension followed by target on large object).

From Fischer et al. (2008) study 1.

Despite grasp congruency being uninformative to target location—and participants were made aware of this—Fischer et al. (2008) discovered a cueing effect whereby response times were speeded at certain time delays. Specifically, after both the 200 ms and the 300 ms delays, participants were faster to detect targets appearing on the object that was congruent with the motor cue (i.e., congruent with the previously shown kind of grasp). The fact that observers quickly inferred the target object (which could be construed as the agent's goal), upon paying attention to the agent's specific hand grasp, makes sense in terms of the motor system being engaged automatically during action simulation. It is possible, however, that the motor-facilitated cuing effect may simply be driven by strategically learning that a whole-hand prehension (large grasp gesture) would be associated with the target being likely to appear on the large object, and that a pincer grasp (small grasp gesture) would be associated with the target being likely to appear on the small object. It might well be that participants' performances reflected strategic inferences rather than spontaneous or automatic representation of motor actions.

In a second experiment, Fisher et al. (2008) re-designed the task stimuli such that the motor cues were anti-predictive (i.e., the whole-hand prehension gesture was followed in 80% of trials by the target appearing over the small object). Participants showed faster detection of the congruent trials only when there was a 300 ms delay interval between motor cue and target onset. During the 100 ms delay interval, participants showed significantly faster detection for the incongruent trials. Fischer et al. explained that the strategic allocation of attention seemed to occur first, and this high-level (or top-down) strategic control of attention was only later over-ridden by automatic motor processing. Fisher et al.'s interpretation of their second experiment echoes points raised by Csibra (2008), that meaning may be generated outside the motor system first; only then is the motor system able to simulate the action that was observed, thereby producing motor-representation effects.

4.6.1. Static versus dynamic

It is worthwhile to note here that Fischer et al.'s (2008) task made use of still frames that implied movement (shown one after the other), unlike the Ambrosini et al. (2011; 2012; 2013) tasks, that used video stimuli. While many studies looking at the mirror neuron system use either video sequences or live experimenters (dynamic methods), there is evidence supporting the use of combined still images (static method). Stürmer, Aschersleben, and Prinz (2000; Experiments 1 and 2) compared static visual primes with dynamic visual primes. They found that the visuo-motor priming effect was, in fact, stronger in the static condition. The grasp cuing paradigm—in utilising a combination of still frames, rather than moving videos—may be effective in controlling for differences in timing and visual stimulation across stimuli conditions. There may well be a higher chance of replication success with the Fisher et al. grasp cueing paradigm.

Overall, the present behavioural evidence from Fisher et al.'s (2008) grasp cueing paradigm is consistent with the findings from Ambrosini et al.'s (2011, 2012, 2013) action observation paradigm, with both methods revealing that adults encode relevant motor information to latch onto the goals of others' actions. Consequently, to forecast, for its second experiment, this thesis will seek to replicate the findings of Fischer et al.'s findings (i.e., their Experiment 1 findings), but using stimuli from Ambrosini et al. (2013). This will act as a second point of reference for triangulating the research methods across this thesis. The third point of reference will come from Experiment 3, which will look at explicit, rather than implicit evaluations of the same stimuli.

4.7. Replicate, replicate!

Replication is not only important for mindreading, but also for the wider psychological sciences in general. In the midst of recurring replication issues (see section 1.6.), it is important that conceptual replications are conducted, as well as direct replications. Take, for instance, Kulke et al. (2018a; 2018b) who carried out replication studies of anticipatory looking false belief tasks, making sure to use original stimuli and sufficient sample sizes. Furthermore, the use of different methodological procedures to test a hypothesis allows for converging evidence, which bolsters the theoretical claims drawn from such studies. The current thesis will include both direct and conceptual replications.

One thing to note is that, while adults are definitely more capable of mindreading than infants and children, researching adults is still a useful endeavour to tease out the possible system structure of mindreading (Apperly, 2011). Just as there is hype around the mindreading that children *can* do, it is informative to find out the limits of mindreading in adults, too. The experiments of the current thesis hence tested adult participants.

As briefly mentioned in this chapter (see Chapters 5, 6, and 7 for in-depth detail), Experiment 1 of this thesis is a direct replication of Ambrosini et al.'s (2013) study that investigated goal-tracking using the action observation paradigm; it used the same stimuli as Ambrosini et al., and a sufficiently large sample size. Given that there are already replication problems with the Ambrosini et al. action observation paradigm, securing motor effects from a different but related paradigm would go some way towards establishing and providing convergent evidence for motor representations actually being involved in some cases of goal ascription. Thus, Experiment 2 is a conceptual replication whereby Fischer et al.'s (2008) grasp cueing paradigm is combined with Ambrosini et al.'s stimuli. Experiment 3 takes yet a different approach, by garnering adults' explicit judgements of the Ambrosini et al. stimuli.

4.8. Conclusion

From looking more closely at the motor system, it is apparent that motor representations are an important part of how we represent actions. Mirror neurons and the direct matching account offer an explanation for the process through which we represent actions motorically. While the recent and bold idea that mindreading may be motorically grounded has some evidence, it requires more empirical backing. Crucially, Butterfill, Low, and colleagues' recent experiments on the possibility of motorically grounded mindreading—which already jumps ahead to investigate belief-tracking—rests on the assumption that Ambrosini et al.'s collection of studies (2011; 2012; 2013) reliably tests goal-tracking. However, Ambrosini et al.'s action observation paradigm meets both theoretical and practical challenges. In this regard, the structural layout of Fischer et al.'s (2008) grasp cueing paradigm offers a way in which to further test and extend the action observation paradigm. Now that these first four chapters have set up the rationale for the experiments of the current thesis, the next three chapters describe those experiments in detail.

Chapter 5.

5.1. Introduction

The current chapter will detail Experiment 1, going over key background literature to set up the rationale and context of the experiment. After this, the experimental hypothesis is stated. Then, the methodology, including procedural information, is described. Results of the current experiment are then presented. Finally, the results are briefly discussed before leading on to the next experiment (more in-depth discussion awaits in Chapter 8).

Experiment 1 was a direct replication of Ambrosini et al.'s (2013) study, albeit with adults only (i.e., did not include infants). To briefly summarise the key details of Ambrosini et al.'s study from section 4.5.3., their adult participants completed the action observation paradigm, whereby they viewed an agent reaching for one of two objects. The response measured was gaze proactivity to the target object, meaning how fast participants' gaze landed on the appropriate object before the agent's hand reached it. The authors found a main effect of pre-shape advantage, that is, participants' eye movements were faster to latch onto the target objects when the hand pre-shaped (displayed a pincer grasp gesture or whole-hand prehension gesture), as opposed to the no-shape (displayed closed fist) control condition. The authors interpreted this as evidence that motor cues (in this case, specific hand gestures) facilitate action anticipation when observing others. That is to say, we can identify the target object of a goal-directed action more quickly when cued by an appropriate grasp. This is theorised to happen because of the involvement of motor representations that become activated in the observer. More broadly speaking, some researchers have considered that Ambrosini et al.'s findings may feed into the possibility that mirror neurons may also play a part in this motor matching effect (e.g., Rizzolatti & Sinigaglia, 2010).

The 2013 study was spotlighted because Vannuscorps and Caramazza (2017) already failed to replicate Ambrosini et al.'s initial (2011) study. It is worth briefly reminding ourselves of these findings: D.C., an individual born without upper limbs, showed proactive gaze even though direct matching processes would predict this to be theoretically impossible. More pertinently, Vannuscorps and Caramazza's typicallydeveloping adult control participants were just as proactive (eyes reaching target object before the hand) in the no-shape trials as they were in the pre-shape trials. This contradicts Ambrosini et al.'s motor cueing finding, as the pre-shaped hand gestures did not facilitate gaze proactivity more so than the no-shape hand gesture.

Ambrosini et al.'s 2011 (and 2012) study showed an agent reaching to one of two tomatoes, whilst their 2013 study showed an agent reaching to one of two green plastic balls. The 2013 study is also versatile in that it can be used to test various age groups across development. More relevantly though, it is possible for the Ambrosini group to claim that the 2013 stimuli offer the most methodological (and theoretical) merit because the study revealed ontogenetic evidence for direct matching: there was proactive tracking of an agent's goal-related action in infants (as young as 8 months of age), as well as adults. Their 6-month-olds also showed a motor-matching effect, but this effect was only most apparent amongst individual infants who were already capable of precision grasping. Experiment 1 thus sought to replicate Ambrosini et al.'s (2013) findings. Following on from Ambrosini et al.'s findings of a pre-shape advantage in adults, it was hypothesised that, in the current experiment, adults should show faster gaze proactivity in the pre-shape trials (compared to the no-shape trials), with significant proactivity revealed for both types of pre-shaping.

5.2. Methodology

5.2.1. Participants

Using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007), an a priori power analysis of Ambrosini et al.'s (2013) findings (the effect size of hand shape for adults was $\eta_{p^2} = 0.49$) indicated that a sample size of 47 participants was sufficient to detect a main effect of a pre-shape advantage in a one-way within-subjects analysis of variance (ANOVA). As a precautionary step, to offset data attrition from following Ambrosini et al.'s (2011; 2013) strict criteria regarding what is counted as a valid gaze trial for data analysis, a larger sample size was recruited. The final sample were comprised of 74 tertiary students (age range: 18 -29 years; mean age: 18.67 years, SD = 1.55) recruited through the *Introduction to Psychology Research Programme* at Victoria University of Wellington. All participants were naïve to the aims of the experiment and had normal or corrected-to-normal vision.

Ethical approval was obtained from the School of Psychology Ethics Committee, under delegation of the Victoria University Human Ethics Committee. Participants provided written informed consent prior to commencing the experiment and were debriefed on the purpose of the experiment after completion.

5.2.2. Materials

Stimuli

A copy of the original stimuli was obtained through communication with Ambrosini (stimuli from Ambrosini et al., 2013; Figure 14). The videos show a hand reaching and contacting either a small green ball or a large green ball. The small ball and the large ball acted as the target objects. The four possible hand gestures were as follows: (1) pincer grasp to the small object, (2) whole-hand prehension to the large object, (3) closed fist to the small object as a control, and (4) closed fist to the large object as a control. Importantly, the beginning portion of the reaching movement was the same for all grasp types, meaning participants could not simply guess in advance which object would be congruent. All combinations were semi-randomised and counterbalanced, with half the trials being no-shape (closed fist) and half being preshape (pincer grasp or whole-hand prehension). Since the different hand gestures were randomised, that meant grasp was unpredictable to participants. Each trial began with a central fixation cross that lasted for a random variable delay of either 3500, 4000, or 4500 ms (to avoid mechanistic responding), followed by a fixation cross on the actor's hand that lasted 1000 ms. The experiment was fully computerised.

77



Figure 14 | Snapshots of stimulus videos involving two green ball objects. This figure shows the target layout whereby the large object is near the hand (object position was counterbalanced). The rows depict the following variations (from top to bottom: no-shape gesture to small target, pincer grasp to small target, no-shape gesture to large target, whole-hand prehension to large target). The white circle in the left-hand column depicts the area of interest superimposed on the target object.

From Ambrosini et al. (2013).

Apparatus

Participants' eye movements were recorded using Tobii X2-30 eye-tracking software installed on a laptop computer. Participants were individually tested in a quiet room. Each participant was seated and had their head in a chin rest, centred at an approximate viewing distance of 54 cm from the colour screen display, for the duration of the trials. At the beginning of the experiment, gaze position was calibrated with Tobii software using a standard 9-point procedure (as per Ambrosini et al., 2011).

5.2.3. Procedure

Once gaze calibration was completed, the stimulus videos were presented to participants. Similar to Ambrosini et al. (2013), the total number of trials per participant was 64.

The experiment adopted a within-subjects design. Before the formal start of the trials, participants were shown four practice videos (one from each of the four conditions). They were given time to ask any questions about the task. Participants were instructed to look at the fixation crosses whenever they appeared on screen until their disappearance, and then to simply watch the videos.

5.3. Results

5.3.1. Data reduction

Areas of interest (AOIs) were created for each stimulus video: the hand AOI was dynamic, covering the actor's hand (whilst resting and then as it moved towards the objects); the *small* AOI was static, covering the small object; the *large* AOI was also static, covering the large object. The data were reduced according to the following exclusion criteria. Data were excluded if participants (1) looked at the hand for less than 0.2 seconds, (2) looked at the target object for less than 0.2 seconds in pre-shape trials (i.e., small object in pincer trials or large object in whole-hand trials), (3) looked at neither target for 0.2 seconds or longer in no-shape trials, (4) did not look at the hand during the fixation cross on the hand, and/or (5) already prematurely looked at either object within the first 500 ms of the hand moving. Based on these criteria, 17% of the total trials were excluded. It is worth mentioning that mindreading studies that use eve gaze often face the problem of large exclusion rates; recall, for instance, Kulke et al.'s (2019) high exclusion rates noted in section 1.6., when replicating several action anticipation studies. Effectively, while infant studies report participant exclusion due to general inattention or fussiness, there will be cases of inattention with adult participants, but we can only control for this through the proxy of exclusion criteria.

Following Ambrosini et al. (2013), two 'time to first fixation' values were calculated for each participant: one for their time to first fixation within the small AOI, and another for their time to first fixation within the large AOI. Gaze proactivity for each trial was also calculated, by subtracting the time to first fixation within target AOI (i.e., small AOI for pre-shape–small target trials; large AOI for pre-shape–large target trials) from the time at the last frame (end of movement phase; see Figure 14). Thus, negative looks were regarded as proactive since gaze arrived within the target AOI before the actor's hand.

5.3.2. Data analysis

The initial analyses followed the analyses reported by Ambrosini et al. (2013). Analyses were performed using SPSS (version 27), and values have been rounded to two decimal places where applicable. Action anticipation was initially analysed by performing a repeated-measures ANOVA on gaze proactivity with hand shape (preshape vs. no-shape) and object (small vs. large ball) as within-subjects factors. There was a significant main effect of hand shape (F(1,65) = 10.64, p = .01, $\eta_{p^2} = .14$) and object (F(1,65) = 131.13, p < .001, $\eta_{p^2} = .68$; Figure 15). The shape × object interaction was not significant (F(1,65) = .09, p = .77, $\eta_{p^2} = .01$).



Figure 15 | Time of gaze proactivity (with 1 standard error mean, henceforth SEM) at the target object, relative to the end of the trial video. Gaze proactivity is plotted as a function of target size and hand shape. The end of the trial video is represented by the horizontal line at zero seconds. Negative values represent proactive eye movements.

A follow-up paired t-test confirmed that proactive eye movements were faster to the large object than to the small object (t(64) = 12.81, p < .001, Cohen's d = 1.59). A paired t-test also confirmed that, contrary to the hypothesis based on Ambrosini et al. (2013), proactive eye movements were faster for the no-shape (closed fist gesture) condition than for the pre-shape condition (t(65) = 3.26, p = .01, Cohen's d = .40). In fact, Figure 15 also showed that gaze proactivity was fastest when participants observed the closed fist (no-shape control) moving towards the large object; here, a simple perceptual strategy could be driving participants' gaze responses (i.e., the visuospatial properties of the closed fist may perceptually match the large object). Overall, it appears that gaze responses during action observation are modulated by object size (or goal salience), rather than necessarily by the motor properties of the agent's particular hand grasping gesture.

It is important to note that Ambrosini et al. (2013) did not consider distance of the object from the agent's hand as a factor worthy of investigating. It is reasonable to assume that, because of the stimulus layout (recall that object position was counterbalanced), the position of the target object as being closer to or farther away from the reaching hand could influence the proactivity of participants' eye gaze. Thus, a supplementary 3 (grasp: pincer grasp vs. whole-hand prehension vs. fist) × 2 (object) × 2 (distance: target nearer to approaching hand vs. target farther from approaching hand) ANOVA on gaze proactivity was conducted. There was a significant main effect of object (F(1,3) = 81.31, p < .01, $\eta_{p^2} = .96$). Main effects of grasp (F(2,6) = 2.94, p = .13, $\eta_{p^2} = .50$) and distance (F(1,3) = 8.28, p = .06, $\eta_{p^2} = .73$) were not significant. In addition, there was a significant distance × object interaction (F(1,3) = 17.70, p = .03, $\eta_{p^2} = .86$; Figure 16), but no further interactions were significant.



Figure 16 | Time of gaze proactivity (with 1 SEM) at the target object, relative to the end of the trial video. Gaze proactivity is plotted as a function of target size and distance from the agent's hand. The end of the trial video is represented by the horizontal line at zero seconds. Negative values represent proactive eye movements.

Follow-up paired t-tests, with Bonferroni corrections, confirmed that proactive eye movements were significantly different when the target object was near versus far from the hand for the small object t(64) = -6.75, p < .001, Cohen's d = -.84, but not for the large object t(65) = -.37, p = .71, Cohen's d = -.05. In other words, when the small object was the target, participants showed—regardless of the availability or type of motor cue—earlier gaze arrival times when the object was near the hand compared to when the same object was far from the hand. Figure 16 also showed that when the large object was the target, participants showed early gaze arrival times regardless of where the object was located in relation to the hand, suggesting that a large target is attentionally salient.

5.4. Discussion

To recapitulate, at the highest level (the first ANOVA), the current experiment found effects of shape and object. The main effect of shape showed that participants' gaze was more proactive in no-shape trials than pre-shape trials. Note that this is the opposite to what Ambrosini et al. (2013) found. The main effect of object showed that participants' gaze was more proactive to the large object than the small object. This was in line with Ambrosini et al.'s findings; the authors did note that this object effect could be due to the visual salience of the large object (this is discussed further in Chapter 8). Of particular interest was the finding that proactive looks were significantly faster for no-shape trials than pre-shape trials, and significantly faster to the large object than to the small object. This means that, overall, fastest proactivity occurred in what was supposed to be the control trial (i.e., no-shape, closed fist hand gesture).

Importantly, when adding the distance of the target to the reaching hand to the analyses—which Ambrosini et al. (2013) did not consider as a factor—proactive eye movements to the small object were significantly faster when the object was nearer to the agent's hand, compared to when it was farther away. This means that, regardless of the kind of grasp, the small object garnered more proactive looks when it was closest to the hand. This does not support the direct matching hypothesis, according to which

motor cues and congruency between grasp type and object size should be what facilitates gaze proactivity.

Whilst the results were mixed compared to Ambrosini et al. (2011; 2012; 2013), the data dovetailed with the failed replication attempt by Vannuscorps and Caramazza (2017). The findings align with other research suggesting that responses during action observation are sensitive to object type and salience (such as physical size), instead of motor cues (e.g., Henrichs, Elsner, Elsner, & Gredebäck, 2012; Naish, Reader, Houston-Price, Bremner, & Holmes, 2013; Proulx, 2010). Indeed, Henrichs et al. found that infants show gaze shifts significantly earlier when they observed a hand (open palm facing downwards) reach for a large object, compared to when it reached for a small object.

The current findings raise significant problems for Ambrosini et al. at a task analysis level, but also present theoretical worries for other researchers (e.g., Butterfill, 2020; Low et al., 2020; Zani et al., 2020), who assume that the evidence on the action observation paradigm reliably showcases motor representations playing a role in adults' efficient mindreading of others' goal-directed actions (at the very least). Further, what does this mean for the theoretical argument used to explain the original study's findings, namely, the direct matching hypothesis? This non-replication also has implications for the idea that mirror neurons form the basis of a motor matching account, whereby we simulate the observed actions of others which in turn facilitates our action prediction abilities.

Despite these findings, all is not yet lost for the action observation paradigm. It could be that the presentation format of the stimuli creates noise in the data, since one cannot precisely control for differences between videos using a live experimenter. Perhaps using the same stimuli, but altering how they are shown to participants via an attention cueing paradigm (Fischer et al., 2008), may show motor effects. Experiment 2 goes on to test this idea, by way of a conceptual replication of Ambrosini et al. (2013).

84

5.5. Conclusion

Experiment 1, a direct replication of Ambrosini et al.'s (2013) study, showed that their classic action observation paradigm is not as robust as previously thought. The next chapter details an experiment to further probe the validity of this paradigm.

Chapter 6.

6.1. Introduction

The current chapter details Experiment 2, starting with summarising a key background study to set up the rationale and context of the experiment. After this, the experimental hypothesis is stated. Then, the methodology, including procedural information, is described. Results of the current experiment are shown via statistical analyses. Finally, the results are briefly discussed. Chapter 8 will contain a more indepth discussion.

Due to the failure of Experiment 1 in replicating the action observation paradigm, the next step was to investigate whether presenting the same stimuli in a different format would result in a successful replication. This is one way to tease apart whether it is the stimuli themselves that are problematic, or whether alternative procedural changes would allow the stimuli to show the desired effects, according to the direct matching hypothesis and motor simulation. Recall Fischer et al.'s (2008) grasp cueing paradigm from section 4.6. (see previous Figure 13), in which motor cues (i.e., specific hand gestures) directed participants' attention to the corresponding object outcome.

To briefly summarise the key details of Fischer et al.'s (2008) study, adults completed a target detection task. In this task they observed static images of a hand displaying either a pincer grasp or a whole-hand prehension, followed by a resting state gesture (an uninformative closed fist). Subsequently, a target symbol (yellow star) appeared on a large or small object and participants' reaction time to indicate the target's appearance was measured. A grasp cueing effect was found, such that response times were speeded at certain time delays; after 200 ms and 300 ms delays (but not at the other 0, 100, or 450 ms delays). Participants were faster to detect targets appearing on the object that was congruent with the motor cue (e.g., pincer grasp paired with yellow star on the small object).

This grasp cueing effect—whereby speed of inferring the target object aligned with attention to the agent's specific hand gesture—makes sense in terms of the motor

86

system being engaged automatically during action simulation. Additionally, the fact that Fischer et al. (2008) found motor-relevant effects with static images is useful from a procedural standpoint. In the Ambrosini et al. (2013) video stimuli used in Experiment 1, there may be subtle differences in the timing of the hand's movement to the objects between the various conditions. The use of static (still images), rather than dynamic (videos) stimuli in Fischer et al.'s study makes it easier to control for timing differences across conditions.

This experiment weaves together the stimuli from Ambrosini et al. (2013) and the procedure of Fischer et al. (2008). While Experiment 1 was a direct replication, Experiment 2 is a conceptual replication. The current experiment aims to extend the action observation paradigm by taking a different procedural approach than the original (Ambrosini et al., 2013) task. Will there be a grasp cueing effect when Ambrosini et al.'s stimuli are presented to participants in the format of Fischer et al.'s target detection task? Extrapolating from the direct matching hypothesis (i.e., a motor simulation account), the benefits of the capacity to use grip aperture for target prediction should be apparent in a target detection task. Therefore, the expectation here was that grasp type would still act as a motor cue and orient participants' attention to the congruently shaped object. It was hypothesised that a pincer grasp would facilitate anticipation towards the small object, and a whole-hand prehension would facilitate anticipation towards the large object.

6.2. Methodology

6.2.1. Participants

An a priori power analysis using G*Power 3.1 (Faul et al., 2007) revealed that a sample size of 75 would be sufficient in order to run analyses similar to Fischer et al. (2008). As a precautionary step, the sample size was slightly increased to offset attrition.

The final sample were 78 tertiary students (age range: 17 - 34 years; mean age: 19.39 years, SD = 2.64) recruited through the *Introduction to Psychology Research*

Programme at Victoria University of Wellington. All participants were naïve to the aims of the experiment and had normal or corrected-to-normal vision.

Ethical approval was obtained from the School of Psychology Ethics Committee, under delegation of the Victoria University Human Ethics Committee. Participants provided written informed consent prior to commencing the experiment, and were debriefed on the purpose of the experiment after completion.

6.2.2. Materials

Stimuli

Stimuli were taken from Ambrosini et al. (2013); screenshots of the original videos supplied by the authors were shown in succession, as depicted in Figure 17, in line with the procedure of Fischer et al.'s (2008) grasp cueing paradigm. As before, the two objects were a large green ball and a small green ball; the hand (always the right hand of the actor) displayed three different gestures: closed fist (control), pincer grasp, or whole-hand prehension. Participants watched scenes of the stimuli in colour.

In each trial, following the event sequence from Fischer et al. (2008), the hand was first shown in a resting posture (closed fist resting on table) for 1000 ms, alongside the small and large object. Next, the hand was shown further along the reaching movement trajectory, in either a pincer grasp or whole-hand prehension, for 150 ms. Then, the resting posture (fist in same position along movement trajectory) was shown for delays of 0, 100, 200, 300, or 450 ms, before the target appeared superimposed on one of the two objects. The purpose of this variable delay interval was to test for automaticity of the grasp cueing effect. Just as in Fischer et al.'s study, the target was a yellow star, and all stimulus scenes were assembled from still images. The experiment was fully computerised.

All variables (grasp type, target delay, target location, image orientation, and object orientation) were counterbalanced, randomised, and appeared equally often. In this way, note that grasp type was unpredictive of target object. Each participant completed six practice trials and 160 experimental trials. In addition, 10% catch trials

88

prevented anticipatory responding. Catch trials involved no target appearing, therefore participants were not to make a button-press response in such instances.

Apparatus

The experiment was created and run through Open Sesame software (version 3.2; Mathôt, Schreij, & Theeuwes, 2012). Participants were seated in front of a 19-inch colour monitor and standard keyboard.

6.2.3. Procedure

This experiment adopted a within-subjects design. Participation took place in a quiet room at the same time of day (i.e., morning). Following Fischer et al.'s (2008) procedure, participants were instructed to hover over the response key in order to reduce response time by eliminating any movement time (M. Fischer, personal communication, February 12, 2019; as a shorter response time means less variability). Participants responded by pressing the space bar of the keyboard with their right hand when they detected the target (yellow star). They were instructed to respond fast without making errors in the catch trials (where no target appeared). Reaction times were recorded in milliseconds, from target onset to keyboard response. Before the formal start of the trials, participants were shown six practice trials (randomised). They were given time to ask any questions about the task. Then, participants were instructed to follow the prompts on screen to start the experiment. The stimulus clips (collated still images in quick succession; Figure 17) were presented, and there was a rest break halfway through the trials. All instructions were displayed on screen as part of the experiment.



Figure 17 | Sample event sequence for Experiment 2, depicting variations of incongruent trials (top panels show incongruent pincer grasp trial; lower panels show incongruent whole-hand prehension trial). Congruent trials would have target on small object for pincer grasp and target on large object for whole-hand prehension.

6.3. Results

Catch trial responses were infrequent (there was a 5% response rate across all participants for catch trials), suggesting participants were concentrated on the task and following instructions.

6.3.1.Data reduction

As per Fischer et al.'s (2008) criteria, we accepted response times for analysis that were greater than 100 ms and within 2.5 standard deviations of each participant's mean. Catch trial data were not included in analyses, as their purpose was to serve as an attentional check.

6.3.2. Data analysis

The initial analyses, as far as possible, followed the analyses reported by Fischer et al. (2008). Analyses were performed using SPSS (version 27), and values have been rounded to two decimal places where applicable. A 2 (congruency: congruent vs. incongruent) × 5 (target delay: 0, 100, 200, 300, 450 ms) ANOVA confirmed a significant delay effect F(4,308) = 76.63, p < .001, $\eta_{p^2} = .50$. These results are depicted in Figure 18. The congruency effect was not significant F(1,77) = .57, p = 45, $\eta_{p^2} = .01$. The delay × congruency effect was also not significant F(4,308) = 1.50, p = .20, $\eta_{p^2} = .02$.



Figure 18 | Mean reaction times (RTs, in ms, with 1 SEM) to detect targets in Experiment 2 with unpredictive grasps that were congruent or incongruent with the object over which the target appeared (i.e., congruent: pincer grasp and small object or whole-hand prehension and large object; incongruent: pincer grasp and large object or whole-hand prehension and small object). Target delay refers to the onset of a yellow star superimposed over target object, in ms.

It is important to note that in the Ambrosini et al. (2013) stimuli, the objects (large ball and small ball) were not apparently equidistant to the hand; one of the objects appears closer to the hand and this is exacerbated by the camera angle adopted by Ambrosini et al. when creating their stimuli. Consequently, rather than grip aperture, attentional capture by the object that appears closest to the hand may be driving potential congruency × grasp reaction time effects. Therefore, a supplementary congruency × grasp × distance (object near to hand) ANOVA was performed to check for near-space advantages. There was a significant three-way congruency × grasp × distance (small object vs. large object near to hand) interaction F(1,77) = 49.38, p < .001, $\eta_p^2 = .39$. Separate two-way congruency × distance ANOVAs were carried out for each of the grasp types (pincer and whole-hand) to interpret this three-way interaction.

For pincer grasp trials, there was a significant congruency × distance interaction F(1,77) = 23.01, p < .001, $\eta_{p^2} = .23$ (see Figures 19 and 20). When the large object was near the hand, mean reaction times were faster for incongruent trials (when target star appeared on large object despite pincer grasp; M = 307.36, SD = 49.22), than congruent trials (when target star appeared on small object matching pincer grasp; M =318.55, SD = 49.45). When the small object was near the hand, mean reaction times were faster for congruent trials (when target star appeared on small object matching pincer grasp; M = 312.82, SD = 51.76), than incongruent trials (when target star appeared on large object despite pincer grasp; M = 317.64, SD = 53.13).



Figure 19 | Mean reaction times (RTs, in ms, with 1 SEM) for pincer grasp trials to detect targets in Experiment 2 that were congruent or incongruent with the object over which the target appeared, when the object near (closest to) the hand was large or small. Points A–D correspond to the images in Figure 20.



Figure 20 | Visual representation of four possible outcome configurations for pincer grasp trials. Prior to each of these images, the hand displayed a pincer grasp. (A) depicts a congruent trial with the large object near to hand; (B) depicts an incongruent trial with the small object near to hand; (C) depicts a congruent trial with the large object near to hand; (D) depicts an incongruent trial with the large object near to hand; (D) depicts an incongruent trial with the large object near to hand; (D) depicts an incongruent trial with the large object near to hand; (D) depicts an incongruent trial with the large object near to hand; (D) depicts an incongruent trial with the large object near to hand.

For whole-hand prehension trials, there was a significant congruency × distance interaction F(1,77) = 32.37, p < .001, $\eta_p^2 = .30$ (see Figures 21 and 22). When the large object was near the hand, mean reaction times were faster for congruent trials (when target star appeared on large object matching whole-hand prehension; M = 305.61, SD =47.19), than incongruent trials (when target star appeared on small object despite whole-hand prehension; M = 316.43, SD = 50.12). When the small object was near the hand, mean reaction times were faster for incongruent trials (when target star appeared on small object despite whole-hand prehension; M = 309.37, SD = 50.61), than congruent trials (when target star appeared on large object matching whole-hand prehension; M = 318.50, SD = 49.80).



Figure 21 | Mean reaction times (RTs, in ms, with 1 SEM) for whole-hand prehension trials to detect targets in Experiment 2 that were congruent or incongruent with the object over which the target appeared, when the object near (closest to) the hand was large or small. Points A–D correspond to the images in Figure 22.



Figure 22 | Visual representation of four possible outcome configurations for whole-hand prehension trials. Prior to each of these images, the hand displayed a whole-hand prehension. (A) depicts a congruent trial with the small object near to hand; (B) depicts an incongruent trial with the large object near to hand; (C) depicts a congruent trial with the large object near to hand; (D) depicts an incongruent trial with the small object near to hand; (D) depicts an incongruent trial with the small object near to hand; (D) depicts an incongruent trial with the small object near to hand; (D) depicts an incongruent trial with the small object near to hand; (D) depicts an incongruent trial with the small object near to hand.
Follow-up paired t-tests confirmed that, for pincer grasp, mean reaction times were significantly different between congruent and incongruent trials when (1) the large object was near the hand t(77) = 4.64, p < .001, Cohen's d = .53, and (2) the small object was near the hand t(77) = -1.97, p = .05, Cohen's d = -.22. Similarly, for whole-hand prehension, mean reaction times were significantly different between congruent and incongruent trials when (1) the large object was near the hand t(77) = -4.39, p < .001, Cohen's d = -.50, and (2) the small object was near the hand t(77) = 4.27, p < .001, Cohen's d = .48.

6.4. Discussion

To recapitulate, at the highest level (the first ANOVA), the current experiment only found a significant delay effect, while Fischer et al. (2008) also found a significant congruency effect as well as a congruency \times delay interaction. This is plausibly due to the different stimuli used (see Chapter 8 for more detail). Ultimately, the finding that stands out is that participants' reaction times were faster when the target object was closer to the agent's hand. There was no evidence that grasp type speeded reaction times to congruent target objects, which is what the direct matching hypothesis predicts.

In pincer grasp trials, reaction times were significantly faster when the congruent target (small object) was closer to the hand. On the face of it, this seems to follow a motor matching account, because the motor cue of the hand making a pincer grasp seems to facilitate attention towards the matching small object. If a motor matching account were the parsimonious explanation, then this finding should hold no matter the spatial layout of the objects; however, this is not the case. In pincer grasp trials, when the objects were swapped around (i.e., when the incongruent target object, the large object, was closer to the hand), reaction times were significantly faster for this incongruent large object. This negates a grasp cueing effect. According to a motor-matching account, grasp type should cue faster responding in congruent trials regardless of the position of the objects.

Similarly, in whole-hand prehension trials, reaction times were significantly faster when the congruent target (large object) was closer to the hand. Again, this

finding alone may seem in line with a motor matching account, because the motor cue of the hand making a whole-hand prehension seems to facilitate attention towards the matching large object. However, in whole-hand prehension trials, when the objects were swapped around (i.e., when the incongruent target object, the small object, was closer to the hand), reaction times were significantly faster for this incongruent small object. This does not support a grasp cueing effect. Rather, in this experiment, the object closest to the hand plays a bigger role in drawing the attention of participants.

Thinking back to Ambrosini et al.'s (2013) study, recall that the authors did not consider object layout (i.e., which object was closest to the hand) as a factor in their analyses. However, Experiment 2 discovered that the spatial layout of the two target objects in relation to the agent's hand is indeed a relevant factor. These results do not bode well for the action observation paradigm or the direct matching hypothesis. Alas, it seems that rather than participants' mirror systems enabling rapid action anticipation based on grasp and object congruency, participants' reaction times were speeded when the target object was closer to the hand. It was the mere spatial layout of the objects that swayed participants' responses, not any motor matching effect related to the type of grasp. What does this mean for the involvement of mirror neurons in action observation?

There are supporting research studies that align with the findings here. For instance, Garrido-Vásquez and Schubö (2014) found that adults were faster at executing right-hand computer-key responses to a probe that appeared on a mug located in near space (at the front of a table) than on a mug located in far space (at the back of a table). However, they also found that participants were just as fast at target detection when the near mug was presented with its handle oriented to the left as compared to when the mug's handle was oriented to the right. Strictly speaking, the direct motor matching account would have predicted that a reaction time advantage would be higher when the mug's handle orientation was congruent with the reaching hand. The current findings of this chapter's experiment fit with broader evidence suggesting the following: rather than being purely driven by specific motor representations of the pre-shaping hand, tracking the outcomes of someone else's reaching behaviour may partly operate on the basis of

97

visual attention to the locations of objects in an environment. More specifically, near spaces seem more likely to attract potential action.

Seeing as even the same stimuli, when presented differently, did not manage to successfully replicate any motor effects, Ambrosini et al.'s (2013) stimuli may not be the best way to test motor matching. That is not to say their action observation paradigm as a whole is not valid; instead, maybe differently configured stimuli would help achieve results in line with the direct matching hypothesis. Possibilities for future research, regarding how the stimuli could be altered, are discussed in the final chapter (Chapter 8).

6.5. Conclusion

By combining Ambrosini et al.'s (2013) stimuli and Fischer et al.'s (2008) grasp cueing procedure, Experiment 2 effectively demonstrated a variation in how one can test the classic action observation paradigm. However, the results of this conceptual replication did not support the direct matching hypothesis. It seems likely that the configuration of the stimuli draws participants' attention, hindering any motor matching effect (Chapter 8 discusses in more depth why this may be). The next chapter details the final experiment of the current thesis, which takes a closer—and more explicit—look into how adults perceive and understand the action observation stimuli.

Chapter 7.

7.1. Introduction

The current chapter will detail Experiment 3. Firstly, relevant background literature will be recapped in order to set up the rationale and context of the experiment. After this, the experimental hypotheses are stated. Then, the methodology, including procedural information, is described. Finally, results of the current experiment are presented and briefly discussed before leading on to the general discussion in Chapter 8.

Experiments 1 and 2 both revealed an effect of distance (of target object to the hand), even though they utilised quite different measures and paradigms; Experiment 1 measured gaze proactivity in the action observation paradigm, whereas Experiment 2 measured reaction time by adapting an attentional grasp cueing paradigm. To further unpack this effect, Experiment 3 took a different approach, while still making use of the same Ambrosini et al. (2013) stimuli. The current experiment asked adults to make probability judgements for snapshots of all the different Ambrosini et al. stimulus configuration, participants were shown snapshots that spanned the trajectory of the agent's hand movement (i.e., beginning with the agent's hand at rest and ending with the agent's hand in flight just prior to contacting one of the objects). These probability judgements required of participants were to predict the likelihood of the agent's hand reaching for the small or large object, for each snapshot.

There is precedent—even within the mindreading literature—to elicit adults' explicit judgements about stimuli that originated to be used in more implicit paradigms, as a way of validating whether the stimuli are successful at tapping into the construct they supposedly test. Consider, for instance, Low and Edwards' (2018) study. They recreated a video based on Onishi and Baillargeon's (2005) object location false belief scenario, as well as a movie based on Scott and Baillargeon's (2009) object identity false belief scenario. Adults watched videos showing either an *expected* outcome (i.e., an ending where the agent searched in a manner consistent with her false belief), or an

99

unexpected outcome (i.e., and ending where the agent searched in a manner inconsistent with her false belief). Adults rated the unexpected ending for the Onishi and Baillargeon scenario as being just that: unexpected. This aligned with the manner of interpretation intended by the task developers themselves. It is therefore possible to contend that there is some degree of construct validity to the Onishi and Baillargeon scenario being used with infants.

However, adults were confused by the meaning of Scott and Baillargeon's (2009) object identity false belief scenario. Many participants perceived that the unexpected ending was in fact *expected*, and likewise that the expected ending was actually *unexpected*! These findings underscore how challenging it can be to constrain which aspects of an agent's situation may be relevant for interpreting their behaviour. Low and Edwards' (2018) approach—of soliciting adults' explicit judgements about stimuli used in mindreading tasks with infants—can help researchers constrain and contextualise how infants might potentially interpret the same stimuli, and in so doing, help researchers gain some insight into how and where reasoning changes. Complementing analyses of adults' rapid eye movements and latencies of target detections (as per Experiments 1 and 2) with analyses of adults' deliberations about outcome probabilities (as per Experiment 3), will naturally raise the following issue: does implicit understanding tap distinct processes from explicit understanding, or are implicit and explicit understandings situated along a continuum? This will be further elaborated upon in Chapter 8.

Given the distance effect found in both previous experiments of this thesis, it is hypothesised that in the current study, participants will explicitly judge the target of the agent's reaching hand to systematically be the closest object. Based on Experiment 1 and 2, the hypothesis extends that even in the first picture (depicting the start of the reaching movement where there is not yet any motor cue present), participants will still judge that the nearer object is a more probable target of the action sequence.

7.2. Methodology

7.2.1. Participants

Using G*Power 3.1 (Faul et al., 2007) it was determined that a sample size of at least 34 would provide 90% statistical power for detecting a medium-sized effect (d = 0.5) in the planned comparisons, assuming a paired (two-tailed) t-test with an alpha level of .05. Given the online nature of the study due to Covid-19 interruptions, in order to guard against potential data loss and participant drop out, a total of 60 participants were recruited.

The final sample were 60 tertiary students, as all participants were included in analyses (age range: 17 - 36 years; mean age: 19.25 years, SD = 2.90) recruited through the *Introduction to Psychology Research Programme* at Victoria University of Wellington. They were naïve to the aims of the study.

Ethical approval was obtained from the School of Psychology Ethics Committee, under delegation of the Victoria University Human Ethics Committee. Participants provided online informed consent prior to commencing the experiment and were virtually debriefed on the purpose of the experiment after completion.

7.2.2. Materials

Stimuli

Scenes from Ambrosini et al.'s (2013) video stimuli were extracted using the video-to-picture conversion function in VLC media player. These extracted scenes had the same width and height as the original videos, and the recording ratio was set such that five scenes were extracted from each video (there were eight different videos in total, corresponding to the eight different trial types). These five scenes were selected to span the trajectory of the hand movement, with the exception of the end contact phase. In other words, participants did not see an image showing the outcome scenes where the hand actually contacted one of the objects, because getting participant evaluations when the hand had already made a choice would not have provided useful data. The five

scenes, for four of the eight trials, are depicted in Figure 23. The experiment was fully computerised.



Figure 23 | Stimulus images involving two green ball objects. This figure shows the target layout whereby the large object is near the hand (object location was counterbalanced). The rows depict the following variations (from top to bottom: no-shape gesture to small target, no-shape gesture to large target, pincer grasp to small target, whole-hand prehension to large target).

From Ambrosini et al. (2013).

Apparatus

Participants were given a link to the online experiment

(https://www.qualtrics.com), which was administered using Qualtrics^{XM} software

(Qualtrics, Provo, Utah). They completed the experiment remotely.

7.2.3. Procedure

This study adopted a within-subjects design. Participants signed up for a 30minute experiment. The instructions conveyed that the task comprised of eight trials and that each trial presented five separate pictures for evaluation. Therefore, it was clear that they would be shown a total of 40 pictures, presented one at a time. The instructions also conveyed that each picture showed an actor sitting in front of a small object and a large object, and that for each picture, participants had to indicate—by moving sliders from 0% to 100%—what they perceived might be the numerical chance of the actor reaching to each of the two objects. The sliders were pre-programmed such that the ratings for each sub-question for each picture evaluation would sum to 100%. In this way, there was no need for participants to calculate and ensure that their ratings for a picture summed to 100%; see Figure 24.



Figure 24 | Snapshot of one of the 40 (five scenes × eight trial types/videos) possible images, depicting sliders used to elicit participants' probability judgements (%). It is not necessary to read the small text instructions, this figure simply depicts how participants evaluated each image.

The eight trials were presented in a semi-random fashion (e.g., to avoid having the two pincer trials or the two whole-hand trials presented sequentially) and fixed across participants. Finally, all trials counterbalanced the order of questions regarding each object, that is, whether the first question in each trial asked for an evaluation of potential reach to the small object or the large object. There was no time limit for the evaluations and all participants completed the task within the 30-minute session.

7.3. Results

7.3.1. Data reduction

Given the significant and continuing stress experienced by the potential student participants over the Covid-19 period, the setup of this experiment was designed to ensure that the online task instruction was as clear as possible and that participants would find the task seamless, quick, and straightforward to complete. Indeed, all 60 participants completed the entire task; there was no participant attrition and no need for data exclusion.

7.3.2. Data analysis

Analyses were performed using SPSS (version 27), and values have been rounded to two decimal places where applicable (percentages in-text have been rounded to a whole number). Due to the increased volume of analyses conducted for the current experiment, this section is segmented according to analyses for the three different grasp types (i.e., whole-hand prehension, pincer grasp, and closed fist) and distance of target object to hand (i.e., small object near vs. far from hand and large object near vs. far from hand).

Combined analysis for whole-hand prehension

The means of participants' percentage predictions of the hand contacting either the near or far object in the whole-hand prehension sequence (averaged across items where the large object was near to hand and far from hand) are reported in Table 1. Data were analysed in a 2 (distance: near vs. far) × 5 (scene in sequence: first, second, third, fourth, fifth) repeated measures ANOVA. When Mauchly's test indicated that the assumption of sphericity was violated, the degrees of freedom were corrected using Greenhouse-Geisser estimates. The main effect of distance was significant, *F*(1,59) = 6.37, *p* = .01, η_{p^2} = .10. The main effect of scene was not significant (*p* = 1.00). There was a significant two-way distance × scene interaction, *F*(3.09, 182.25) = 68.59, *p* < .001, η_{p^2} = .54.

Scene	Distance	Mean	Standard deviation	Standard error
1	Near	61.98	13.30	1.72
	Far	38.03	13.30	1.72
2	Near	62.81	14.14	1.82
	Far	37.19	14.14	1.82
3	Near	62.42	16.35	2.11
	Far	37.58	16.35	2.11
4	Near	47.54	19.15	2.47
	Far	52.46	19.15	2.47
5	Near	35.38	16.67	2.15
	Far	64.62	16.67	2.15

Table 1 | Mean chance (with standard deviations and standard errors of means) of hand contacting near or far object for each scene in whole-hand grasp sequence.

Paired samples t-tests were performed to interpret the two-way interaction. Henceforth, Bonferroni adjustments were applied to correct for five paired comparisons (therefore *p* must be < .01; .05/5 = .01). For scene 1, participants set a higher percentage for the hand contacting the near object (62%), compared to the far object (38%; *t*(59) = 6.98, *p* < .001, Cohen's *d* = .90). For scene 2, participants set a higher percentage for the hand contacting the near object (63%), compared to the far object (37%; *t*(59) = 7.02, *p* < .001, Cohen's *d* = .91). For scene 3, participants set a higher percentage for the hand contacting the near object (62%), compared to the far object (38%; *t*(59) = 5.88, *p* < .001, Cohen's *d* = .76). For scene 4, there was no difference between the percentage set for the hand contacting the near object (48%) and the far object (52%; *t*(59) = -.99, *p* = .32, Cohen's *d* = -.13). For scene 5, participants switched and set a higher percentage for the hand contacting the far object (65%), compared to the near object (35%; *t*(59) = -.679, *p* < .001, Cohen's *d* = -.88). Next, the analyses turned to unpack whether the switch in the allocation of percentages (from initially favouring the near object to later favouring the far object) could be driven by whether the target of the whole-hand prehension (i.e., the large object) happened to be positioned near to, or far from, the hand.

Whole-hand prehension when large target near to hand

The means of participants' percentage predications of the hand contacting either the near large target object, or far small (non-target) object—when the target was positioned near to the hand—are illustrated in Figure 25. Raw data were analysed in a 2 (distance: near vs. far) × 5 (scene in sequence: first, second, third, fourth, fifth) repeated measures ANOVA. The main effect of distance was significant F(1,59) = 46.61, p < .001, $\eta_{p^2} = .44$. The main effect of scene was not significant (p = 1.00). The two-way distance × scene interaction was not significant, F(2.40,141.50) = 2.82, p = .05, $\eta_{p^2} = .05$. For the whole-hand prehension, when the large target appeared nearer to the hand, participants' bias to allocate a higher percentage to the near object, as compared to the far object (a bias that was present from the first scene in the event sequence), did not change across the five scenes presented.



Event Sequence of Whole-Hand Prehension to Near Large Target

Figure 25 | Mean percentage predictions (with 1 SEM) of hand contacting near large object for each scene in whole-hand prehension sequence. To-T4 refer to the 5 different time points/scenes corresponding to the stimulus images below the graph (i.e., To: scene 1, T1: scene 2, etc.).

Whole-hand prehension when large target far from hand

The means of participants' percentage predictions of the hand contacting either the near small (non-target) object, or far large target object—when the target (i.e., large object) appeared far from the hand—are illustrated in Figure 26. Data were analysed in a 2 (distance: near vs. far) × 5 (scene in sequence: first, second, third, fourth, fifth) repeated measures ANOVA. The main effect of distance was significant F(1,59) = 35.63, p < .001, $\eta_{p^2} = .38$. The main effect of scene was not significant (p = 1.00). The two-way distance × scene interaction was significant, F(3.25,191.77) = 180.16, p < .001, $\eta_{p^2} = .75$.



Event Sequence of Whole-Hand Prehension to Far Large Target

Figure 26 | Mean percentage predictions (with 1 SEM) of hand contacting far large object for each scene in whole-hand prehension sequence. To-T4 refer to the 5 different time points/scenes corresponding to the stimulus images below the graph.

For scene 1 (To, or time point zero), participants set a higher percentage for the hand contacting the near small object (58%), compared to the far large object (42%; t(59) = 3.65, p < .001, Cohen's d = .47). For scene 2 (T1), participants set a higher percentage for the hand contacting the near small object (59%), compared to the far large object (41%; t(59) = 4.15, p < .001, Cohen's d = .54). For scene 3 (T2), there was no difference between the percentage set for the hand contacting the near small object (50%) and the far large object (50%; t(59) = .06, p = .95, Cohen's d = .01). For scene 4 (T3), participants set a lower percentage for the hand contacting the near small object (31%), compared to the far large object (69%; t(59) = -8.05, p < .001, Cohen's d = -1.04). For scene 5 (T4), participants set a lower percentage for the hand contacting the near small object (6%), compared to the far large object (94%; t(59) = -36.46, p < .001, Cohen's d = -4.71). Overall, for the whole-hand prehension event—when the small non-target object happened to be near to the hand and the large target object happened to be far from the hand—participants' showed a clear bias from the outset: they allocated a

higher percentage of contact to the near (non-target) object, and only switched to allocate a higher percentage of contact to the far (target) object late in the event sequence (i.e., from scene 4 onwards).

Combined analysis for pincer grasp

The means of participants' percentage predictions of the hand contacting either the near or far object in the pincer grasp sequence (averaged across items where the small target was near to hand and far from hand) are reported in Table 2. Data were analysed in a 2 (distance: near vs. far) × 5 (scene in sequence: first, second, third, fourth, fifth) repeated measures ANOVA. The main effect of distance was not significant, F(1,59) = .35, p = .55, $\eta_{p^2} = .01$. The main effect of scene was not significant (p = 1.00). The two-way distance × scene interaction was significant, F(2.64,155.94) = 98.40, p < .001, $\eta_{p^2} = .63$.

Scene	Distance	Mean	Standard deviation	Standard error
1	Near	60.71	15.19	1.96
	Far	39.29	15.19	1.96
2	Near	57.58	15.75	2.03
	Far	42.42	15.75	2.03
3	Near	57.93	18.73	2.42
	Far	42.07	18.73	2.42
4	Near	44.72	18.00	2.32
	Far	55.28	18.00	2.32
5	Near	24.30	13.85	1.79
	Far	75.70	13.85	1.79

 Table 2 | Mean chance (with standard deviations and standard errors of means) of hand contacting near or far object for each scene in pincer grasp sequence.

Paired samples t-tests were performed to interpret the two-way interaction. For scene 1, participants set a higher percentage for the hand contacting the near object (61%), compared to the far object (39%; t(59) = 5.46, p < .001, Cohen's d = .71). For scene 2, participants set a higher percentage for the hand contacting the near object (58%), compared to the far object (42%; t(59) = 3.73, p < .001, Cohen's d = .48). For scene 3, participants set a higher percentage for the hand contacting the near object (58%), compared to the far object (42%; t(59) = 3.28, p = .01, Cohen's d = .42). The contrast for scene 4 did not meet the Bonferroni correction for multiple comparisons; for scene 4, there was no difference between the percentage allocated for the hand contacting the near object (45%) and the far object (55%; t(59) = -2.27, p = .03, Cohen's d = -.29). For scene 5, participants set a lower percentage for the hand contacting the near object (24%), compared to the far object (76%; t(59) = -14.37, p < -14.37.001; Cohen's d = -1.86). Next, the analyses turned to unpack whether the switch in the allocation of percentages (from initially favouring the near object to later favouring the far object) could be driven by whether the target of the pincer grasp (i.e., the small object) happened to be positioned near to, or far from, the hand.

Pincer grasp when small target near to hand

The means of participants' percentage predictions of the hand contacting either the near small target object or far large (non-target) object—when the target was positioned near to the hand—are illustrated in Figure 27. Raw data were analysed in a 2 (distance: near vs. far) × 5 (scene in sequence: first, second, third, fourth, fifth) repeated measures ANOVA. The main effect of distance was significant F(1,59) = 11.38, p <.001, $\eta_p^2 = .16$. The main effect of scene was not significant (p = 1.00). The two-way distance × scene interaction was significant, F(2.65,156.60) = 22.25, p < .001, $\eta_p^2 = .27$.



Event Sequence of Pincer Grasp to Near Small Target

Figure 27 | Mean percentage predictions (with 1 SEM) of hand contacting near small object for each scene in pincer grasp sequence. To-T4 refer to the 5 different time points/scenes corresponding to the stimulus images below the graph.

For scene 1, participants set a higher percentage for the hand contacting the near small object (60%), compared to the far large object (40%; t(59) = 4.07, p < .001, Cohen's d = .53). For scene 2, participants set a higher percentage for the hand contacting the near small object (63%), compared to the far large object (37%; t(59) = 4.85, p < .001, Cohen's d = .63). For scene 3, participants set a higher percentage for the hand contacting the near object (68%), compared to the far object (32%; t(59) = 5.86, p < .001, Cohen's d = .76). The contrast for scene 4 did not meet the Bonferroni correction for multiple comparisons; for scene 4, there was no difference between the percentage set for the hand contacting the near small object (57%) and the far large object (43%; t(59) = 2.04, p = .05, Cohen's d = .26). Surprisingly, at scene 5, participants set a lower percentage for the hand contacting the near small object (39%), compared to far large object (61%; t(59) = -3.15, p = .01, Cohen's d = -.41). It seemed that participants found the Ambrosini et al. (2013) stimuli ambiguous, even when the

scene (scene 5) showed the hand displaying a pincer grasp just before contact; it appeared as if participants' bias towards a near object was replaced by a bias for a large object.

Pincer grasp when small target far from hand

The means of participants' percentage predictions of the hand contacting either the near large (non-target) object or far small target object—when the target was positioned far to the hand—are illustrated in Figure 28. Raw data were analysed in a 2 (distance: near vs. far) × 5 (scene in sequence: first, second, third, fourth, fifth) repeated measures ANOVA. The main effect of distance was significant F(1,59) = 22.58, p < .001, $\eta_{p^2} = .28$. The main effect of scene was not significant (p = 1.00). The two-way distance × scene interaction was significant, F(2.93,173.09) = 161.74, p < .001, $\eta_{p^2} = .73$.



Event Sequence of Pincer Grasp to Far Small Target

Figure 28 | Mean percentage predictions (with 1 SEM) of hand contacting far small object for each scene in pincer grasp sequence. To-T4 refer to the 5 different time points/scenes corresponding to the stimulus images below the graph.

For scene 1, participants set a higher percentage for the hand contacting the near large object (62%), compared to the far small object (38%; t(59) = 5.44, p < .001, Cohen's d = .70). For scene 2, there was no difference between the percentage set for the hand contacting the near large object (53%) and the far small object (47%; t(59) = 1.00, p = .32, Cohen's d = .13). For scene 3, there was also no difference between the percentage set for the hand contacting the near large object (52%; t(59) = -.59, p = .55, Cohen's d = -.08). For scene 4, participants set a lower percentage to the near large object (33%), compared to the far small object (67%; t(59) = -5.87, p < .001, Cohen's d = -.76). Similarly, for scene 5, participants set a lower percentage to the near large object (9%), compared to the far small object (91%; t(59) = -30.83, p < .001, Cohen's d = -3.98).

Combined analysis for closed fist when target object large

The means of participants' percentage predictions of the hand contacting either the near or far object in the closed fist control sequence (averaged across items where the large target was near to hand and far from hand) are reported in Table 3.

Data were analysed in a 2 (distance: near vs. far) × 5 (scene in sequence: first, second, third, fourth, fifth) repeated measures ANOVA. The main effect of distance was significant, F(1,59) = 23.29, p < .001, $\eta_{p^2} = .28$. The main effect of scene was not significant (p = 1.00). The two-way distance × scene interaction was not significant, F(2.98,175.80) = 2.62, p = .05, $\eta_{p^2} = .04$. Given the non-significant distance × scene interaction, there were no follow-up analyses into whether the interaction may be further distinguished by whether the large target object was near to, or far from, the hand. Overall, participants allocated a higher percentage to the hand contacting a near object (M = 56.54) than compared to a far object (M = 43.46; mean difference = 13.08; t(59) = 4.83, p < .001, Cohen's d = .62).

Scene	Distance	Mean	Standard deviation	Standard error
1	Near	59.84	12.58	1.62
	Far	40.16	12.58	1.62
2	Near	58.31	14.65	1.89
	Far	41.69	14.65	1.89
3	Near	56.13	14.75	1.90
	Far	43.87	14.75	1.90
4	Near	54.09	16.62	2.15
	Far	45.91	16.62	2.15
5	Near	54.31	16.12	2.08
	Far	45.69	16.12	2.08

Table 3 | Mean chance (with standard deviations and standard errors of means) of hand contacting near or far target for each scene in closed fist sequence (when target was the large object).

Closed fist when large target near to hand

The means of participants' percentage predictions of the hand contacting either the near far object or far small object—when the large target object was positioned near to the hand—are illustrated in Figure 29. Raw data were analysed in a 2 (distance: near vs. far) × 5 (scene in sequence: first, second, third, fourth, fifth) repeated measures ANOVA. The main effect of distance was significant F(1,59) = 95.20, p < .001, $\eta_p^2 = .62$. The main effect of scene was not significant (p = 1.00). The two-way distance × scene interaction was significant, F(2.28,134.78) = 34.21, p < .001, $\eta_p^2 = .37$.



Event Sequence of Closed Fist to Near Large Target

Figure 29 | Mean percentage predictions (with 1 SEM) of hand contacting near large object for each scene in closed fist sequence. To-T4 refer to the 5 different time points/scenes corresponding to the stimulus images below the graph.

For scene 1, participants set a higher percentage for the hand contacting the near large object (61%), compared to the far small object (39%; t(59) = 5.27, p < .001, Cohen's d = .68). For scene 2, participants set a higher percentage for the hand contacting the near large object (63%), compared to the far small object (37%; t(59) = 6.53, p < .001, Cohen's d = .84). For scene 3, participants set a higher percentage for the hand contacting the near large object (64%), compared to the far small object (36%; t(59) = 6.08, p < .001, Cohen's d = .79). For scene 4, participants set a higher percentage for the hand contacting the near large object (65%), compared to the far small object (35%; t(59) = 5.63, p < .001, Cohen's d = .73). For scene 5, participants returned to setting a higher percentage for the hand contacting the hand contacting the near large object (12%; t(59) = 11.94, p < .001, Cohen's d = 1.54).

Closed fist when large target far from hand

The means of participants' percentage predictions of the hand contacting either the far large object or near small object—when the large target object was positioned far from the hand—are illustrated in Figure 30. Raw data were analysed in a 2 (distance: near vs. far) x 5 (scene in sequence: first, second, third, fourth, fifth) repeated measures ANOVA. The main effect of distance was significant F(1,59) = 7.38, p = .01, $\eta_{p^2} = .11$. The main effect of scene was not significant (p = 1.00). The two-way distance × scene interaction was significant, F(3.31,195.25) = 34.84, p < .001, $\eta_{p^2} = .37$.



Event Sequence of Closed Fist to Far Large Target

Figure 30 | Mean percentage predictions (with 1 SEM) of hand contacting far large object for each scene in closed fist sequence. To-T4 refer to the 5 different time points/scenes corresponding to the stimulus images below the graph.

For scene 1, participants set a higher percentage for the hand contacting the near small object (59%), compared to the far large object (41%; t(59) = 3.28, p = .01, Cohen's d = .42). For scenes 2 and 3, there was no difference between the percentage set for the hand contacting the near small object (scene 2: 53% and scene 3: 48%) and the far large object (scene 2: 47% and scene 3: 52%; scene 2 t(59) = 1.08, p = .29, Cohen's d = .14; and scene 3, t(59) = -.61, p = .54, Cohen's d = .08). For scenes 4 and 5, participants set a lower percentage for the hand contacting the near small object (43% and 21%,

respectively), compared to the far large object (57% and 79%, respectively; t(59) = -2.24, p = .03, Cohen's d = -.29 and t(59) = -9.94, p < .001, Cohen's d = -1.28, respectively).

Combined analysis for closed fist when target object small

The means of participants' percentage predictions of the hand contacting either the near or far object in the closed fist control sequence (averaged across items where the small target was near to hand and far from hand) are reported in Table 4. Data were analysed in a 2 (distance: near vs. far) × 5 (scene in sequence: first, second, third, fourth, fifth) repeated measures ANOVA. The main effect of distance was not significant, $F(1,59) = .01, p = .93, \eta_{p^2} = .00$. The main effect of scene was not significant (p = 1.00). However, the two-way distance × scene interaction was significant, F(2.89,170.23) =115.89, $p < .001, \eta_{p^2} = .66$.

Table 4 | Mean chance (with standard deviations and standard errors of means) of hand contacting near or far target for each scene in closed fist sequence (when target was the small object).

Scene	Distance	Mean	Standard deviation	Standard error
1	Near	59.20	13.05	1.68
	Far	40.80	13.05	1.68
2	Near	59.68	12.54	1.62
	Far	40.33	12.54	1.62
3	Near	57.92	14.45	1.87
	Far	42.08	14.45	1.87
4	Near	39.90	14.64	1.89
	Far	60.10	14.64	1.89
5	Near	32.69	11.98	1.55
	Far	67.31	11.98	1.55

For scene 1, participants set a higher percentage for the hand contacting the near small object (59%), compared to the far large object (41%; t(59) = 5.46, p < .001, Cohen's d = .71). For scene 2, participants set a higher percentage for the hand contacting the near small object (60%), compared to the far object (40%; t(59) = 5.98, p < .001, Cohen's d = .77). For scene 3, participants set a higher percentage for the hand contacting the near small object (58%), compared to the far large object (42%; t(59) = 4.25, p < .001, Cohen's d = .55). For scene 4, participants set a lower percentage for the hand contacting the near small object (40%), compared to the far large object (60%; t(59) = -5.35, p < .001; Cohen's d = -.69). For scene 5, participants set a lower percentage for the hand contacting the near small object (67%; t(59) = -11.19, p < .001, Cohen's d = -1.45). Given the significant distance × scene interaction, follow-up analyses were carried out to examine whether the interaction may be modulated by whether the small target object was close to, or far from, the hand.

Closed fist when small target object near to hand

The means of participants' percentage predictions of the hand contacting either the near small object or far large object—when the small target object was positioned near to the hand—are illustrated in Figure 31. Raw data were analysed in a 2 (distance: near vs. far) × 5 (scene in sequence: first, second, third, fourth, fifth) repeated measures ANOVA. The main effect of distance was significant F(1,59) = 8.81, p = .01, $\eta_{p^2} = .13$. The main effect of scene was not significant (p = 1.00). The two-way distance × scene interaction was significant, F(2.28,134.26) = 5.64, p = .01, $\eta_{p^2} = .09$.



Event Sequence of Closed Fist to Near Small Target

Figure 31 | Mean percentage predictions (with 1 SEM) of hand contacting near small object for each scene in closed fist sequence. To-T4 refer to the 5 different time points/scenes corresponding to the stimulus images below the graph.

For scene 1, participants set a higher percentage for the hand contacting the near small object (57%), compared to the far large object (43%; t(59) = 4.18, p < .001, Cohen's d = .54). For scene 2, participants set a higher percentage for the hand contacting the near small object (56%), compared to the far large object (44%; t(59) = 3.03, p < .001, Cohen's d = .39). For scenes 3 and 4, there was no difference between the percentage set for the hand contacting the near small object (scene 3: 52% and scene 4: 51%) and the far large object (scene 3: 48% and scene 4: 49%; scene 3 t(59) = .71, p = .48, Cohen's d = .09; and scene 4, t(59) = .62, p = .54, Cohen's d = .08). For scene 5, participants returned to setting a higher percentage for the hand contacting the near small object (60%), compared to the far large object (40%; t(59) = 3.43, p < .001, Cohen's d = .44).

Closed fist when small target object far from hand

The means of participants' percentage predictions of the hand contacting either the near large object or far small object—when the small target object was positioned far from the hand—are illustrated in Figure 32. Raw data were analysed in a 2 (distance: near vs. far) x 5 (scene in sequence: first, second, third, fourth, fifth) repeated measures ANOVA. The main effect of distance was significant F(1,59) = 12.11, p < .001, $\eta_p^2 = .17$. The main effect of scene was not significant (p = 1.00). The two-way distance × scene interaction was significant, F(3.22,189.78) = 244.25, p < .001, $\eta_p^2 = .81$.



Event Sequence of Closed Fist to Far Small Target



For scene 1, participants set a higher percentage for the hand contacting the near large object (61%), compared to the far small object (39%; t(59) = 5.26, p < .001, Cohen's d = .68). For scene 2, participants set a higher percentage for the hand contacting the near large object (64%), compared to the far small object (36%; t(59) = 5.26).

6.17, p < .001, Cohen's d = .80). For scene 3, participants continued to set a higher percentage for the hand contacting the near large object (64%), compared to the far small object (36%; t(59) = 5.80, p < .001, Cohen's d = .75). For scenes 4 and 5, participants set a lower percentage for the hand contacting the near large object (28% and 5%, respectively), compared to the far small object (72% and 95%, respectively; t(59) = -7.79, p < .001, Cohen's d = -1.01 and t(59) = -41.00, p < .001, Cohen's d = -5.29, respectively).

7.4. Discussion

To recapitulate, analyses revealed that for whole-hand prehension, when the large target was near the hand (see Figure 25), participants were consistently biased to judge the near large object as being the target of the agent's reach. This alone may seem to support a motor matching account; however, for whole-hand prehension when the large target was far from the hand (see Figure 26), participants started off with a bias to judge the near small (non-target) object as being the target of the agent's reach, and only later in the sequence did this switch in favour of the actual (far large) target object. While this is not consistent with motor matching, it supports this experiment's initial hypothesis.

For pincer grasp, when the small target was near the hand (see Figure 27), participants started off with a bias to judge the near small object as being the target of the agent's reach, but for some reason this switched in the last scene, in favour of the far large (non-target) object. It seems this stimulus configuration was particularly ambiguous for observers. One possibility is that when distance is held constant (i.e., when the target of the pincer grasp is close to the hand), participants increasingly notice motor cues (the motor profile of the grasp itself) and then consider other cues (such as the forward trajectory of the arm), which may lead to conflict or ambiguity in determining the goal of the reach-to-grasp action. For pincer grasp, when the small target was far from the hand (see Figure 28), participants started off with a bias to judge the near large (non-target) object as being the target of the agent's reach, and only later in the sequence did this switch in favour of the actual (far small) target object. Which cues (e.g., grasp type, distance, trajectory) are considered, as well as when these cues are considered during participants' action observation requires further attention; this theme will be picked up again in the final chapter (Chapter 8).

For closed fist, when the large target was near the hand (see Figure 29), participants were consistently biased to judge the near large object as being the target of the agent's reach, despite the lack of a pre-shaped grasp. Recall that the closed fist trials were meant to act as a neutral control hand gesture. For closed fist, when the large target was far from the hand (see Figure 30), participants started off with a bias to judge the near small (non-target) object as being the target of the agent's reach, and only later in the sequence did this switch in favour of the actual (far large) target object. If the closed fist really was a neutral control, participants should have predicted equal (50%) probability to both objects across all scenes. Clearly, this was not the case.

Lastly, for closed fist, when the small target was near the hand (see Figure 31), participants tended to judge the near small object as being the target of the agent's reach. For closed fist, when the small target was far from the hand (see Figure 32), participants started off with a bias to judge the near large (non-target) object as being the target of the agent's reach, and only later in the sequence did this switch in favour of the actual (far small) target object.

These numerous analyses tell a similar story, and confirm the hypothesis that distance (the object's position as near or far from the hand) plays a far larger role than the motor cue of grasp type, in participants' judgements of which object is deemed the goal of the reaching movement. If the stimuli themselves were not biased in any way, then participants should predict—at least at the very beginning of the sequence, before the hand's grasp has taken shape—that both objects are equally likely targets (thus assigning 50% to the small and large object). Nevertheless, it is clear that participants begin their observation of the stimuli with a bias towards the near object being the target. Since the distance effect has proven itself to be rather robust, Ambrosini et al.'s (2013) failure to take this factor into account undermines their interpretation of their findings as being in line with the direct matching hypothesis. Another key take-away

from this experiment is the confirmation that the closed fist does not act as an accurate control. This is discussed in depth in Chapter 8.

7.5. Conclusion

The final experiment of this thesis, an explicit investigation into Ambrosini et al.'s (2013) stimuli, found that they are not fit for purpose. That is, the stimuli elicit biased judgements from observers regarding which object the agent is judged as being likely to contact. Seeing as the position of the (small or large) object from the reaching hand has a definite effect on observers' predictions, does this spell the end of the action observation paradigm? The next chapter will discuss the implications of all three experiments and suggest possible solutions that might just save the paradigm.

Chapter 8.

8.1. Introduction

The current chapter is a general discussion. The results of this thesis' three experiments will be discussed, as well as explanations for the pattern of findings. After this, methodological and theoretical limitations are explored, and possible solutions to these limitations are addressed, in the form of suggestions for research avenues. Then, the distinction between implicit and explicit is discussed, followed by an assessment of the future of motor matching based on the current state of the evidence. Finally, a summary and conclusion of the thesis as a whole is offered.

8.2. Big picture

This thesis began with the infamous story of false belief, and how it took over the field of mindreading at large, leaving other-particularly low-level-types of mindreading in its empirical dust. However, with the realisation of this myopic view of the field, scholars have had to take a step back and theorise more about goals and actions, instead of leaping ahead to exciting, but complex, mental states like false belief. The replication crisis also reared its ugly head, and many mindreading paradigms and effects were left wanting. Apperly and Butterfill's (2009) two-systems account offered an enticing answer to some paradoxical findings. Alongside this, the recent view that motor representations may underpin human mental-state-tracking (e.g., Butterfill, 2020; Zani et al., 2020) is a novel theoretical advance, because motor processes may be cognitively efficient in the sense required for minimal mindreading (a core principle of the two-systems account). However, the motor theory for *goal*-tracking had to be investigated more thoroughly before belief-tracking could be theorised to have a motoric basis. This thesis therefore aimed to test and validate a classic action observation paradigm. In order to do so, three experiments were conducted to investigate action observation for goal-directed movements.

8.2.1. Results refresher

Let us remind ourselves of the results across all three experiments of this thesis. Experiment 1 was a direct replication of the Ambrosini et al. (2013) study that investigated goal-tracking using the action observation paradigm. Adults' eye gaze was tracked as they viewed reach-to-grasp movements. Contrary to standard predictions, results did not replicate gaze proactivity according to a motor matching account. Rather, adults' eyes latched on to targets that were larger and/or nearer to the agent's hand. A motor matching account would have predicted that grasp type (pincer grasp or wholehand prehension) should cue gaze proactivity to the congruent target (i.e., small object or large object, respectively). The results of Experiment 1 were consistent with another failed replication of Ambrosini et al. (2011) by Vannuscorps and Carmazza (2017), in which less than half of their adult participants showed proactive gaze in the pre-shape trials and, further, there was a non-significant difference of less than 20 ms between proactivity in the pre-shape trials and proactivity in the no-shape trials.

In Experiment 2, adults' reaction times were measured as they viewed the same stimuli, but presented in a static, rather than dynamic, format. Similarly to Experiment 1, adults response times were faster to detect a target when it appeared over an object closer to the reaching hand, rather than an object farther away from the hand. Again, this was at odds with a motor matching account.

Finally, in Experiment 3, adults' explicit probability judgements were solicited for still frames taken from the original video stimuli. Yet again, a distance effect prevailed, whereby adults explicitly predicted that the agent's hand would contact the closer of two objects, even when the hand was at rest (in the absence of any motor cue). In the pincer trial of this experiment, when the small object was located near the hand, participants appeared to change their mind in the last instance (i.e., that the hand would be more likely to contact the far large object instead of the near small one). This shows that trajectory of the reaching arm may also be an important factor or cue which observers use to predict action outcomes. One hypothesis leading on from this is that different strategies can be utilised in action observation, such as distance or motor cues, then subsequently trajectory cues as well. This is discussed in section 8.5.2.

Overall, these results imply that adults are applying non-motor heuristics during action observation. As such, the stimuli tested in these classical experiments may not be fit for purpose, that is, they do not reliably show a motor matching effect during action observation in adults.

8.3. Triangulation

Experiment 1 measured gaze proactivity in the action observation paradigm, Experiment 2 measured reaction time by adapting an attentional grasp cueing paradigm, and Experiment 3 measured probability predictions/judgements in an explicit evaluation task. These three distinct approaches, all using the same stimuli in some form, can be thought of as three points of a triangle (see Figure 33 for a visual representation of the current thesis' experiments). All three approaches appear to point to the same underlying features, showing converging evidence that the Ambrosini et al. (2013) stimuli do not reliably replicate the original action observation paradigm's motor matching effect. The findings and approaches from this thesis fit with the process of triangulation (e.g., see Bekhet & Zauszniewski, 2012; Ma & Norwich, 2007), which helps to understand phenomena via the use of multiple methods. The resulting convergence or divergence of results gives a richer understanding than simply using one method or measure.



Figure 33 | Visual representation of the current thesis' use of triangulating multiple methods (across Experiments 1-3). A mix of implicit and explicit measures were used.

The usefulness of a more holistic approach, such as triangulation, harks back to the importance of replication (see section 1.6. and 4.7.). As mentioned, due to the current replication crisis in Psychology, it is important that non-replications of studies are also recognised for their utility in helping us understand underlying mechanisms, such as the cognitive processes that underpin mindreading. Replication attempts of experiments are good scientific practice. In the case of the current thesis, direct and conceptual replications helped to identify deep-seated problems with certain classical action observation stimuli, that will require closer reflection and discussion.

8.4. Lessons from this thesis about (non-)functionality of action observation stimuli

Experiment 1 and 2, albeit non-replications, may have turned out to identify a procedural salvo to salvage theoretical expectations that depend on the validity of the action observation paradigm. A key finding from the current thesis' experiments was

that there was no clear differentiation for participants between the no-shape control trials and the pre-shape experimental trials. All experiments of the current thesis took inspiration from studies by Ambrosini et al. (2011; 2012; 2013) and Fischer et al. (2008), all of which had a closed resting fist as the control hand position. There is, however, reason to believe that such a control condition is a problem for the action observation paradigm. It may not be the experimental pre-shape trials that were problematic; rather, it may be that the problem strikes at the heart of Ambrosini et al.'s control trials themselves, where the actor contacts one of the objects with a closed fist.

Recall, also, that even participants in Ambrosini et al.'s (2011) study looked first to the larger object, when they could not use the motor cues of the hand to predict the goal. As the authors themselves stated: "the situation is different in the no-shape condition where participants proactively gazed towards the large target object *only* [emphasis added]" (Ambrosini et al., 2011, p. 1441). However, in the Ambrosini et al. (2013) study, adults in the no-shape condition proactively gazed towards the large target and small target object. Consequently, Ambrosini et al.'s own work, post-2011, indicates that adults also show the presence of proactive gazing in the control trials themselves.

On a more theoretical note, once an appropriate control condition is established (as per section 8.4.), a conceptual task extension to this could involve participants completing the task with their hands constrained (as per Ambrosini et al., 2012). Their hands could be temporarily constrained in order to test a direct matching hypothesis, which would predict impaired gaze proactivity under physical constraint only. Several possibilities are unpacked below to conjecture as to why the no-shape closed fist gesture might have introduced noise that obscured differences in gaze proactivity (between the control and experimental trials).

8.4.1. Overlap in muscle use

While neurons in the primary motor cortex fire during a pincer grasp movement and will not fire during the closed fist movement, the two tasks still overlap in terms of muscle use (Muir & Lemon, 1983). That is to say, even though—visually—a pre-shaped grasp and a closed fist may not seem similar, on a muscular level the two gestures are less mutually exclusive when considering they both contract many of the same handand finger-muscles. Further, although the outcomes of pincer grasps and whole-hand prehensions are to grasp differently-sized objects, Figure 34 depicts how both grasp types (and even the closed fist) nonetheless share the similar 'C' curve of the fingers. There are also suggestions in the literature indicating that—depending on context and the nature of the target objects—the grip aperture of a hand performing a pincer grasp towards a small object, that is flanked by a large object, can be still sufficient for the large target to be grasped (e.g., Sartori, Camperio-Ciani, Bulgheroni, & Castiello, 2014).



Figure 34 | Visual depiction (from left to right) of closed fist, pincer grasp, and whole-hand prehension.

8.4.2. Iconicity

A closed fist gesture may not be straight-forwardly non-goal-directed, nonpurposeful, or non-informative as Ambrosini et al. assume. There is evidence showing that adults perceive a closed fist and pincer grasp as having similar affordances for object interactions (e.g., Perilli, Barrada, & Maiche, 2013). Adults may also perceive the movement of a closed fist as affording the contacting or displacement of an object (e.g., Perilli et al., 2013), and even perceive congruency between a clenched fist and objects such as an 'egg', a 'hairbrush', and a 'telephone', so much so that the gesture can facilitate object recognition (e.g., Cooperrider & Goldin-Meadow, 2017; Pine, Reeves, Howlett, & Fletcher, 2013). The closed fist has what gesture researchers call *iconicity*: gestures in which the hand shapes can be taken to represent visuospatial features of the world. Whilst a closed fist shares associative visuospatial properties with the large ball (as found in Experiments 1 and 2), gesture researchers have also commented on how there are contexts where fist formation can be taken to represent a small round shape (Cooperrider & Goldin-Meadow, 2017). Therefore, if a closed fist can indicate goal-directedness the way a grasping hand does, then the closed fist might also facilitate tracking. There are also low-level perceptual associations at work, as the closed fist matches the large object in terms of size.

8.4.3. Familiarity

Another relevant criticism of the closed fist control condition is the fact that such a hand gesture is unfamiliar with respect to reach-to-grasp goals, whereas pincer and whole-hand grasps are much more naturalistic in the context of actually picking up objects. Even if Experiment 1 *had* replicated Ambrosini et al.'s (2013) effects, there would still be significant problems in interpreting why there should be less proactivity in the no-shape condition (closed fist), than in the pre-shape condition. Instead of Ambrosini et al.'s effects showcasing motor matching, their results could still be confounded with differences in attentional engagement (e.g., Southgate, 2013). It is possible that, by virtue of the relative novelty of the no-shape trials, it may be more challenging for adults to disengage from—and thus more challenging to generate proactive saccades (eye movements) away from—the closed fist.

A possible test would be to occlude a portion of the movement in both conditions. If Southgate's attentional engagement possibility has merits, and if the current thesis findings are reliable, one prediction is that occluding a portion of the trajectory in the no-shape and pre-shape trials would show proactivity in *both* conditions (supporting low-level attention to cues or heuristics). And if proactive gazing genuinely results from access to corresponding motor representations, then occlusion of the movement trajectory should have no effect, thereby preserving Ambrosini et al.'s (2013) findings.

8.4.4. Spatial layout of objects

In the Ambrosini et al. (2013) stimuli used across the experiments of this thesis, the physical distance between the two objects is relatively close (this also applies to all of Ambrosini et al.'s original 2011 stimuli), and target position information is important for the control of goal-directed arm movements (Sarlegna et al., 2003). This makes it possible for observers to respond reactively by shifting their gaze to some mid-point between objects, then monitoring the trajectory and likely contact of the hand, and only then making adjustments to the gaze as necessary. Due to this, it is not clear the extent to which participants' gaze is predictive/anticipatory (as opposed to reactive), or what the benefits—of taking advantage of grip aperture cues for target prediction—are for the observers.

Moreover, as seen from still frames extracted from an example video from the Ambrosini et al. (2011) study (Figure 35), the objects are both positioned to one side of the screen and placed very close to each other. Because of this, participants' gaze in half of the pre-shaped trials must travel from the large to the small object (if the target is the small object) and must travel from the small to the large object (if the target is the large object). This potentially contaminates analysis of gaze proactivity. To remedy this, one could use a layout more common in anticipatory looking tasks used by mindreading researchers, in which the two target locations are distinctly separate (to avoid confusion caused by gaze inevitably passing over one object to get to the other; Paulus, Schuwerk, Sodian, & Ganglmayer, 2017). Doing so would help determine the robustness of the paradigm, and potentially allow the paradigm to be expanded upon in the future to test other factors relevant to mindreading, such as belief-tracking.


Figure 35 | Video excerpts, displaying a hand reaching towards two objects (a small tomato and a large tomato). The blue line in the central image indicates an example of proactive gaze moving ahead to the objects before the hand has reached its target.

From Corrado Sinigaglia (https://www.youtube.com/watch?v=Vjoid5tVW64).

8.5. Current experiments: problems and solutions

There are, as to be expected, limitations with the current research experiments found in this thesis. These limitations are discussed and solutions for future experiments are offered.

8.5.1. Engagement and motivation

The format through which the studies of the current thesis were shown to participants is worth noting. Participants viewed computer-based stimuli (Experiment 1: dynamic video clips; Experiments 2 and 3: static images that, when viewed together consecutively, gave the illusion of apparent motion). Such presentation of stimuli may not in fact elicit naturalistic behaviour (Laidlaw, Foulsham, Kuhn, Kingstone, & Purves, 2011), due to video clips not being as engaging or interactive as real-life scenarios. Kulke, Wübker, and Rakoczy (2019) observe that when infant tasks are used with adults or even older children, the unsuitability of the stimuli may lead to a lack of motivation in following along with the event sequence. Although the action observation paradigm was initially used with adults, this criticism still holds; having more engaging videos in eye gaze studies may increase motivation of non-infant participants.

To this end, there is a case for using novel, unfamiliar objects that participants have never come into contact with visually or physically, as opposed to typical, everyday objects that participants are accustomed to seeing and grasping often. Chua, Bub, Masson, and Gauthier (2018) found that novel objects evoke representations of grasping in a relatively automatic manner. Pictorial representations or actual 3-dimensional novel objects can also be tailored to a particular grasp type, controlled for object size, and even tested for saliency. According to Handy, Tipper, Borg, Grafton, and Gazzaniga (2006) and van Elk, Viswanathan, van Schie, Bekkering, and Grafton (2012), unfamiliar objects can even yield greater motor cortex activation than familiar objects, due to the increase in analytic processing that arises from the object's novelty.

That said, it is difficult to control for differences in movement across all trials in the case of real-life social scenarios. Virtual and augmented reality have the potential to bridge this gap, by immersing people in different realities that also allow for carefully controlled stimuli. Future research could compare stimuli formats such as video presentation, live action, and virtual/augmented reality, to distinguish whether videobased stimuli can indeed stand in as a proxy for live scenarios, or whether virtual/augmented reality is a viable next step.

8.5.2. Assuming homogeneity in behaviour and cue processing

The current experiments assumed homogeneity across all participants, with regards to how they engaged in action observation. However, observers might not all track objects and goal-directed movements in the same way. If it were the case that there exist inter-individual differences in the strategies used during action observation, much larger sample sizes would be required in order to carry out cluster analyses and investigate these separate groups, or profiles. To make this more feasible, it may be necessary to commission large-scale, multi-laboratory, collaborative studies to pool resources. Pre-registering is also a good way forward, to ensure that even null findings be given their righteous place as published material to contribute to the current state of certain paradigms and theoretical stances.

A critical finding from this thesis was that there are other cues which participants relied on during action observation. Experiments 1 to 3 revealed that distance of target to hand and size of target were cues (or strategies) that underpinned goal-tracking. Experiment 3 even revealed that there were instances—when distance was held constant (i.e., the goal target was closer to the hand)—whereby participants would first consider grasp cues and then consider arm trajectory cues. For the purposes of the current thesis, it could be conjectured that there are different strategies adults use when attempting to make sense of others' actions, including cues such as distance, arm trajectory, object size or salience, and so on; perhaps motor pre-shaping cues are simply one of these strategies—or waves—used in conjunction with the numerous other cues available during action observation. This idea is reminiscent of Siegler's (2016) overlapping waves model of cognitive development, in which different processing strategies wax and wane in how frequently they are used across time. Figure 36 illustrates this nicely.



Time (as reach-to-grasp action unfolds)

Figure 36 | Exemplar of diverse strategy use across development, taking inspiration from Siegler's (2016) overlapping waves model. The frequency of each strategy's use (%) varies as an action unfolds across time.

Further to the existence of multiple strategies over which cue or coalition of cues are focused upon, different individuals may use each strategy more or less frequently than others, or in a different order. For instance, some may scan available and graspable objects first, before focusing on the hand to potentially make use of motor cues such as the hand pre-shaping. Others may focus on a reaching hand until another cue, such as the arm's trajectory or the distance to an object, becomes more salient. If such a model really does represent strategies underlying action understanding, then treating a sample as a homogenous group does not allow for the nuances of these distinct profiles to shine through (i.e., the distinct waves get lost and become noise, if the research aim is to find one wave when there are, in fact, several).

In a similar vein to the previous point, the (overly) strict exclusion criteria of some gaze proactivity studies (e.g., Ambrosini et al., 2011, 2013) may end up wiping out the diversity that underpins natural looking behaviour. If individuals employ multiple gaze strategies, then overly strict criteria run the risk of relegating potentially hidden clusters/profiles as noise or outliers. That said, if we allow naturalistic behaviour to come through in laboratory settings, it does make it difficult to balance this with controlling data in order to make experimental claims.

8.5.3. What about from *your* perspective?

Another limitation is that the motor account cannot be merely dismissed, because the current experiments' stimuli were not from a first-person viewpoint; rather, the stimuli showed the agent from a side-on, third-person view. Additionally, regarding Butterfill's (2020)'s theoretical confidence that Ambrosini et al.'s basic paradigm taps a core underlying phenomenon (i.e., motor-processing for tracking mental-state-related actions), the task must be re-designed to control for the following confounding variable: observers may be relying on arm trajectory, instead of grip aperture, for tracking goalrelated actions. As Dushanova and Donoghue (2010) report, neuronal activity in the primary motor cortex helps to predict both direction and trajectory of observed actions as they unfold. The Ambrosini et al. (2011; 2012; 2013) stimuli show the reaching hand emerging horizontally from the side of the frame (side counterbalanced; see previous Figures 11 and 14). It could be argued that seeing actions in the first-person would facilitate direct motor matching the most, because of the similarity to one's own perspective.

One such improvement to this, is that the actor be shown reaching from the bottom of the frame (i.e., from a bird's eye view directly above hand and objects), in a

way that maximises the potential for direct matching. When viewers watch such stimuli, it appears as if they themselves might be reaching for a target (see also Fischer et al., 2008). An expansion of this would be to have two conditions: a *self* versus *other* perspective, whereby the *self* condition shows grasping from a first-person perspective, and the *other* condition shows the same grasp rotated 180° (as if the agent was sitting opposite the observer). To further tease apart whether the orientation of the reaching agent effects direct motor matching, one could also add conditions that show grasping from the left- and right-hand-side (i.e., 90° to the left and to the right of the *self* condition). Figure 37 weaves these various suggestions together (i.e., bird's eye view, occluder to hide trajectory cue, and different angles of observation), to depict an improved stimulus layout for future research.



Figure 37 | Visual depiction of different action observation layouts from bird's eye view, with distinctly separate target object locations and occluders to visually block motor cue of the hand before the outcome/contact phase. Four rotated perspectives are shown, which vary in respect to how closely they map on to the observer.

Interestingly, Fischer and Szymkowiak (2004) found that the viewpoint or perspective of the hand (they showed both a *self*-perspective, similar to a first-person view, and an *other*-perspective, rotated by 180°) was not related to the facilitation of encoding visual information from hand gestures. If action observation does indeed involve direct matching, it would make sense for a variation of different hand angles and hand orientations to evoke the same matching processes, since our everyday observations of others' motor actions occur from many visually diverse perspectives. For instance, we anticipate the actions of agents when those agents are beside us, opposite us, and so on. However, stimuli with a bird's eye view of the hand reaching towards the objects should be directly compared with the side-on view displayed in the Ambrosini et al. (2011; 2013) stimuli. If there is no difference between the angles of display, then the angle of observation would not seem to influence the direct matching process. But, if there is a difference between the angles of display, this may shed light on the optimal circumstances for direct matching to occur.

8.5.4. On using more appropriate control trials

As section 8.4. highlighted, there are multiple areas for improvement in the action observation paradigm, particularly regarding methodological concerns (namely, the closed fist not being neutral enough to be considered a suitable control hand gesture). That said, there are also many solutions that provide rich ideas for future research. An avenue for further research would be to investigate whether, and to what extent, adults take advantage of hand pre-shaping (pincer grasp and whole-hand prehension) to track goal-related outcomes, when the control trials are suitably adjusted.

Looking to other studies, a visuo-motor priming study by Borghi et al. (2005) found a similar congruency effect for hand gesture as that found by Fischer et al. (2008); Borghi et al. used an open hand gesture as their control/neutral condition, alongside the two grasp types of pincer grasp and whole-hand prehension. More specifically, the closed fist could be replaced with a non-purposive back-of-hand movement (i.e., palm facing upwards, Kanakogi & Itakura, 2011; Woodward, 1998). Additionally, control trials could show the grasping hand reaching towards one of two objects of the same size (i.e., two identical small objects or two identical large objects), in either a pincer grasp or a whole-hand prehension as a more informative control condition. This layout elegantly eliminates the need for a control hand gesture.

In terms of what might be expected from these changes, one possibility is that the proactive effect (of proactivity in the experimental trials being greater than the proactivity in the control trials) will be consistent across all the aforementioned control trial adaptations. Such consistency would offer a strong test of the view that motor-processing for tracking goals and mental-state-related action is a core phenomenon (e.g., Butterfill, 2020; Butterfill & Apperly, 2016). It is also possible that one of the control trial adaptations yields clearer evidence than the others, when it comes to the role of the motor system in action observation.

8.5.5. Weirdness strikes again

A final limitation concerns the '*weirdness*' of mindreading. Historically, the study of human mindreading and its constituent processes has been undertaken with largely western, educated, industrialised, rich, and democratic (W.E.I.R.D.) populations—as is the case in many areas of Psychology, and also Science more broadly (Muthukrishna et al., 2020). Though such samples constitute the majority of research samples, these characteristics are in fact the minority if you consider them globally (Norenzayan, Henrich, & Heine, 2010). For instance, there are cross-cultural differences in the use of gesture: in the Afro-Swedish context, the closed fist gesture can be indicative of power, defiance, anger, or solidarity, among other things (Osei-Kofi, Licona, & Chávez, 2018). Although, Cotton (2013) stipulates that open-handed gestures (such as palm with fingers together) tend to be neutral and non-offensive across many cultures and countries. Yet, there is a need for more cross-cultural research into the different ways by which non-WEIRD cultures might track and attribute others' goals in action observation paradigms; however, this was not a focus of the current thesis.

8.6. Implicit versus explicit goal-tracking

Psychological science is rife with battles; the classic nature versus nurture debate may spring to mind. Thinking back to the history of mindreading from Chapter 1, there are rich versus lean interpretations of how mindreading is acquired, and develops. More nuanced are the arguments between top-down and bottom-up processing. Of particular relevance here is the notion of implicit versus explicit goal-tracking of an agent's reachto-grasp action. In Experiments 1 and 2, we measured adults' proactive gaze and adults' reaction times over narrow stimulus-onset-asynchronies, to tap (bluntly speaking) implicit indicators of action understanding. In Experiment 3, we measured adults explicit verbal answers (probability judgements) to questions about the outcome of an agent's react-to-grasp action. There is a scarcity of work being conducted to assess adults' implicit as well as explicit tracking of others' goals; therefore, it is worthwhile to discuss whether implicit and explicit measures are separate in tapping distinct representations, or whether they may lie on a continuum.

One possibility is that the implicit and explicit *worlds*, for want of a better word, are vastly different and overlap only slightly (Apperly & Butterfill, 2009). There is room for caution with such an explanation; for instance, regarding Apperly and Butterfill's (2009) two-systems account, De Bruin and Newen (2014) cite the lack of depth into how the two systems are functionally related, as a drawback. They add that it is unclear whether there is any direct inter-system interaction. Delineating the relationship between implicit and explicit goal-tracking would require similar theoretical input into the exact nature of this relationship if it is not entirely mutually exclusive.

Another possibility is that the implicit and explicit worlds are much more alike than previously thought, and in fact overlap a great deal; some studies show that implicit and explicit measures of performance in false belief mindreading tasks are correlated (e.g., Low, 2010). Indeed, a theoretical conjecture is that the application of strategies or rules found in adults' implicit (or quick) processing are "re-represented" (Karmiloff-Smith, 1992) for supporting explicit analysis (see also Gönül & Paulus, 2021). Given these two possibilities, how could one tease apart which one is closer to the 'truth', or at least, which possibility is more parsimonious? The answer lies in the careful use of methods and measures.

One way to probe this distinction might be as follows. A future study could occlude the outcome of reach-to-grasp movements, and then measure both anticipatory eye movements as well as verbal or written answers to the question "Which object will the hand contact?". In this way—similarly to how triangulation can provide converging or diverging evidence—the comparison between eye movements and explicit answers could reveal whether the two response types tap into separate or similar underlying systems (e.g., see Paulus et al., 2017). This method also reflects the action-prediction dissociation found in children's mindreading (mentioned in section 1.5.2.), whereby young children show a dissociation between correct eye movements but incorrect verbal predictions in mindreading tasks (e.g., Clements & Perner, 1994).

Another way to probe the distinction is to test for process dissociations. For instance, participants could be instructed to complete a secondary task (e.g., a workingmemory task or a language-based task) during action observation, to see whether there is an interference effect on one or both of the response types, when executive functioning and/or language resources are depleted. This could also reveal whether measures typically thought of as implicit and explicit are relatively separate from each other (i.e., if only one response type is hindered), or linked (i.e., if both response types are hindered). Ultimately, the theoretical demarcation between implicit and explicit calls for further empirical investigation.

8.6.1. If it ain't broke, don't fix it

Critics may argue that implicit tasks and stimuli are just that—implicit—and should not be investigated in an explicit manner as this is irrelevant to the intended purpose of that task or stimulus. In response to this, Low and Edwards (2018; see also Poulin-Dubois et al., 2018) aptly point out the following: in order to properly determine what a certain task is measuring, independent validation of a task's analysis is required. One way to do so is to ask for adults' explicit judgements about the stimuli (as Experiment 3 did). Additionally, testing for convergence across tasks or measures helps, as this thesis' experiments did via the process of triangulation (section 8.3.). Lastly, if adults' explicit understandings of a set of stimuli reveal biases (such as distance or size effects), this potentially tells us that adults enter an implicit task (with this same stimuli) with these biases already at play. In other words, such perceptual biases may impact implicit understandings. If this is the case, surely it is worth noting in order to avoid misinterpretation of data, or incorrect theoretical justification of findings.

There are avenues for moving away from the historical implicit-explicit binary distinction. For instance, as mentioned briefly in section 2.5.1., Gönül and Paulus (2021) favour a theoretical framework whereby children's representations progressively explain action in more detail as they go through development. This gradual progression treats implicit and explicit understanding as more continuous, each at opposite ends of the same spectrum. Ultimately, the theoretical demarcation between implicit and explicit warrants further research. More specifically, future investigation must parse out whether implicit mindreading is a distinct phenomenon that has hitherto been difficult to detect, or whether implicit mindreading is not robust on its own, but rather much more linked to explicit mindreading (Kulke, Wübker, & Rakoczy, 2019).

8.7. Whither motor matching: what's the prognosis?

There is a re-energisation in the field of cognition to emphasise the importance of sensorimotor processes for goal understanding, and even belief understanding (Ambrosini et al., 2013; Edwards & Low, 2017; Low et al., 2020; Zani et al., 2020). Motor processes are now starting to be taken seriously again as a foundation for wide-ranging social cognition (with inevitable pushbacks, e.g., see Heyes & Catmur, 2021). The process of matching an observed movement onto our own motor representations of that action is considered as providing us with a richer sense of, for instance, the goal of others' actions from the inside (Butterfill, 2020; Rizzolatti & Sinigaglia, 2010). A significant platform for holding up the direct matching account for goal-tracking comes from proactive gaze studies that include the series of findings from Ambrosini and colleagues' laboratory, but given the unsuccessful replications, where does this leave motor matching? Is movement really a window to the mind?

While it may seem that motor matching is withering away amongst numerous methodological and theoretical issues, this thesis does not aim to dismiss the role of motor representations all together. There is still a case for motor theory within action observation and mindreading; it may just not be front and centre, especially as the hype around mirror neurons dies down (while there has been excitement around mirror neurons in the cognitive neuroscience field, some are sceptical that the hype is exaggerated, e.g., Heyes & Catmur, 2021). Motor cues are useful and used by observers of action, but it is also clear there are many other cues vying for our attention. What dictates when we use a motor versus another type of cue? On the one hand, perhaps adjusting and controlling for all the methodological issues (from section 8.4.) would result in much clearer motor matching effects, that is, perhaps the paradigm—as it currently stands—simply does not serve as a good measure of proactive gaze. On the other hand, what if—even after fixing the numerous methodological issues—motor matching still remains elusive? Then, perhaps low-level motor theory is simply not the correct angle from which to tackle action understanding.

8.7.1. Very peripersonal...

It is tempting for a supporter of motor theory to argue that the current thesis' findings—of gaze, response latencies and probability predictions being biased towards the target closest to the reaching hand—may be consistent with the primacy of motor processing; perhaps motor effects are more robust for actions towards objects in peripersonal space (i.e., within reachable distance; Costantini, Ambrosini, Scorolli, & Borghi, 2011; Higuchi, Imanaka, & Patla, 2006), as opposed to extrapersonal space. However, this thesis would suggest caution in straightforwardly falling prey to that temptation, because the peripersonal space argument is not straightforwardly applicable since both the small and large object in Ambrosini et al.'s (2013) stimuli are within peripersonal space (both objects are graspable and within reaching distance).

Of course, one could be further tempted to argue that the motor matching effect only occurs—or is stronger—for objects in very-near space (perhaps "peri-peripersonal space"), as opposed to generally within reaching distance. Future research could, of course, digitally manipulate the videos to test the argument (e.g., stretch the scene lengthways so that the set of objects are varying distances from the reach), but this thesis maintains that the argument seems rather a hollow post-hoc stance. How does one then decide at what distance objects must be for motor matching to occur, if the distinction between peripersonal and extrapersonal alone is not enough?

8.7.2. Conjecture: perceptual analysis and conceptual processing

The findings, across the three experiments in this thesis, suggest that it is possible to account for the pattern of adults' proactive eye movements without appealing to observers having to encode and resonate with motor features of the agent's hand shape. One conjecture is that the current findings suggest we can just as effectively rely on perceptual analysis of, and conceptual processes about the observed scene to determine the goal of someone else's action (Csibra, 2008; Eshuis, Coventry, & Vulchanova, 2009). Perceptual analysis can provide descriptions about the spatial arrangements of objects in the environment (e.g., which object is closer to, and which object is farther away from the agent) and their saliency (e.g., size and colour). Such perceptual analysis can provide input for conceptual or inferential processing.

Conceptual processing may then draw upon internal models that specify information about others, the cultural environment, and the statistical regularities of others' behaviours (e.g., learned rules such as: people are likely to reach for things that are closer to them). These internal models—with specifications of rules about actions or behaviours—are likely to be built up through years of constructing an understanding of how others are likely to act, reach, or gesture towards others and towards objects in the context of social interaction (Carpendale & Lewis, 2004; Ruffman et al., 2012). It is also worth reminding ourselves of Paulus et al.'s (2011) study, mentioned in section 2.6., in which statistical frequency information was found to aid action prediction in infants. Such statistical learning is an example of a low-level strategy—not classed as mindreading—because it does not involve even the attribution of a goal to the agent. Interestingly though, experiential learning can come across as both a lean or a rich process. On the one hand, previous learning can be boiled down to statistical

143

information such as frequency, and simple behaviour rules or non-motor heuristics devoid of mentalising. On the other hand, previous experience is seen to be situated within complex social systems of interaction between parents, peers, teachers, and so on. This idea will be picked up again in section 8.8.

In this manner, goal-tracking in adults (as well as in infants and children) can occur outside of the motor system. Of course, it is possible that information identified and processed outside of the motor system may be subsequently passed on to the motor system (Jacob, 2008; Southgate, 2013); this could explain how observers (as found in Experiment 3) may first expect that the agent's goal is to reach for the closest (or largest) target, and only later attend to motor cues (like the hand shaping) to maintain or adjust their prior expectations, as they witness the events unfold. In other words, adults could perceptually and conceptually make expectations of what an agent's goal might be, and then motorically simulate a way in which the agent might fulfil that goal.

In terms of cognitive decision-making, such a conjecture of how goal-tracking is accomplished may be consistent with *predictive coding* frameworks of human decisionmaking, whereby action anticipation and prediction are both experience-based and data-driven (Donnarumma, Costantini, Ambrosini, Friston, & Pezzulo, 2017; Kilner, Friston, & Frith, 2007a, 2007b; Nijssen, Pletti, Paulus, & Müller, 2021). According to such predictive coding models, there is synergy between lower-level and higher-level processes in that they are bidirectional; there is constant interplay of reciprocal topdown and bottom-up information. Incoming sensory input continuously informs action prediction, which is then compared to the actual observed event (e.g., an agent's movement). The difference, or gap, between the prediction and what was actually observed, is thought to be what drives the motor system to keep on updating its predictions for upcoming actions. This links back to Csibra's (2008) top-down action reconstruction model and criticisms of the direct matching hypothesis (in section 4.5.4.).

144

8.7.3. Support for perceptual-conceptual conjecture

The conjecture that goal-tracking can be performed by relying on perceptual and conceptual processes, which are experience-dependent, can also accommodate a range of other evidence. First, the conjecture fits with evidence that individuals born without arms, for instance, can correctly visually anticipate and verbally predict the goals of others' reaching actions (e.g., Vannuscorps & Caramazza, 2017). Second, it fits with evidence that while damage to various parts of the motor cortex can cause movement disorders, the damage does not necessarily affect one's perception and understanding of others' actions or movements (e.g., Negri et al., 2007). Third, it fits with evidence showing that infants' exhibited motor activations when they saw a human hand as well as a mechanical claw move towards the target object but not the distractor object (including even when the human hand moved in a biologically impossible manner towards the target object; e.g., Southgate, 2013; Southgate & Begus, 2013). Overall, this thesis takes the moderate view that we can anticipate and predict others' goals from a coalition of cues and using a range of strategies, and that perhaps one of these might focus on motor terms. Such a view would even align with a dynamical developmental systems perspective whereby, from infancy onwards, we only gradually master grasp-toreach actions with experience, and also by coordinating perception (focusing and stabilising attention), gesture, and motor control skills (e.g., Smith & Thelen, 1993).

8.8. Time to put the 'inter' back in inter-action

In the previous section, this thesis discussed how an individual's sensitivity to a coalition of action cues (e.g., distance and size of objects, trajectory, hand shaping) that are encoded perceptually, cognitively, and motorically, may facilitate attribution of others' goals. To get a better grip upon the foundational processes for goal anticipation and prediction, it is worthwhile to consider putting the *inter* back in inter-action, so to speak, so as not to undermine the important roles shared experience and social interaction play in how we come to navigate our worlds (Carpendale & Lewis, 2004). Scientists need to take more seriously how the meanings that we as adults attribute to others' actions towards outcomes (i.e., goals), may be rooted in our early learning and

experiences about how others' respond to our actions (Carpendale & Lewis, 2004; Mead, 1934).

Our knowledge is learned within social systems, as opposed to in isolation. Infants pick up meaning from social interaction, such as through "mini customs" or routines involving gestures and actions their parents often make towards them (Lewis, Carpendale, & Stack, 2013). To illustrate, consider an infant reaching towards two toys on a table – in response, the caregiver may hand over the closer toy and may even help the infant to adapt their fingers to the toy's physical properties. Likewise, infants are also experiencing how others respond to others reaching to grasp objects. Through such shared routines, customs, or interaction patterns (which are themselves dynamic), we come to gradually learn about how others respond to our grasp-to-reach movements. In so doing, we may also gradually learn that various actions may have different outcomes that can be anticipated, and that there might be ways of acting which require certain hand gestures or requests.

Indeed, a sensible reminder from Carpendale and Lewis (2008) is that imitatively mirroring an action does not necessarily translate to action understanding. That is, even if an action is motorically simulated in an observer, it is unclear how this motor resonance in fact helps the observer to comprehend the *meaning* of that action within the given context. There is often a myriad of social and contextual cues that give rise to correctly understanding someone's goal-directed gestures. Reach-to-grasp actions may be simpler to understand, but a coherent theory must also account for more complex and socially nuanced actions (unless different levels of action complexity are underpinned by different systems or processes).

If scientists can have a better understanding of the meanings which we take from the ways that others respond to our actions over development, this can help them to better understand how adults impose meaning onto others' grasp-to-reach movements in the kinds of action observation paradigms tested in this thesis. The system of interaction or activities—upon which the meanings about others' goals are discovered and constructed over the course of development—may even be the bridge that connects and transforms rudimentary goal-tracking to deeper reflective understandings of mental

146

states as such (like intentions). Following a long-standing plea from Lewis et al. (2013), the current thesis hopes that grounding meaning about actions that are directed to outcomes (goals) in the natural world of interaction—for developing theories, designing experiments and interpreting data—may finally help scientists transcend unproductive dichotomous debates (such as nativism vs. empiricism; sensorimotor vs. representationalism) regarding how humans gain insight into minds and actions.

8.9. Summary and Conclusion

Most of us have a fundamental curiosity about people. The field of mindreading helps satisfy this, and will keep doing so as the field moves and grows alongside new discoveries about social and cognitive developmental psychology. That said, mindreading researchers are still debating about the kinds of representations that modulate expectations of how actions will unfold. To summarise the current thesis as a whole, the action observation paradigm that was investigated could not be directly or conceptually replicated. Mainly, instead of motor hand cues triggering proactive gaze, it was found that more prosaic cues such as distance and salience were behind participants' eye movements. While there are many methodological improvements that may rectify the situation, there lies a deeper theoretical question beneath these nonreplications: is there indeed a motor basis to action and goal understanding? If not, then higher-order mindreading such as false-beliefs seems an unlikely candidate for being motorically grounded (contrary to recent research, e.g., Low et al., 2020). Ultimately, the role of motor processes in action observation and action prediction must be further unpacked. It is wise to tackle this area via different measures, such as gaze proactivity and reaction time (implicit), alongside more deliberate judgements such as likelihood predictions (explicit), to probe the underlying mechanisms via which action observation is conducted in adults.

- Ambrosini, E., Costantini, M., & Sinigaglia, C. (2011). Grasping with the eyes. *Journal of Neurophysiology*, *106*(3), 1437-42.
- Ambrosini, E., Reddy, V., de Looper, A., Costantini, M., Lopez, B., & Sinigaglia, C. (2013). Looking ahead: Anticipatory gaze and motor ability in infancy. *PLoS ONE*, *8*(7), E67916.
- Ambrosini, E., Sinigaglia, C., & Costantini, M. (2012). Tie My Hands, Tie My Eyes. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 263-266.
- Apperly, I. (2011). *Mindreaders: the cognitive basis of "theory of mind"*. Psychology Press.
- Apperly, I., & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970. https://doi.org/10.1037/a0016923
- Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8
- Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science*, *8*(6), 492–499. https://doi.org/10.1111/j.1467-7687.2005.00440.x
- Bekhet, A. K., & Zauszniewski, J. A. (2012). Methodological triangulation: an approach to understanding data. *Nurse Researcher*, *20*(2), 40–43. https://doi.org/10.7748/nr2012.11.20.2.40.c9442
- Borghi, A., Bonfiglioli, C., Lugli, L., Ricciarelli, P., Rubichi, S., & Nicoletti, R. (2005). Visual hand primes and manipulable objects. *Proceedings of the Cognitive Science Society*, 322-327.
- Buccino, G., Binkofski, F., Fink, G., Fadiga, L., Fogassi, L., Gallese, V. ... & Freund, H. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *The European Journal of Neuroscience*, *13*(2), 400– 404. https://doi.org/10.1046/j.1460-9568.2001.01385.x
- Burnside, K., Ruel, A., Azar, N., & Poulin-Dubois, D. (2018). Implicit false belief across the lifespan: Non-replication of an anticipatory looking task. *Cognitive Development*, *4*6, 4–11. https://doi.org/10.1016/j.cogdev.2017.08.006

Butterfill, S. (2020). *The developing mind: a philosophical introduction*. Routledge.

- Butterfill, S., & Apperly, I. (2016). Is goal ascription possible in minimal mindreading? *Psychological Review*, *123*(2), 228–233. https://doi.org/10.1037/rev0000022
- Butterfill, S., & Sinigaglia, C. (2012). Intention and Motor Representation in Purposive Action. *Philosophy and Phenomenological Research*, *88*(1), 119-145.

- Cannon, E., & Woodward, A. (2012). Infants generate goal-based action predictions: Infants generate goal-based predictions. *Developmental Science*, *15*(2), 292–298. https://doi.org/10.1111/j.1467-7687.2011.01127.x
- Carlson, S., & Moses, L. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*(4), 1032–1053. https://doi.org/10.1111/1467-8624.00333
- Carpendale, J., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *The Behavioral and Brain Sciences*, *27*(1), 79–96. https://doi.org/10.1017/S0140525X04000032
- Carpendale, J., & Lewis, C. (2008). Mirroring cannot account for understanding action. *The Behavioral and Brain Sciences*, *31*(1), 23–24. https://doi.org/10.1017/S0140525X07003147
- Carruthers, P. (2017). Mindreading in adults: Evaluating two-systems views. Synthese International Journal for Epistemology, Methodology and Philosophy of Science, 194(3), 673.
- Cattaneo, L., Sandrini, M., & Schwarzbach, J. (2010). State-dependent TMS reveals a hierarchical representation of observed acts in the temporal, parietal, and premotor cortices. *Cerebral Cortex*, 20(9), 2252–2258. https://doi.org/10.1093/cercor/bhp291
- Chua, K., Bub, D., Masson, M., & Gauthier, I. (2018). Grasp representations depend on knowledge and attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(2), 268–279. https://doi.org/10.1037/xlm0000453
- Clements, W., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, 9(4), 377–395. https://doi.org/10.1016/0885-2014(94)90012-4
- Cooperrider, K., & Goldin-Meadow, S. (2017). When Gesture Becomes Analogy. *Topics in Cognitive Science*, *9*(3), 719-737.
- Costantini, M., Ambrosini, E., Cardellicchio, P., & Sinigaglia, C. (2014). How your hand drives my eyes. *Social Cognitive and Affective Neuroscience*, *9*(5), 705–711. https://doi.org/10.1093/scan/nst037
- Costantini, M., Ambrosini, E., Scorolli, C., & Borghi, A. M. (2011). When objects are close to me: affordances in the peripersonal space. *Psychonomic bulletin & review*, *18*(2), 302-308.
- Cotton, G. (2013). Gestures to avoid in cross-cultural business: In other words, 'Keep your fingers to yourself!'. *The Huffington Post. Avaiable at: < http://www.huffingtonpost. com/gayle-cotton/cross-cultural-gestures_b_3437653.html>(retrieved July 7, 2017).*
- Craighero, L., Bello, A., Fadiga, L., & Rizzolatti, G. (2002). Hand action preparation influences the responses to hand pictures. *Neuropsychologia*, *40*(5), 492–502. https://doi.org/10.1016/S0028-3932(01)00134-8

- Csibra, G. (2008). Action mirroring and action understanding: an alternative account. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition*. (pp. 435-459). Oxford University Press.
- Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. Developmental science, 1(2), 255-259.
- Cutting, A. L., & Dunn, J. (1999). Theory of Mind, Emotion Understanding, Language, and Family Background: Individual Differences and Interrelations. Child Development, 70(4), 853–865. https://doi.org/10.1111/1467-8624.00061
- Daum, M., Attig, M., Gunawan, R., Prinz, W., & Gredebäck, G. (2012). Actions seen through babies' eyes: A dissociation between looking time and predictive gaze. *Frontiers in Psychology*, *3*, 370–370. https://doi.org/10.3389/fpsyg.2012.00370
- De Bruin, L., & Newen, A. (2014). The developmental paradox of false belief understanding: a dual-system solution. *Synthese (Dordrecht), 191*(3), 297–320. https://doi.org/10.1007/s11229-012-0127-6
- Dixson, H., Komugabe-Dixson, A., Dixson, B., & Low, J. (2018). Scaling theory of mind in a small-scale society: a case study from Vanuatu. *Child Development*, *89*(6), 2157–2175. https://doi.org/10.1111/cdev.12919
- Donnarumma, F., Costantini, M., Ambrosini, E., Friston, K., & Pezzulo, G. (2017). Action perception as hypothesis testing. *Cortex*, *89*, 45–60. https://doi.org/10.1016/j.cortex.2017.01.016
- Dushanova, J., & Donoghue, J. (2010). Neurons in primary motor cortex engaged during action observation. *The European Journal of Neuroscience*, *31*(2), 386–398. https://doi.org/10.1111/j.1460-9568.2009.07067.x
- Edwards, K., & Low, J. (2017). Reaction time profiles of adults' action prediction reveal two mindreading systems. *Cognition*, *160*, 1-16.
- Eshuis, R., Coventry, K. R., & Vulchanova, M. (2009). Predictive eye movements are driven by goals, not by the mirror neuron system. *Psychological Science*, *20*(4), 438–440. https://doi.org/10.1111/j.1467-9280.2009.02317.x
- Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience*, *9*(7), 878–879. https://doi.org/10.1038/nn1729
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Fenici, M. (2015). A simple explanation of apparent early mindreading: infants' sensitivity to goals and gaze direction. *Phenomenology and the Cognitive Sciences*, *14*(3), 497–515. https://doi.org/10.1007/s11097-014-9345-3

- Fischer, M., Prinz, J., & Lotz, K. (2008). Grasp cueing shows obligatory attention to action goals. *Quarterly Journal of Experimental Psychology (2006), 61*(6), 860–868. https://doi.org/10.1080/17470210701623738
- Fischer, M. & Szymkowiak, A. (2004). Joint attention for pointing but not grasping postures. *Cortex, 40*(1), 168–170. https://doi.org/10.1016/S0010-9452(08)70937-7
- Fizke, E., Butterfill, S., van de Loo, L., Reindl, E., & Rakoczy, H. (2017). Are there signature limits in early theory of mind? Journal of Experimental Child Psychology, 162, 209– 224. https://doi.org/10.1016/j.jecp.2017.05.005
- Flanagan, J., & Johansson, R. (2003). Action plans used in action observation. *Nature*, 424(6950), 769–771. https://doi.org/10.1038/nature01861
- Ganglmayer, K., Attig, M., Daum, M., & Paulus, M. (2019). Infants' perception of goal-directed actions: A multi-lab replication reveals that infants anticipate paths and not goals. *Infant Behavior & Development, 57*, 101340. https://doi.org/10.1016/j.infbeh.2019.101340
- Gao, T., Newman, G., Scholl, B. (2009). The psychophysics of chasing: A case study in the perception of animacy. *Cognitive Psychology*, *59*(2), 154–179. https://doi.org/10.1016/j.cogpsych.2009.03.001
- Garrido-Vásquez, P., & Schubö, A. (2014). Modulation of visual attention by object affordance. *Frontiers in Psychology*, *5*, 59–59. https://doi.org/10.3389/fpsyg.2014.00059
- Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165–193. https://doi.org/10.1016/0010-0277(95)00661-H
- Gönül, G., & Paulus, M. (2021). Children's reasoning about the efficiency of others' actions: The development of rational action prediction. *Journal of Experimental Child Psychology*, 204, 105035–105035. https://doi.org/10.1016/j.jecp.2020.105035
- Goswami, U. (2020). *Cognitive development and cognitive neuroscience: the learning brain* (Second edition). Routledge.
- Green, D., Li, Q., Lockman, J., & Gredebäck, G. (2016). Culture influences action understanding in infancy: prediction of actions performed with chopsticks and spoons in Chinese and Swedish infants. *Child Development*, *87*(3), 736–746. https://doi.org/10.1111/cdev.12500
- Hamilton, A., & Grafton, S. (2008). The motor hierarchy: from kinematics to goals and intentions. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition*. (pp. 381-407). Oxford University Press.

- Handy, T., Tipper, C., Borg, J., Grafton, S., & Gazzaniga, M. (2006). Motor experience with graspable objects reduces their implicit analysis in visual- and motor-related cortex. *Brain Research*, *1097*(1), 156–166. https://doi.org/10.1016/j.brainres.2006.04.059
- Haslinger, B., Erhard, P., Altenmüller, E., Schroeder, U., Boecker, H., & Ceballos-Baumann, A. (2005). Transmodal Sensorimotor Networks during Action Observation in Professional Pianists. *Journal of Cognitive Neuroscience*, *17*(2), 282–293. https://doi.org/10.1162/0898929053124893
- Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243–259. https://doi.org/10.2307/1416950
- Henrichs, I., Elsner, C., Elsner, B., & Gredebäck, G. (2012). Goal salience affects infants' goaldirected gaze shifts. *Frontiers in Psychology*, *3*.
- Heyes, C., & Catmur, C. (2021). What happened to mirror neurons? *Perspectives On Psychological Science*.
- Higuchi, T., Imanaka, K., & Patla, A. E. (2006). Action-oriented representation of peripersonal and extrapersonal space: Insights from manual and locomotor actions 1. *Japanese Psychological Research*, *48*(3), 126-140.
- Jacob, P. (2008). What Do Mirror Neurons Contribute to Human Social Cognition? *Mind & Language*, *23*(2), 190–223. https://doi.org/10.1111/j.1468-0017.2007.00337.x
- Jeannerod, M., & Frak, V. (1999). Mental imaging of motor activity in humans. *Current Opinion in Neurobiology*, *9*(6), 735-739.
- Kahneman, D. (2012). *Thinking, fast and slow* (1st ed.). Farrar, Straus and Giroux.
- Kampis, D., Karman, P., Csibra, G., Southgate, V., & Hernik, M. (2020). A two-lab direct replication attempt of Southgate, Senju, & Csibra (2007). *PsyArXiv*.
- Kanakogi, Y., & Itakura, S. (2011). Developmental correspondence between action prediction and motor ability in early infancy. *Nature Communications*, *2*(1), 341– 341. https://doi.org/10.1038/ncomms1342
- Karmiloff-Smith, A. (1992). *Beyond modularity : a developmental perspective on cognitive science*. MIT Press.
- Keysar, B., Lin, S., & Barr, D. (2003). Limits on theory of mind use in adults. *Cognition, 89*(1), 25–41. https://doi.org/10.1016/S0010-0277(03)00064-7
- Kilner, J., Friston, K., & Frith, C. (2007a). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, *8*(3), 159–166. https://doi.org/10.1007/s10339-007-0170-2

- Kilner, J., Friston, K., & Frith, C. (2007b). The mirror-neuron system: a Bayesian perspective. *Neuroreport*, *18*(6), 619–623. https://doi.org/http://doi.org/10.1097/WNR.ob013e3281139ed0
- Kochukhova, O., & Gredebäck, G. (2010). Preverbal infants anticipate that food will be brought to the mouth: An eye tracking study of manual feeding and flying spoons. *Child Development*, *81*(6), 1729–1738. https://doi.org/10.1111/j.1467-8624.2010.01506.x
- Kovács, Á., Téglás, E., & Endress, A. (2010). The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults. *Science*, 330(6012), 1830–1834. https://doi.org/10.1126/science.1190792
- Krogh-Jespersen, S., & Woodward, A. (2018). Reaching the goal: Active experience facilitates 8month-old infants' prospective analysis of goal-based actions. *Journal of Experimental Child Psychology*, *171*, 31–45. https://doi.org/10.1016/j.jecp.2018.01.014
- Kulke L., & Hinrichs M. (2021). Implicit theory of mind under realistic social circumstances measured with mobile eye-tracking. *Scientific Reports*, *11*(1), 1215–1215. https://doi.org/10.1038/s41598-020-80614-5
- Kulke, L., Johannsen, J., & Rakoczy, H. (2019). Why can some implicit Theory of Mind tasks be replicated and others cannot? A test of mentalizing versus submentalizing accounts. *PloS One, 14*(3), e0213772–e0213772. https://doi.org/10.1371/journal.pone.0213772
- Kulke, L., Reiß, M., Krist, H., & Rakoczy, H. (2018a). How robust are anticipatory looking measures of Theory of Mind? Replication attempts across the life span. *Cognitive Development*, *46*, 97–111. https://doi.org/10.1016/j.cogdev.2017.09.001
- Kulke, L., Von Duhn, B., Schneider, D., & Rakoczy, H. (2018b). Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. *Psychological Science*, *29*(6), 888-900.
- Kulke, L., Wübker, M., & Rakoczy, H. (2019). Is implicit Theory of Mind real but hard to detect? Testing adults with different stimulus materials. *Royal Society Open Science*, 6(7), 190068–190068. https://doi.org/10.1098/rsos.190068
- Laidlaw, K., Foulsham, T., Kuhn, G., Kingstone, A., & Purves, D. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences PNAS*, *108*(14), 5548–5553. https://doi.org/10.1073/pnas.1017022108
- Lavelle, J. (2018). *The Social Mind: A Philosophical Introduction* https://doi.org/10.4324/9781315735535
- Lewis, C., Carpendale, J., & Stack, J. (2013). Anticipation and social interaction: Commentary on "Stepping off the pendulum: Why only an action-based approach can transcend the nativist–empiricist debate" by J. Allen and M. Bickhard. *Cognitive Development, 28*(2), 159–163. https://doi.org/10.1016/j.cogdev.2013.01.008

- Low, J. (2010). Preschoolers implicit and explicit false-belief understanding: Relations with complex syntactical mastery. *Child Development*, *81*(2), 597–615. https://doi.org/10.1111/j.1467-8624.2009.01418.x
- Low, J., Drummond, W., Walmsley, A., & Wang, B. (2014). Representing how rabbits quack and competitors act: limits on preschoolers' efficient ability to track perspective. *Child Development*, *85*(4), 1519–1534. https://doi.org/10.1111/cdev.12224
- Low, J., & Edwards, K. (2018). The curious case of adults' interpretations of violation-ofexpectation false belief scenarios. *Cognitive Development, 46*, 86– 96. https://doi.org/10.1016/j.cogdev.2017.07.004
- Low, J., Edwards, K., & Butterfill, S. (2020). Visibly constraining an agent modulates observers' automatic false-belief tracking. *Scientific Reports*, *10*(1), 11311–11311. https://doi.org/10.1038/s41598-020-68240-7
- Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science*, *24*(3), 305–311. https://doi.org/10.1177/0956797612451469
- Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, *105*(3), 489–512. https://doi.org/10.1016/j.cognition.2006.10.007
- Ma, A., & Norwich, B. (2007). Triangulation and Theoretical Understanding. *International Journal of Social Research Methodology*, *10*(3), 211–226 https://doi.org/10.1080/13645570701541878
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324. doi:10.3758/s13428-011-0168-7_
- Mead, G. H. (1934). Mind, self and society (Vol. 111). University of Chicago Press.: Chicago.
- Monroy, C., Gerson, S., & Hunnius, S. (2017). Infants' motor proficiency and statistical learning for actions. *Frontiers in Psychology*, *8*, 2174–2174. https://doi.org/10.3389/fpsyg.2017.02174
- Muir, R., & Lemon, R. (1983). Corticospinal neurons with a special role in precision grip. *Brain research*, *261*(2), 312-316.
- Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, educated, industrial, rich, and democratic (WEIRD) Psychology: Measuring and mapping scales of cultural and psychological distance. *Psychological Science*, *31*(6), 678–701. https://doi.org/10.1177/0956797620916782

- Naish, K., Reader, R., Houston-Price, A., Bremner, T., & Holmes, C. (2013). To eat or not to eat? Kinematics and muscle activity of reach-to-grasp movements are influenced by the action goal, but observers do not detect these differences. *Experimental Brain Research*, *225*(2), 261-275.
- Nawaz, S., & Lewis, C. (2018). Mother–child conversation and social understanding in Pakistan. *International Journal of Behavioral Development*, *42*(5), 496–505. https://doi.org/10.1177/0165025417741365
- Needham, A., Barrett, T., & Peterman, K. (2002). A pick-me-up for infants' exploratory skills: Early simulated experiences reaching for objects using "sticky mittens" enhances young infants' object exploration skills. *Infant Behavior & Development, 25*(3), 279– 295. https://doi.org/10.1016/S0163-6383(02)00097-8
- Negri, G., Rumiati, R., Zadini, A., Ukmar, M., Mahon, B., & Caramazza, A. (2007). What is the role of motor simulation in action and object recognition? Evidence from apraxia. *Cognitive Neuropsychology*, *24*(8), 795–816. https://doi.org/10.1080/02643290701707412
- Nijssen, S., Pletti, C., Paulus, M., & Müller, B. (2021). Does agency matter? Neural processing of robotic movements in 4- and 8-year olds. *Neuropsychologia*, *157*, 107853–107853. https://doi.org/10.1016/j.neuropsychologia.2021.107853
- Norenzayan, A., Henrich, J., & Heine, S. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29–29. https://doi.org/10.1038/466029a
- Oktay-Gür, N., Schulz, A., & Rakoczy, H. (2018). Children exhibit different performance patterns in explicit and implicit theory of mind tasks. *Cognition*, *173*, 60–74. https://doi.org/10.1016/j.cognition.2018.01.001
- Onishi, K., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science* (*American Association for the Advancement of Science*), 308(5719), 255–258. https://doi.org/10.1126/science.1107621
- Osei-Kofi, N., Licona, A., & Chávez, K. (2018). From Afro-Sweden with defiance: the clenched fist as coalitional gesture? *New Political Science*, *40*(1), 137–150. https://doi.org/10.1080/07393148.2017.1418576
- Ostarek, M., & Bottini, R. (2021). Towards Strong Inference in Research on Embodiment -Possibilities and Limitations of Causal Paradigms. *Journal of Cognition*, 4(1), 5–5. https://doi.org/10.5334/joc.139
- Paulus, M., Hunnius, S., van Wijngaarden, C., Vrins, S., van Rooij, I., & Bekkering, H. (2011). The role of frequency information and teleological reasoning in infants' and adults' action prediction. *Developmental Psychology*, *47*(4), 976– 983. https://doi.org/10.1037/a0023785

- Paulus, M., Schuwerk, T., Sodian, B., & Ganglmayer, K. (2017). Children's and adults' use of verbal information to visually anticipate others' actions: A study on explicit and implicit social-cognitive processing. *Cognition*, *160*, 145-152.
- Perilli, F., Barrada, J., & Maiche, A. (2013). Temporal Dynamics of Action Contribution to Object Categorization. *Psicologica: International Journal of Methodology and Experimental Psychology*, 34(2), 145-162.
- Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science (American Association for the Advancement of Science), 308*(5719), 214–216. https://doi.org/10.1126/science.1111656
- Pine, K., Reeves, L., Howlett, N., & Fletcher, B. (2013). Giving cognition a helping hand: The effect of congruent gestures on object name retrieval. *British Journal of Psychology*, *104*(1), 57-68.
- Pomiechowska, B., & Csibra, G. (2017). Motor activation during action perception depends on action interpretation. *Neuropsychologia*.
- Poulin-Dubois, D., Rakoczy, H., Burnside, K., Crivello, C., Dörrenberg, S., Edwards, K. ... & Ruffman, T. (2018). Do infants understand false beliefs? We don't know yet–A commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cognitive Development*, *48*, 302-315. https://doi.org/10.1016/j.cogdev.2018.09.005
- Powell, L., Hobbs, K., Bardis, A., Carey, S., & Saxe, R. (2018). Replications of implicit theory of mind tasks with varying representational demands. *Cognitive Development*, *46*, 40–50. https://doi.org/10.1016/j.cogdev.2017.10.004
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral* and brain sciences, 1(4), 515-526.
- Proulx, M. (2010). Size matters: Large objects capture attention in visual search. *PLoS ONE*, *5*(12), E15293.
- Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. II: Area F5 and the control of distal movements. *Experimental Brain Research*, *71*(3), 491–507. https://doi.org/10.1007/BF00248742
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131–141. https://doi.org/10.1016/0926-6410(95)00038-0
- Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews. Neuroscience*, *11*(4), 264–274. https://doi.org/10.1038/nrn2805

- Ruffman, T., Aitken, J., Wilson, A., Puri, A., & Taumoepeau, M. (2018). A re-examination of the broccoli task: Implications for children's understanding of subjective desire. *Cognitive Development*, *46*, 79–85. https://doi.org/10.1016/j.cogdev.2017.08.001
- Ruffman, T., Perner, J., & Parkin, L. (1999). How Parenting Style Affects False Belief Understanding. Social Development (Oxford, England), 8(3), 395–411. https://doi.org/10.1111/1467-9507.00103
- Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children: Statistical learning and social understanding. British Journal of Developmental Psychology, 30(1), 87–104. https://doi.org/10.1111/j.2044-835X.2011.02045.x
- Sarlegna, F., Blouin, J., Bresciani, J., Bourdin, C., Vercher, J., & Gauthier, G. (2003). Target and hand position information in the online control of goal-directed arm movements. *Experimental Brain Research*, *151*(4), 524–535. https://doi.org/10.1007/s00221-003-1504-7
- Sartori, L., Camperio-Ciani, A., Bulgheroni, M., & Castiello, U. (2014). How posture affects macaques' reach-to-grasp movements. *Experimental Brain Research*, *232*(3), 919–925. https://doi.org/10.1007/s00221-013-3804-x
- Schuwerk, T., Kampis, D., Baillargeon, R., Biro, S., Bohn, M., Byers-Heinlein, K., ... & Frank, M. (2021). Action anticipation based on an agent's epistemic state in toddlers and adults. *PsyArXiv*.
- Schuwerk, T., & Paulus, M. (2016). Preschoolers, adolescents, and adults visually anticipate an agent's efficient action; but only after having observed it frequently. *Quarterly Journal of Experimental Psychology (2006), 69*(4), 800–816. https://doi.org/10.1080/17470218.2015.1061028
- Schuwerk, T., Priewasser, B., Sodian, B., & Perner, J. (2018). The robustness and generalizability of findings on spontaneous false belief sensitivity: A replication attempt. *Royal Society Open Science*, *5*(5), 172273.
- Scott, R., & Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development*, *80*(4), 1172–1196. https://doi.org/10.1111/j.1467-8624.2009.01324.x
- Siegler, R. (2016). Continuity and change in the field of cognitive development and in the perspectives of one cognitive developmentalist. *Child Development Perspectives*, *10*(2), 128–133. https://doi.org/10.1111/cdep.12173
- Smith, L, & Thelen, E. (1993). A dynamic systems approach to development: Applications. In This book grew out of a workshop," Dynamic Systems in Development," held for the Society for Research in Child Development in Kansas City, KS, Apr 1989. The MIT Press.

- Sommerville, J., Hildebrand, E., & Crane, C. (2008). Experience matters: The impact of doing versus watching on infants' subsequent perception of tool-use events. *Developmental Psychology*, *44*(5), 1249–1256. https://doi.org/10.1037/a0012296
- Sommerville, J., Woodward, W., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1–B11. https://doi.org/10.1016/j.cognition.2004.07.004
- Song, H., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology*, *44*(6), 1789–1795. https://doi.org/10.1037/a0013774
- Song, H., Onishi, K., 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. https://doi.org/10.1016/j.cognition.2008.08.008
- Southgate, V. (2013). Do infants provide evidence that the mirror system is involved in action understanding? *Consciousness and Cognition*, *22*(3), 1114–1121. https://doi.org/10.1016/j.concog.2013.04.008
- Southgate, V., & Begus, K. (2013). Motor activation during the prediction of nonexecutable actions in infants. *Psychological Science*, *24*(6), 828–835. https://doi.org/10.1177/0956797612459766
- Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*(7), 587–592. https://doi.org/10.1111/j.1467-9280.2007.01944.x
- Stürmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence Effects With Manual Gestures and Postures: A Study of Imitation. *Journal of Experimental Psychology*. *Human Perception and Performance*, 26(6), 1746–1759. https://doi.org/10.1037/0096-1523.26.6.1746
- Van Elk, M., Viswanathan, S., van Schie, H., Bekkering, H., & Grafton, S. (2012). Pouring or chilling a bottle of wine: an fMRI study on the prospective planning of object-directed actions. *Experimental Brain Research*, *218*(2), 189–200. https://doi.org/10.1007/s00221-012-3016-9
- Vannuscorps, G., & Caramazza, A. (2017). Typical predictive eye movements during action observation without effector-specific motor simulation. *Psychonomic Bulletin & Review*, 24(4), 1152-1157.
- Wang, B., Hadi, N., & Low, J. (2015). Limits on efficient human mindreading: Convergence across Chinese adults and Semai children. *The British Journal of Psychology*, *106*(4), 724–740. https://doi.org/10.1111/bjop.12121
- Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, *72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

- Wellman, H., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5
- Woodward, A. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34. https://doi.org/10.1016/S0010-0277(98)00058-4
- Woodward, A., & Cannon, E. (2013). 16 Online action analysis: Infants' anticipation of others' intentional actions. In M. Rutherford & V. Kuhlmeier (Eds.), Social perception: Detection and interpretation of animacy, agency, and intention (pp.383-404). MIT press.
- Zani, G., Butterfill, S., & Low, J. (2020). Mindreading in the balance: adults' mediolateral leaning and anticipatory looking foretell others' action preparation in a false-belief interactive task. *Royal Society Open Science*, *7*(1), 191167–191167. https://doi.org/10.1098/rsos.191167
- Zhang, W., & Rosenbaum, D. (2008). Planning for manual positioning: The end-state comfort effect for manual abduction–adduction. *Experimental Brain Research*, *184*(3), 383-389.