Theory of mind across biological and artificial embodiment: theory, experiments and computational models

Francesca Bianco

A thesis submitted for the degree of Doctor of Philosophy (PhD)

School of Computer Science and Electronic Engineering

University of Essex

March 2022

Impact of COVID-19

My original PhD plan involved infant neuroimaging studies, computational modelling, and robotics experiments. The pandemic caused great disruption to my research project due to my inability to conduct face-to-face experiments with infants, as well as to collect data from robotics studies due to restricted access to the lab. As a result, I was not able to conduct my originally planned neuroimaging infant studies nor robotics experiments.

My original infant studies would have been the first, to my knowledge, to investigate multisensory integration for ToM development and ability through an infant false-belief study. Furthermore, they would have been a proof-of-concept of my theoretical proposal in this thesis of including complementary measures to looking behaviour in infant studies to measure ToM, such as neuroimaging and computational modelling. Critically, by combining the multisensory nature of these tasks with the neuroimaging methodology, these studies would have also investigated the plausibility of the simulation mechanism underlying infants' success in false-belief tasks, contributing to the ongoing developmental debate on early ToM ability.

The objective of my robotics experiments was twofold. Firstly, following the recent advent of developmental robotics (Sandini et al., 2021), I aimed to create a crosstalk between the developmental psychology and robotics fields for advancing our knowledge of ToM. Specifically, on the one hand, I planned to use knowledge derived from my infant studies to inspire robotic architectures and build robots with increasingly human-inspired social skills. On the other hand, I planned to conduct research in robots, as embodied agents, for inspiring new hypotheses on ToM development, its underlying mechanisms and implicated factors to be validated in infant studies. Secondly, I aimed to equip robots with the computational architecture I present in this thesis to (a) validate my computational modelling results from a simulative environment to an embodied agent, real-world scenario, and (b) test whether my architectural implementation could improve robots’ social skills and HRIs.

The pandemic forced me to implement my experimental infant studies online, where possible. Therefore, I conducted only the first (feasibility) study from my neuroimaging multisensory project, with a few changes made to the experimental paradigm to adapt it to the online setting. To implement this study online, I programmed the paradigm on the Lookit platform, which was quite challenging, as was online research with the infant population itself. Furthermore, given the limited options for online infant research, I decided, together with my supervisors, to extend my experimental studies to the limb difference population. This was done in an attempt to still investigate sensory experiences for ToM development, although from a new perspective. Indeed, adult individuals with limb difference have differing sensorimotor experiences to the general population since birth or acquired during development. Therefore, this possibly affects ToM development to different extents. While this new research project provided a viable research question, the limb difference population is not well represented in scientific research and databases including this population did not exist at the time my project was conceived. For this reason, I had to build a database from scratch by contacting associations supporting individuals with limb difference in the UK and connecting with individuals with limb difference through social media and other advertisement (website design). In-person recruitment was not possible given the COVID-19 restrictions, making this process even more challenging. Nevertheless, I managed to build a database counting more than 250 people and I implemented studies online aimed at the limb difference population. Limitations remain also with online studies for a niche adult population, especially with respect to recruitment and participants' engagement for the whole duration of the task. Therefore, a long time was required to complete studies, and some remain unfinished to this date, although ongoing.

To conclude, COVID-19 highly impacted my original research project, preventing me to conduct face-to-face, neuroimaging and robotics studies. However, it also pushed me to reinvent my project to address my original research questions in new ways and from new perspectives. Although I was not able to conduct studies and use methodologies that I was really looking forward to mastering during my PhD, I am overall happy with my hard work during these years which allowed me to continue my research in this challenging time. Nevertheless, I hope to conduct such studies in the future.

Declaration

I declare that this thesis, ‘Theory of mind across biological and artificial embodiment: theory, experiments and computational models’, represents my own work, except where otherwise stated. None of the work referred to in this thesis has been accepted in any previous application for a higher degree at this or any other University or institution. All quotations have been distinguished by quotation marks and the sources of information specifically acknowledged.

Submitted by Francesca Bianco

Abstract

A recurrent, yet still incredibly interesting, debate surrounds the development of human Theory of Mind (ToM), as well as the identification of the mechanisms and factors implicated in this cognitive ability. In addition to psychology, this debate has attracted several research fields, including robotics for building social robots. Ultimately, while humans represent the best model of ToM for implementing a machine ToM for social robots, unresolved questions regarding human ToM development and ability need addressing. In this thesis, I used a mixed approach involving different disciplines and methodologies to contribute to human and machine ToM. Specifically, through research in the infant (Part 1 of this thesis), as well as the limb difference and general populations (Part 2), and computational modelling (Part 3), I aimed at investigating (i) ToM emergence, (ii) the mechanisms underlying ToM ability and development, and (iii) factors implicated in this cognition. Merging findings from experimental and computational modelling research, I contribute to these topics in three important ways. First, I provide support for early ToM emergence and indicate its advantage towards improved prediction of others' behaviours. Second, I suggest the coexistence of the association, simulation, and teleological for mentalising mechanisms and their collaboration for achieving ToM in different scenarios. Third, I identify sensorimotor-driven embodiment and perspective taking as factors implicated in ToM ability and development, while suggesting that self-other similarity may not be a requirement for ToM. Whether multisensory integration and mental rotation impact ToM remains undetermined. Overall, my multidisciplinary findings provide insights into human and machine ToM. Specifically, experimental findings increase our general knowledge of human ToM, computational findings inform the implementation of social robotic architectures, whilst knowledge from both disciplines is exploited to advance

one another and lead to a more global understanding of ToM. Future studies are warranted to validate and extend these results.

Acknowledgments

These past years have been a journey that I was fortunate to share with a number of encouraging and supporting people. I wish to thank Silvia and Maria Laura, my supervisors from the Psychology department, for their infectious passion for research, their continued support throughout the good and bad times of my project, and for helping me become a more confident researcher. I will be forever grateful for their availability. On this note, I also thank Devon and Eleonora for their interest in my research, I wish them all the best in the future. I would like to thank Dimitri, my main supervisor from the Computer Science department, for taking up the enormous task of teaching me the majority of what I now know about programming and computational modelling. I really appreciate the countless hours he dedicated to guide me through the new world that Computer Science was for me three years ago and tirelessly fix my code. Furthermore, I wish to thank Vishuu, my supervisor from the Computer Science department, for his help in the early stage of my research project. I owe to him the realisation of one of my biggest dreams since I was in high school, that is the incredible opportunity to work with the NAO and Pepper robots.

There are likewise many friends and colleagues to thank from University of Essex: Francesca for being the model for one of my experiments and for the good times we shared; Mehrin for assisting me with functional near-infrared spectroscopy and the many laughs; Priscilla for her collaboration and knowledge about eye-tracking; Aminur for his support with Openpose; Robin for his technical support and kind words; Lena for helping me with recruitment. A particular thanks goes to Helge, who was crucial through the process of creating a database including individuals with limb difference for my research. I would also like to acknowledge our partners supporting

individuals with limb difference: IAMPOSSIBLE Foundation; Steel Bones; Reach Charity and Limb Power.

In addition, I wish to thank all my amazing friends who supported me in various ways through this journey: Noemi, Parmi, Zatte, Babbi, Laura, Ella, Sarianna, Amanjit, Nora. I am also blessed to have acquired an extra family in England who brought a dash of normality to my life during these challenging years. I would like to thank Meena, Pankaj, Aashi, Adam and Enzo for their endless generosity, interminable food packages and fun times shared together.

A colossal thank you goes to my best friend and boyfriend Kishan for all his love, support, and encouragement. This journey simply would have not been the same without him. A person who always found a way to make me smile and who always believed in me and my abilities even when I did not. And of course, for putting up with my craziness and mood-swings throughout.

Last but not least, I wish to thank my whole family around the world, for always being there for me and bringing happiness to my days. A special thank you goes to my parents, Ana and Stefano, without whom none of this would have been possible. I will be forever grateful to them for giving me everything I have needed in life; I owe it to them that I find myself here today completing this PhD, and for making me the person I am today.

Grazie, gracias, thank you!

Table of Contents

List of Figures

List of Tables

List of Boxes

Overview

Introduction

Humans’ ability to represent and predict what others think or want from short interactions is crucial to our social lives (Baker et al., 2017). This ability to attribute mental states, such as intentions, desires and beliefs, to others is referred to as Theory of Mind (ToM), or mentalising (Frith & Frith, 2005). The importance of having a ToM for humans to successfully navigate the social world is widely recognised (e.g. for collaboration, communication, imitation, teaching, deceiving, persuading (Devaine et al., 2014, Frith & Frith, 1999; Kovács et al., 2010; Rakoczy, 2017; Tomasello et al., 2005)). Furthermore, its critical role in social cognition has also been evidenced by studies assessing individuals with neurodevelopmental disorders characterised by social deficits (e.g. autism spectrum disorder (ASD)) who tend to display atypical ToM ability (e.g. Baron-Cohen et al., 1985; Murray et al., 2017; Schneider et al., 2013). This capacity is not limited to the ability to manipulate others’ behaviours (Frith & Frith, 1999). Knowing that other individuals have mental states which may differ or contrast our own, that may be inaccurate with respect to the state of the world, and that drive their behaviour, is fundamental for efficient and fluid human social interaction (Frith & Frith, 1999; Rakoczy, 2017)

Given the implications of ToM in social cognition and human-human interactions, it has been previously suggested that equipping robots with a ToM would also improve human-robot interactions (HRI). Indeed, with robots increasingly becoming part of the society, achieving more natural and successful HRI is important and providing robotic architectures with a ToM is considered one of the “*Grand Challenges of Science Robotics”* (Yang et al., 2018, p. 9). In the last few decades, the fields of artificial intelligence and robotics have greatly advanced, resulting in the development of increasingly sophisticated virtual and physical intelligent agents with complex abilities and behaviours (e.g. Bhat et al., 2016; Görür et al., 2017; Hoffmann et al., 2017; Milliez et al., 2014). Nonetheless, the integration of AI and robots among humans is still far from optimal. Two main explanations can be provided. On the one hand, AI and robotic agents' increasingly humanoid features and human-inspired complex behaviours have enhanced humans' positive attitude towards them. However, the still limited social capabilities continue to hinder humans' acceptance of AI in their daily lives and of robots as social companions (Abubshait & Wiese, 2017). On the other hand, humans have been often seen as a source of complexity, disturbance, and unpredictability that could affect autonomous agents' performance (e.g. Hiatt et al., 2011; Koay et al., 2007; Sisbot et al., 2007), thus limiting AI and robotic agents' application. Equipping robots with a ToM would aid both these issues. Indeed, an artificial ToM would endow robots and intelligent agents with increased social capabilities, by being able to learn, represent, and reason about mental states and appropriately react to them. In addition, it would allow them to understand mental states-driven human behaviour, thus decreasing humans' unpredictability and disturbance, resulting in more fluid and efficient HRI.

Nonetheless, while humans provide the best example of effective ToM (Tomasello et al., 2005), the literature suggests that we still know relatively little about ToM. Although decades of research in various disciplines, from psychology and neuroscience to artificial intelligence and robotics (Baron-Cohen et al., 1985; Devin & Alami, 2016; Frith & Frith, 2006b; Rabinowitz et al., 2018), have been dedicated to human ToM ability, key features of this cognitive skill are yet to be fully described. There are currently several debates that surround ToM, including (1) when does ToM emerge; (2) which biological and computational mechanisms underlie ToM and its development; (3) which factors (e.g. perspective taking, mental rotation, embodiment, self-other similarity, multisensory integration) are implicated in ToM ability and its development (Bianco & Ognibene, 2019, 2020; Butterfill & Apperly, 2013; Kampis et al., 2015; Skerry et al., 2013).

This thesis endeavours to address these questions and study ToM ability and development using both experimental and computational modelling methodologies. Specifically, I employed behavioural measures (e.g. eye-tracking, response accuracy, self-reported questionnaires) of ToM to assess this ability in the infant, as well as limb difference and general populations. Furthermore, I developed artificial architectures differing in their ability to process beliefs (i.e. with or without ToM) to investigate the developmental trajectory of learning explicit beliefs representations for predicting others' intentions and behaviours. Studying ToM while taking advantage of the crosstalk between disciplines allows a more global and complete understanding of this human cognition. It provides new insights into human ToM ability, while exploiting the knowledge derived from the study of human ToM to improve artificial systems and vice versa (see also Cangelosi & Schlesinger, 2018; Hassabis et al., 2017; Lake et al., 2017; Sandini et al., 2021).

For example, the contribution of my experimental studies to the discussion on the mechanisms underlying ToM in human cognition can inform the development of artificial architectures. At the same time, the computational model presented in this thesis provides (a) insights into the representation of beliefs which can be transferred to the psychology debate on ToM, and (b) testable predictions that can be verified in future studies.

Crucially, the focus on ToM ability is grounded on the fact that ToM itself is a requirement for natural human-human and human-robot interaction, and fundamental to virtually every aspect of social life (Rakoczy, 2017).

Research Questions

To summarize, by using a multidisciplinary approach this thesis aims to contribute to the following research questions:

1. When does ToM emerge and what is its developmental trajectory?
2. Which mechanisms underlie ToM ability and development (i.e. association, simulation, teleological, and/or mentalising)?
3. What are the factors implicated in ToM ability and development (i.e. multisensory integration, embodiment, perspective taking, self-other similarity)?

Thesis roadmap and contributions

The primary aims of this thesis were to focus on (1) the absence of a coherent theoretical framework addressing ToM emergence and its underlying mechanisms, and (2) providing empirical evidence of this ability in infants, people with different sensorimotor abilities, and artificial architectures. Therefore, this thesis provided theoretical and methodological contributions, as well as experimental contributions. As shown in Figure 1, the thesis is organised in five main sections: (1) Overview, (2) Part 1, (3) Part 2, (4) Part 3, (5) Conclusions.

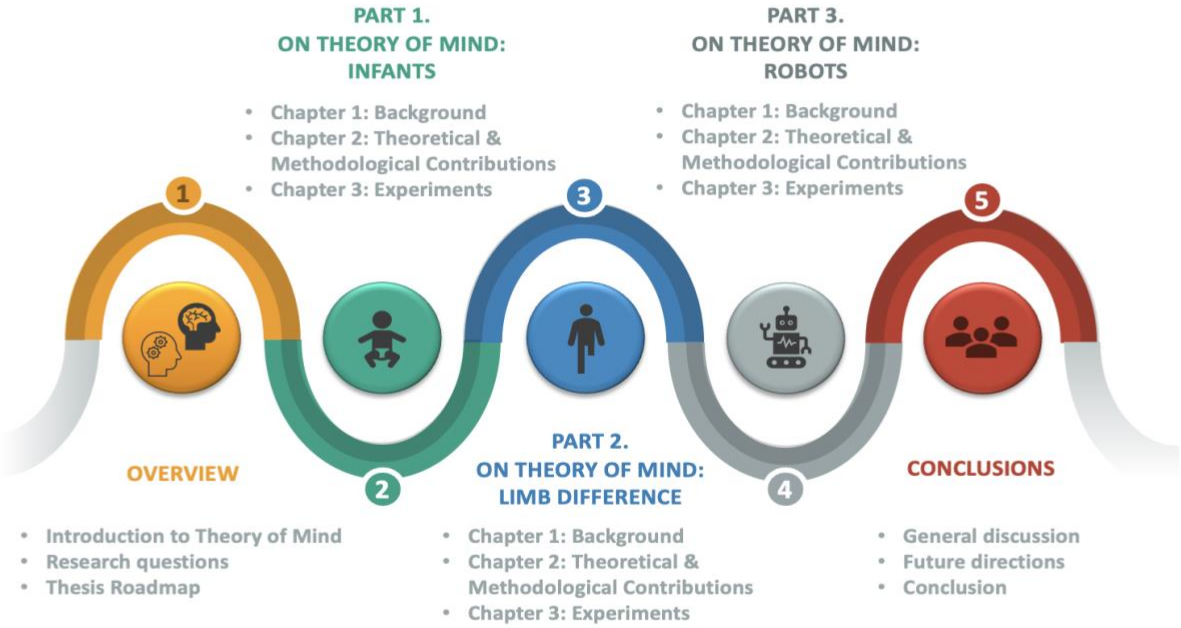


Figure 1. Thesis Roadmap (diagram template from Showeet.com)

Part 1 focuses on evidence of ToM in the infant population and is divided in three chapters.

- ***Chapter 1*** provides the *Background* on theoretical proposals and empirical evidence in the literature of ToM in the infant population.
- ***Chapter 2*** focuses on the *Theoretical and Methodological contributions* of this thesis with respect to the study of ToM through the infant population. Specifically, it introduces innovative methodologies to “standardise” the interpretation of infant false-belief paradigms results. Neuroimaging and computational modelling are proposed as complementary measures to looking behaviour to assess ToM in infant false-belief tasks. Furthermore, two innovative approaches are described, i.e. predictive coding and multisensory integration, which may rely on such methodologies to study ToM from new perspectives.

- ***Chapter 3*** investigates ToM in the infant population. Specifically, it introduces the design and implementation of a new multisensory online false-belief paradigm for the study of ToM in infants. Findings from this chapter contribute to the debate surrounding ToM emergence and factors implicated in ToM (i.e. multisensory integration).

Part 2 focuses on evidence of ToM in the limb difference population and is also divided in three chapters.

- ***Chapter 1*** introduces the limb difference population and provides the *Background* on empirical evidence in the literature of ToM in this population.
- ***Chapter 2*** focuses on the *Theoretical and Methodological contributions* of this thesis with respect to the study of ToM through the limb difference population. It proposes to study the mechanisms behind the development of ToM from a new perspective, i.e. that of individuals whose sensorimotor experiences differ the most from the majority of the population. It introduces different approaches to be used with this population in future studies to further shed light on the mechanisms and factors underlying ToM.
- ***Chapter 3*** investigates ToM in the limb difference population. In more detail, it shows for the first time the utility of several ToM tasks (including both implicit and explicit measures of ToM) for differentiating ToM ability in individuals with differing sensorimotor abilities vs the general population. This chapter is subdivided into three studies, which contribute to the debate surrounding ToM emergence, the mechanisms underlying ToM ability and development, as well as the factors implicated in ToM (i.e. embodiment, self-other similarity,

perspective taking and mental rotation). Furthermore, a unified summary of prevalent characteristics of individuals with limb difference is provided.

Part 3 focuses on evidence of ToM in robotic systems and computational modelling implementations of ToM. This section is also divided in three chapters.

- ***Chapter 1*** provides the *Background* on implementations of ToM in robotic architectures and computational modelling, with a specific interest in beliefs.
- ***Chapter 2*** focuses on the *Theoretical and Methodological contributions* of this thesis with respect to equipping robots with a ToM. In more detail, it describes the approach proposed in this thesis for adaptive ToM architectures, including the "like them" assumption and multi-task learning.
- ***Chapter 3*** introduces the implementation and comparison of two artificial neural networks differing in their ability to learn to explicitly represent others' beliefs. This chapter is subdivided into four series of studies, which contribute to the debate surrounding ToM emergence, the mechanisms underlying ToM ability and development, as well as the factors implicated in ToM (i.e. self-other similarity). Furthermore, it informs the development of architectures for social robots, by presenting a simple, yet advantageous ToM architecture able to outperform a non-ToM architecture in various scenarios involving the prediction of others' intentions and behaviours.

The **Conclusions** section provides a unified discussion of the (developmental) psychology and computational modelling experimental contributions that were presented in this thesis. These are organised by sections representing the original research questions on ToM. Furthermore, it presents future research directions that this thesis may inspire.

This thesis contains work reported in the following peer-reviewed publications or in preparation for submission:

- **“Transferring Adaptive Theory of Mind to Social Robots: Insights from Developmental Psychology to Robotics”. Bianco F.**, Ognibene D. (2019). In: Salichs M. et al. (eds) Social Robotics. ICSR 2019. Lecture Notes in Computer Science, vol 11876. Springer, Cham.

- **“From Psychological Intention Recognition Theories to Adaptive Theory of Mind for Robots: Computational Models”. Bianco F.**, Ognibene D. (2020). In: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20). Association for Computing Machinery, New York, NY, USA, 136–138.

- **“Functional advantages of an adaptive Theory of Mind for robotics: a review of current architectures". Bianco F.**, Ognibene D. (2019). In 11th Computer Science and Electronic Engineering (CEEC), pp. 139-143

- **“Innovative ways to measure Theory of Mind in infants: A multidisciplinary contribution to the developmental psychology debate”. Bianco F.**, Filippetti M.L., Ognibene D., Rigato S. **(in preparation)**

- **“Enhanced Theory of Mind in Individuals with Limb Difference: Embodiment for Theory of Mind Development and Ability”. Bianco F.**, Filippetti M.L., Ognibene D., Rigato S. **(in preparation)**

- **“Prevalent characteristics among individuals with limb difference: a population-based report”.** Bianco F., Rigato S., Ognibene D., Filippetti M.L. **(in preparation)**

- **“’Like Them’: Developmental synergy between behaviour prediction and explicit representations of others’ beliefs in a deep-learning model of Theory of Mind”. Bianco F.**, Filippetti M.L., Rigato S., Ognibene D. **(in preparation)**

Part 1

On Theory of Mind: infants

1. Background

Several studies have attempted to address ToM emergence by assessing children of different age groups in tasks specifically aimed to investigate their understanding of others' mental states (Brooks & Meltzoff, 2015; Carr et al., 2018; Peterson & Wellman, 2019; Poulin-Dubois et al., 2020). While it is now widely accepted that children older than 4 years of age show evidence of ToM ability, whether this cognitive ability can be observed at younger ages is yet to be confirmed (Apperly & Butterfill, 2009; Ruffman & Perner, 2005; Sodian & Kristen, 2016; Wellman et al., 2001). Findings in support of early ToM ability in infants come from behavioural, neuroimaging, and computational studies (see Table 1 for further details on these studies and measures).

For example, a study by Southgate and Vernetti (2014) investigated early ToM ability in 6-month-old infants by assessing their brain activity through electroencephalography (EEG) in response to an agent's varying belief states and related actions. Results from this study suggested that 6-month-old infants are able to both create and update representations of another person's beliefs, even when these are incongruent to their own beliefs, and that such representations guide their predictions of that person's future behaviour. In a behavioural study, Luo (2011) investigated early ToM ability in 10-month-old infants using a preferential looking paradigm, where infants' looking times in response to a person's actions were considered as a measure of belief understanding. Infants correctly interpreted a person's choice of toys based on the person's varying belief states, thus indicating belief understanding and tracking in infants. Overall, together with many other studies in infants under 3 years of age (see Table 1), these findings suggest that young infants can understand and update their beliefs about others' beliefs, thus supporting early

ToM ability. However, to what extent the claim that infants possess ToM can be made is an open discussion in developmental psychology due to controversies surrounding methodology, result interpretation and failed replications. In the next section, I outline the main points of debate surrounding ToM first emergence, including the biological and computational processes proposed to underlie ToM in infancy, contrasting theories and findings, as well as non-replication studies.

Table 1. Evidence supporting early infants' Theory of Mind ability through implicit tasks during the first three years of life.

Study	Infants' age (mo)	Task Type	Paradigm Type (implicit vs explicit)	ToM Measure (behavioural vs neural vs computational)	Main Findings
Behavioural					
Kovács et al. (2010)	7	Visual object detection task	Implicit	Looking time	Beliefs of an agent (irrelevant to performing the task) modulated infants' looking times, even after the agent had left the scene.
Luo (2011) 05/07/2022 20:29:00	10	False-belief task (presence)	Implicit	VOE	Infants associated a preference to the agent with respect to a toy and looked reliably longer when the agent acted in a way that was inconsistent with her preference.
Onishi & Baillargeon (2005)	15	False-belief task (location)	Implicit	VOE	Infants expected the actor to search on the basis of her belief about the toy's location and looked reliably longer when this expectation was violated.
Scott et al. (2010)	18	False-belief task (non-obvious properties)	Implicit	VOE	Infants attributed to an agent a false belief about an object's non-obvious property (rattling noise) and looked longer when they acted in a way that was inconsistent with her false belief.
Scott et al. (2015)	17	False-belief task (identity)	Implicit	VOE	In the deception condition, the infants who saw a deceiving agent replace the rattling test toy with a non-matching silent toy looked reliably longer than those who saw her substitute a matching silent toy.
Senju et al. (2011)	18	False-belief task (location)	Implicit	AL	Anticipatory eye movements revealed that infants who experienced the opaque blindfold expected the actor's action in accord with her having a false belief about the object's location, but infants who

					experienced the trick blindfold did not.
Song & Baillargeon (2008)	14	False-belief task (identity)	Implicit	VOE	Infants expected the agent to be misled by the tuft's resemblance to the doll's hair and to falsely perceive it as belonging to the doll, as they looked longer when she did not search for the doll in the hair box.
Southgate et al. (2007)	25	False-belief task (location/ presence)	Implicit	AL	Infants correctly anticipated an actor's actions when these actions could be predicted only by attributing a false belief to the actor.
Surian et al. (2007)	13	False-belief task (location)	Implicit	VOE	Infants expected searches for an object to be effective when--and only when--the agent knew the location of the desired object.
Träuble et al. (2010)	15	False-belief task (location)	Implicit	VOE	Infants accepted visual as well as manual information access as a proper basis for belief induction and looked longer when the agent behaved in a way that was inconsistent with her belief.
Yott & Poulin-Dubois (2012)	18	False-belief task (location)	Implicit	VOE	After habituating infants to the atypical behavioural rule of looking into box B after seeing the object being places in location A, infants looked significantly longer at the display when the experimenter looked for the toy in the full box (box with the toy) compared to infants who observed the experimenter search in the empty box (box without the toy).
Moriguchi et al. (2018)	18	False-belief task (location)	Implicit	VOE	Infants expected the actor to search in a specific box on the basis of her belief about the toy's location and looked reliably longer when this expectation was violated.
Neural					
Hyde et al. (2018)	7	False-belief task (location)	Implicit	fNIRS	Infants' TPJ activity distinguished between scenarios when another person's belief about the location of the object was false compared with scenarios when the belief was true.

Kampis et al. (2015)	8	Occlusion events from multiple perspectives	Implicit	EEG	Gamma-band activity was observed (a) when an object was occluded from the infants' perspective, as well as (b) when it was occluded only from the other person, and (c) when subsequently the object disappeared, but the person falsely believed the object to be present.
Southgate & Vernetti (2014)	6	False-belief task (presence)	Implicit	EEG	When an agent had a false belief that a ball was in the box, motor activity in the infant brain (sensorimotor alpha suppression) indicated that infants predicted she would reach for the box. The same was not valid when the agent had a false belief that a ball was not in the box.
Computational + Behavioural					
Hamlin et al. (2013)	10	Social Evaluation Task	Implicit	Looking time + Bayesian modelling	Comparison of computational models involving a mentalistic vs simpler vs non-mentalistic inferences suggested that infants are most likely to engage in mentalistic social evaluation. These results are in concordance with infants' looking times in response to the false-belief task.

Abbreviations: VOE: violation of expectation; AL: anticipatory looking; EEG: electroencephalography; fNIRS: functional near-infrared spectroscopy; TPJ: temporo-parietal junction.

Debates around the emergence of ToM

1.1 Experimental paradigms

The false-belief task was first introduced by Wimmer and Perner (1983) and has since been considered the standard paradigm to assess the presence of ToM ability (Poulin-Dubois et al., 2020). Briefly, false-belief tasks assess a person's ability to understand that they may have a different belief compared to another individual in the same situation, and that their representation about the world may contrast reality (Bauminger-Zviely, 2013). The traditional false-belief paradigm (which has been described elsewhere: Wimmer and Perner, 1983) has been modified through the years to evaluate different aspects of ToM and to consider varying cognitive maturity (e.g. language competence, executive functioning) in children. (For an exhaustive overview of the different versions of false-belief tasks and which aspects of ToM these aim to assess in infants, I redirect the reader to Scott and Baillargeon (2017)). The *implicit* false-belief paradigm is the most successful and utilised variation of this task in infant studies as it relies on spontaneous responses (e.g. looking behaviour) to the unfolding scene as an indication of infants' belief understanding (see BOX A below for a description of *explicit* vs *implicit* false-belief tasks). However, contrasting results, and a variety of theories and explanations have been proposed to elucidate infants' performance in virtually the same implicit false-belief task. Therefore, conclusive remarks on the age at which ToM first appears in humans, as well as the mechanisms supporting its development, have not yet been agreed upon.

Box A. Explicit vs implicit false-belief tasks.

EXPLICIT PARADIGM

- Child is asked to either make an ***explicit statement*** about the belief of the agent (e.g. what does the agent think?) or to predict how the agent is going to act (e.g. where will the agent look for an object?) or to perform some action for the agent.
- Typically, children do not pass these tests before the age of 4 years.
- Responding accurately requires sufficient verbal and executive control abilities, which might mask false belief understanding in younger children.

Explicit Elicited-Response:
In this paradigm, children first listen to a verbal narrative about an agent who holds a false belief about some aspect of a scene. Next, children are asked a direct question about the mistaken agent's likely behaviour.

IMPLICIT PARADIGM

- Child or infant's understanding of belief is inferred from their ***spontaneous responses*** to the unfolding scene.
- Different methods for measuring spontaneous responses (e.g. neural, behavioural measures such as AL, VOE).
- Typically seen from 14-18 months of age (behavioural) and from 6 months of age (neural).
- Debated whether success at these tasks reflect infants' understanding of others' beliefs or alternative accounts, e.g. minimalist, behavioural rules or low-level processing.

Implicit Violation of Expectation (VOE):
VOE paradigms rely on the natural tendency of infants to look longer at events that violate, as opposed to confirm, their expectations. In a VOE false-belief paradigm depicting an agent holding a false belief about a situation, infants' looking times increase when an agent behaves in a way that is inconsistent with his false belief.

Implicit Anticipatory Looking (AL):
AL paradigms rely on the fact that infants will predict others' actions by visually anticipating the expected events. In an AL false-belief task context depicting an agent holding a false belief about the location of an object, infants' successful performance entails visually anticipating that the agent will look for the object in the wrong location - according to their false belief.

1.2 Biological processes underlying ToM

The proposition of ToM presence from a very young age has been challenged with theories on alternative, non-ToM-related biological processes that may underlie infants' behaviour in false-belief studies. In particular, three key accounts have been put forward to explain infants' behaviour in such tasks using a non-ToM interpretation, i.e. the minimalist, the low-level processing and behavioural rules accounts.

First, the minimalist account argues against full-blown mentalistic explanations of the findings from infant false-belief studies, suggesting instead the existence of an earlier and simpler form of mental state representation which only later in development results in ToM. For example, Apperly and Butterfill (2009) propose in their paper that infants under 4 years of age can efficiently represent and track belief-like states, but this early cognitive ability coexists with a later-developing ToM ability, which is sufficiently flexible to represent more cognitively demanding concepts.

Second, low-level processing suggests that good performance in false-belief tasks does not rely on successful representation of mental states (e.g. beliefs) but only on memory, attention and perception instead. Therefore, it states that evidence of ToM ability cannot be ultimately determined based on infants' success in such tasks (Heyes, 2014a, 2014b; Perner & Ruffman, 2005).

Third, the behavioural account suggests that infants' ToM competence evidenced from false-belief tasks is not a result of their ability to represent others' mental states, but it is instead due to simpler expectation of behaviour or statistical regularities. For example, Perner and Ruffman (2005) argue that infants' looking behaviour may be guided, rather than by mental state inference, by behavioural rules, e.g. people look for an object in the last place they saw it.

Despite some studies have directly addressed the plausibility of such simpler interpretations (He et al., 2011; Onishi & Baillargeon, 2005; Scott et al., 2015; Surian et al., 2007; Surian & Franchin, 2020; Yott & Poulin-Dubois, 2012 - see Table 1), the debate remains unresolved, as does the discussion regarding the possibility of ToM presence in young infants.

1.3 Computational processes underlying ToM

A second point of discussion concerns the non-ToM-specific computational processes that may underlie infants' observed behaviour in ToM studies. Specifically, three main psychological intention recognition theories, i.e. action-effect association, simulation and teleological, have been proposed in an attempt to explain how infants' performance in false-belief tasks may result from intention recognition abilities that are considered different to ToM. These theories differ from ToM in that the computational processes they describe do not fully explain humans' ability to infer abstract mental states (ToM components). Indeed, while they allow the understanding and prediction of others' intentions (or goals) from their observable behaviour, ToM pinpoints to the unobservable, internal causal structures underlying behaviour, i.e. mental states, which go beyond action observation.

First, the action-effect association theory states that goals are simply inferred by associations between an observed action and the effects that such action has produced (Csibra & Gergely, 2007). An example of the action-effect association theory in infants is provided by Woodward (1998), who utilised a visual habituation paradigm to determine that 6-month-old infants are able to encode others' actions in a consistent way with the more mature understanding of goal-directed action. Specifically, following the habituation of an actor reaching for and grasping one of two objects, infants looked longer when changing, in the test trials, the object grasped, rather than the path of motion taken by the actor to reach for the object. Overall, this indicates goal inference directed by an association created between previously observed actions and the effects of that action.

Second, the simulation theory proposes that actions are understood when the observer directly matches, or mirrors, the observed action onto their own motor system (Rizzolatti et al., 2001). An example of the simulation theory in infants is provided by Southgate et al. (2009), who suggested, through neuroimaging studies, that 9-month-old infants are able to engage in simulation for others' action prediction. Specifically, using EEG during a reaching paradigm, the authors reported attenuation of sensorimotor alpha band activity (i.e. neural indicator of action prediction) in infants both when reaching for an object themselves and when observing an agent reaching for an object. Interestingly, the authors observed this simulated motor activation in the infant brain to initiate before the observation of the agent's action, i.e. once it could be anticipated, thus implicating simulation for prediction of others' behaviour.

Third, the teleological theory proposes that outcomes of actions can be recognised as their goals only if they are performed efficiently (Csibra & Gergely, 2007). An example of the teleological theory in infants is provided by Southgate et al. (2008), who utilised a familiarization paradigm to show that 6-8-month-old infants attributed goals to actions as long as they were efficient. Specifically, the authors familiarised infants to a human arm retrieving a ball through an efficient, goal-directed action (i.e. moving an obstructing object out of the way to retrieve a ball). Thereafter, the authors did not observe increased looking behaviour in test trials that showed the arm executing a biomechanically impossible, although still goal-directed action (i.e.

snaking around the obstructing object to retrieve the ball). These results suggest that infants extended their attributions of goal-directedness formed during familiarization even to the biomechanically impossible action in test trials, as this was efficient. Table 2 includes an overview of some developmental psychology experiments addressing infants' prediction of goal-directed behaviour with respect to the different computational theories.

Table 2. Summary table of developmental psychology experiments describing infants' cognitive ability to predict actions supported by different computational theories.

Study	Task	Findings	Supported Computational Theory
Woodward (1998)	Prediction of grasping action	After seeing a hand repeatedly grasping one of two objects, infants anticipate that the same object would be grasped again, even when the spatial location of the objects are rearranged	Association
Monroy et al. (2017)	Prediction of action sequence	Observing actions, but not visual events, influenced toddlers' action choices when associated with an effect	Association
Southgate et al. (2009)	Prediction of grasping action	Infants display overlapping neural activity during execution and observation of actions, but this activation, rather than being directly induced by the visual input, is driven by infants' understanding of a forthcoming action	Simulation
Skerry et al. (2013)	Prediction of grasping action	Infants apply a general assumption of efficient action as soon as they have sufficient information (possibly derived from their own action experience) to identify an agent's goal in a given instance	Possibly Simulation for Teleological
Gergely & Csibra (1997)	Goal attribution to uncompleted action	Infants use the principle of rational action for the interpretation and prediction of goal-directed actions, but also for making productive inferences about unseen aspects of their context	Teleological
Southgate et al. (2008)	Prediction of goal-directed action	Infants appear to extend goal attribution even to biomechanically impossible actions as long as they are efficient	Teleological

These intention recognition theories have been previously used to rule out early ToM in infants and children undertaking false-belief tasks, in an attempt to explain infants' success in such tasks with simpler, non-ToM interpretations. For example, as opposed to a ToM interpretation, the teleological theory was used to interpret false-belief task results in 18-month-old infants by Priewasser et al. (2018), as well as the association (De Bruin & Newen, 2012) and simulation (Asakura & Inui, 2016) theories in 3- to 6-year-old children (see Part 1, chapter 2.3 - "Computational Modelling" for more details). However, it is still unclear whether these theories can be accounted as computational processes underlying ToM or be considered as precursors of ToM. Some authors (e.g. Asakura & Inui, 2016; Keysers & Gazzola, 2007) have previously supported this stance in the literature. Specifically, Asakura and Inui (2016) proposed an integration in the same framework of the simulation and teleological theories for false-belief reasoning. Keysers and Gazzola (2007) brought forward the idea of a continuum between simulation and ToM. Nonetheless, this remains to be empirically proven. Box B presents a discussion on simulation and teleological theories as precursors of ToM; whereas Table 3 presents a summary of studies describing precursor computational models of ToM.

Box B. Discussion on simulation and teleological theories as precursors of ToM.

- **Teleological theory for ToM**

The *teleological theory* is one of the principal accounts utilized to describe the intention recognition ability based on observable actions in both adults and infants (Gergely & Csibra, 1997; Southgate et al., 2008). However, whether this teleological account is a suitable candidate mechanism underlying ToM remains questioned.

The concepts represented in the ToM account can be considered more complex compared to those of the teleological account. In fact, although the teleological account is able to process actions to derive the goal of an agent in various situations, it is unlikely that the rationality principle may provide access to the unobservable, abstract mental states (Frith & Frith, 2006a; Koster-Hale et al., 2017). In addition, a recent review noted that there are kinds of mindreading contexts that have nothing to do with rationality or efficiency (Goldman, 2012).

Furthermore, rationality (the teleological account) is not very effective when trying to infer mental states which are subjective, as efficiency may not be the prerogative of the agent observed.

In addition, while the teleological account suggests that infants should not be able to distinguish their representation of a scene from that of an agent (thus, reality should be as construed by the infants), the ToM account presupposes the attribution of a perspective to another agent, which may be similar or differ (Luo & Baillargeon, 2010).

Nonetheless, Gergely and Csibra (1997) proposed a continuum between the teleological constructs (i.e. action, goal-state and situational constraints) and the ToM ones. They suggested that the former supposes the same computations and constructs as the latter, but the latter represents more sophisticated, abstract constructs (i.e. intentions, desires and beliefs). Furthermore, Baker et al. (2017) developed a Bayesian computational model for ToM (BToM) based on the teleological principle that was successfully validated on adult, as well as infant data (Hamlin et al., 2013).

- **Simulation theory for ToM**

The *simulation theory* is the other principal account utilized to describe the intention recognition ability based on observable actions in both adults and infants (Gallese & Goldman, 1998; Southgate et al., 2009). However, similarly to the teleological account, whether the simulation account is a suitable candidate mechanism underlying ToM remains questioned.

The simulation theory proposes that others' mental states can be understood account for the subject-specific nature of the mental states only when the observer and the observed person are similar (Frith & Frith, 2006b). This has also consequences on ToM ability. In fact, while similarity is essential to permit the transfer of mental perspectives, it may also be a disadvantage and lead to the quarantine failure (Goldman, 2012). This is described as the failure to (a) exclude own mental states which are lacking in the observed agent when making predictions and (b) consider the observed agent's mental states as self-representations.

on the basis of our own mental states (Brass et al., 2007; Frith & Frith, 2006b; Gallese & Goldman, 1998; Goldman, 2012). Therefore, in contrast to the teleological account, the simulation theory permits the representation of the same abstract mental states, given that we experience our own mental states. However, having the same desire as another person does not necessarily permit the inference of their intentions. Hence the simulation theory can be considered only a first step for ToM (Frith & Frith, 2006b).

Furthermore, the simulation account can

Against the simulation theory as a base for ToM is also some evidence of its inability to support action understanding in novel situations (Brass et al., 2007), which does not support the context-specific nature of mental states.

Nonetheless, Keysers and Gazzola (2007) proposed a model integrating the simulation and mentalistic accounts based on neural evidence. More specifically, the authors suggested that the brain areas associated with both accounts reflect simulation, even though at different levels, rather than radically different processes.

- **Integration for ToM**

Although the *teleological* and *simulation* models in some respects rely on different representations and computations, they may be important in different ToM-related situations or when dealing with specific mental states.

For example, while the teleological model might be useful to predict mental states early in development given its more innate nature (given the central rationality principle) compared to the simulation model, the latter may become valuable when humans start learning from experience and relating to other people.

Similarly, while the simulation approach may be more suitable to infer mental states triggered by bottom-up stimuli, the teleological model may be important when an increasing top-down control is necessitated (see also Part 2, Chapter 3 of this thesis). Such top-down control enabled by ToM and teleological models may enable different preparation strategies for interaction (e.g. Ognibene & Demiris, 2013).

In Bianco and Ognibene (2019) an integration of the simulation and teleological models for ToM was suggested. A complementary view of the models, rather than a contrasting one, was proposed.

For example, although mirroring does not necessarily imply inference and prediction of the final intentions and beliefs of an agent, it may help with the action sequences necessary to reach that goal state (i.e. the trajectory to reach the final state). This may favour the teleological reasoning which may provide further information to infer and predict the mental states of the agent observed. The same might be true also in the opposite direction.
While the teleological model might provide information on possible trajectories of observed actions to infer the agent’s mental states, simulation may allow the correct inference of intentions, desires or beliefs by choosing between such options through internal simulation.

Table 3. Summary of psychology experiments describing precursor computational models of human Theory of Mind ability.

Study	Precursor of Theory of Mind	Proposal	Experiments
Keysers & Gazzola (2007)	Simulation as precursor of Theory of Mind	Based on neural evidence: brain areas associated with both accounts represent simulation (even though at different levels)	N/A
Gergely & Csibra (1997)	Teleological as precursor of Theory of Mind	Continuum between teleological constructs (i.e. action, goal-state and situational constraints) and mentalistic ones (i.e. intentions, desires and beliefs), with the latter supposing same computations and constructs of the former but representing more sophisticated, abstract constructs	12-month-old infants had to infer a goal state to rationalize the incomplete action, whose end state was occluded from them, as an efficient 'chasing' action
Baker et al. (2017)	Teleological as precursor of Theory of Mind	Bayesian computational model for ToM based on the teleological principle: adults were suggested to follow this model to infer the mental states behind an agent's behaviour using priors	'Food-trucks' scenario, using animated two-dimensional displays of an agent navigating through simple grid-worlds. Observers had to infer agents' mental states according to their trajectory
Hamlin et al. (2013)	Teleological as precursor of Theory of Mind	Bayesian computational model for ToM based on the teleological principle: infants were suggested to follow this model to infer the mental states behind an agent's behaviour	Social evaluation task assessing infants' judgement of intentions (helpful and harmful) of observed agents during interaction

1.4 Failed replications

Finally, further alimenting the psychological debate on ToM emergence is the presence of replication studies that failed to observe previously suggested ToM ability in infant false-belief tasks. For example, a seminal paper by Southgate et al. (2007) used eye-tracking to measure 25-month-old infants' AL behaviour following familiarisation in response to an agent holding true or false beliefs regarding an object location. Results from this study indicated that 25-month-olds had the ability to understand false beliefs and predict others' actions based on others' beliefs; however, these findings could not be replicated in a recent two-lab direct replication attempt including two of the original authors (Kampis et al., 2020). The findings from this paper have been used several times by later studies as evidence of infant false belief understanding in support of potential ToM competence. However, as the authors mention following their failed replication study, their paradigm does not reliably replicate the original results and thus cannot be used as evidence for further ToM ability. Similarly, Kulke et al. (2018), who attempted to replicate four influential implicit ToM tasks, could only reliably replicate 1 of the 4 studies. Previous results indicating 18-month-old infants' ability to take into consideration false beliefs when predicting someone else's actions were also not replicated by Powell et al. (2018). (I refer the reader to Kulke and Rakoczy (2018) for a recent overview of replication success in infant false-belief studies for the assessment of ToM ability). Nonetheless, explanations behind these failed replications have been proposed, including (a) slight methodological changes and (b) participants' individual differences (Crivello & Poulin-Dubois, 2018); as well as (c) differences in study procedure, participant motivation and attention between studies (Baillargeon et al., 2018).

However, of particular interest are some further points raised by Kampis et al. (2020) and Baillargeon et al. (2018) regarding the utility of some paradigms to elicit the behaviours on which they rely. Specifically, Kampis et al. (2020) concluded that their paradigm does not reliably elicit the behaviour of interest (i.e. belief-based action prediction), likely resulting in the failed replication. This seems to be in line with Baillargeon et al. (2018)'s commentary which focuses on the inability of the AL false-belief paradigm to consistently elicit spontaneous action prediction or belief tracking in different age groups and populations. This is likely because AL does not elicit basic action prediction even during familiarisation trials in several failed replication studies. In agreement with this proposition, some previous studies have questioned the suitability of the current false-belief paradigms to assess ToM in infants, indicating that success in such tasks may not require ToM ability (Bloom & German, 2000). However, this may not be the case, as I discuss with my proposal in the next section.

2. Theoretical and Methodological Contributions

2.1 Proposal

Given the existent contrasting results and theories, as well as failed replications, in this thesis I suggest that it may not be the paradigm itself the reason why ToM-related behaviour is not observed in some infant studies. Rather, I believe that the behavioural measure that is mostly used for infant ToM assessment in false-belief tasks (i.e. looking behaviour) may not necessarily be the most appropriate measure of ToM in infants *when used alone*. Indeed, most of the mentioned papers employed looking behaviour to determine ToM presence in infants, generally by measuring AL, looking times, eye-gaze and pupil dilation. However, Dörrenberg et al. (2018) investigated the replicability and validity of implicit ToM tasks using 4 different measures in 24-month-olds, including AL, looking times and pupil dilation in a violation-of-expectation (VOE) task. The authors indicated that AL is not a reliable measure of ToM as they could not replicate previous findings. Similarly, looking times and pupil dilation were found to be sensitive to certain control conditions of the task (e.g. presentation order of outcomes) and thus to have limited suitability to describe implicit ToM processes. In accordance with this view, the sole looking behaviour measure used to investigate infant ToM in this thesis (Part 1, 3.1) may not have been sufficient to fully capture ToM ability in 18-month-olds.

To put my proposal further into context, on the one hand, Kampis et al. (2020) described their AL paradigm not to reliably elicit the desired behaviour in infants, i.e. belief-based action prediction, thus failing to show early belief understanding. On the other hand, a study by Southgate and Vernetti (2014) managed to identify activation of brain regions implicated in such behaviour (belief-based action prediction) in 6-month-old infants using neuroimaging. Attempts to replicate the results from this electrophysiological study are not yet available. However, infants' success in this task may indicate that rather than the paradigm not being able to elicit the desired behaviour, the caveat may lie in the chosen measure to assess such behaviour, i.e. the looking behavioural measure. Overall, the evidence here reported has important implications for ToM research. First, it suggests that different measures may have a varying ability to access *implicit* ToM processing. Second, eye-looking behaviour as a behavioural measure for ToM may not fully capture infants' ToM ability and explain ToM-related behaviour *when used alone*.

Overall, additional methodologies to "standardise" the interpretation of the results obtained with false-belief paradigms are needed to determine the presence (or absence) of ToM in infants in a way that can be considered effective and reliable. In the next sections, I introduce two methodologies that might be used as complementary to looking behaviour when assessing ToM in infants through implicit false-belief tasks, i.e. computational modelling and neuroimaging. These are discussed in light of the current literature on false-belief tasks in infants. Furthermore, I describe two innovative approaches to the study of ToM, i.e. predictive coding and multisensory integration, which may rely on such methodologies. My aim is to show the potential of using complementary methodologies for the study of ToM from a new perspective and to inspire future studies in this direction.

Innovative methodologies to interpret infant ToM results in false-belief tasks

2.2 Computational Modelling

In the last few decades, computational modelling has been increasingly applied to various research fields including psychology (Blohm et al., 2020), providing great insights and increasing our knowledge on the mechanisms underlying cognition in health and disease (Ask & Reza, 2016). Although different computational models can be developed with different goals in mind (Blohm et al., 2020), they overall represent a very powerful tool to describe brain mechanisms and computations, as well as test existing competing theories and develop new ones to explain observed behaviours more reliably. This is promising for the present discussion, given the many contrasting theories and findings surrounding infant ToM in false-belief tasks. I here suggest introducing computational modelling in infant false-belief studies to help shed light on the debate on ToM emergence and its development.

One way in which computational modelling could aid the existent debate is by comparing models representing processes supported by contrasting theories. Specifically, implementing computational models of infants' performance in false-beliefs tasks may provide a means to more reliably determine whether ToM is the process underlying infants' success in these tasks. To my knowledge, only Priewasser et al. (2018) utilised computational modelling to examine the brain computations that may underlie infant behaviour in false-belief tasks. The authors adapted the study by Buttelmann et al. (2009) and investigated the helping behaviour of 18-to-32-month-olds during a false-belief task that required infants to help an agent retrieve a hidden toy. Specifically, the authors reasoned that infants' helping behaviour evidenced in this false-belief task could be interpreted as either (1) infants understanding others' beliefs and acting based on these beliefs (mentalising), or (2) infants utilising reasoning based on others' previous actions to recognise their goals, which does not require tracking others' beliefs (teleological reasoning). According to the first interpretation, infants' helping behaviour would be limited to helping the agent when she had a false belief about an object's location. However, according to the second interpretation, infants' helping behaviour would be driven by their knowledge of the agent (i.e. looking for the toy in the wrong box), thus without attributing a false belief to the agent. To determine which reasoning best described the infants' helping behaviour observed in their false-belief task, the authors used Bayesian model comparison, which involves the testing of competing theoretical explanations given a behaviour, to identify which model best described the observed data (*a posteriori*). Their model comparison indicated that the teleological account was more likely to support the observed data, and to describe the computational processing underlying infants' behaviour in their task than a mentalistic account. This work therefore shows that by testing contrasting theories on the same set of data, computational modelling can help to shed light into the processes underlying infants' early competence in the false belief task. I believe that this approach should inspire future infant false-belief studies and may contribute towards the solution of the current debates. Indeed, none of the above studies contesting the findings of ToM presence in infant false-belief tasks (Apperly & Butterfill, 2009; Heyes, 2014a, 2014b; Perner & Ruffman, 2005) have provided concrete evidence of its absence. While they suggested alternative minimalist, behavioural rule or low-processing accounts, such proposals have not been directly tested, thus preventing any conclusion that is warranted by the data. A few other studies compared ToM models with other simpler models to provide evidence of ToM as the mechanism underlying the behaviours observed in their task, although this was done in the context of a social evaluation, rather than false-belief task (Hamlin et al., 2013) and in older children and adults only (Baker et al., 2017). In their study, Hamlin et al. (2013) evaluated infants' understanding of others' beliefs by testing which of the following strategies would drive 10-month-old infants' judgements in a social evaluation task: (1) the analysis of the mental states at play (i.e. the protagonist's preference for an object and the helper’s knowledge of such preference), (2) the assignment of the same knowledge about the protagonist’s mental states to both the protagonist and the helper, or (3) the reliance on low-level cues in the scene. To do so, the authors compared a Bayesian computational ToM model with both simpler mentalistic models and non-mentalistic models, which were used in combination with the more classical behavioural measure of ToM (looking behaviour). The authors concluded that the Bayesian computational ToM model best described infant looking behaviour in their task, providing evidence of ToM processing rather than non-ToM computations. Therefore, these findings support the interpretation of an early ToM ability in infants as seen in this false-belief task. This provides an example of how computational modelling can be used to aid the psychological debate on ToM emergence. While these results contrast the previously described findings from Priewasser et al. (2018), it is worth mentioning that in the latter study an older infant population was recruited (18-32- vs 10-month-olds) and a different task was used (helping vs social evaluation task). Therefore, it is possible that these differences could have led to such contrasting results. Conversely, it is also possible that age differences between studies could reveal varying presence of ToM-related behaviours throughout development. The coexistence of different computational mechanisms supporting ToM in different scenarios (e.g. different experimental settings and populations) and at different developmental stages is further reinforced in the experimental sections of this thesis (Parts 2 and 3). Indeed, this thesis supports the presence of a shared computational framework which supports all previously indicated computational mechanisms underlying ToM when considering different scenarios, e.g. association for simple environments with limited factors influencing others’ behaviours; simulation, teleological or mentalising for more complex environments and more challenging scenarios (see Part 2, 2.3 and Part 3, 3.3 for further details). While it is evident that there exists a range of studies contributing to the modelling of ToM, it is also clear that computational modelling could provide useful result interpretation. Overall, this suggests that computational modelling represents a technique of choice in infant false-belief tasks in support of more classical behavioural measures, e.g. looking behaviour.

In addition to testing contrasting theories, there are other ways in which computational modelling could aid the debate on ToM emergence. First, computational models are useful for drawing new hypotheses and advancing theories able to explain an observed behaviour in a given task. In more detail, computational models can infer hidden variables and lead to new hypotheses that would otherwise not have been considered. Therefore, they can prove extremely useful for the interpretation of data, hypothesis generation, as well as the discovery of key factors underlying higher-order cognitive abilities, including ToM. Computational modelling has not yet been used for such purposes in infant ToM false-belief tasks, or ToM in general. However, I would like to mention Rabinowitz et al. (2018) as an example of the discovery of new factors which may influence the behaviour under investigation, and which may not have been captured otherwise. Briefly, Rabinowitz et al. (2018) trained a neural network to investigate prediction of others' behaviours based on their past behaviours. Interestingly, their neural network independently discovered patterns in others' behaviours, which were used to categorise agents into groups. These patterns had not been captured by Rabinowitz and colleagues; thus, modelling enriched the authors' interpretations of their results.

Computational models may also provide a means to determine constraints and factors affecting ToM development, which may improve the debate on ToM emergence and, in turn, improve infant false-belief studies. This can also help designing more efficient experimental frameworks and providing more precise predictions. An example of this approach could be building a computational model combining an innate (e.g. automatic processing of beliefs or top-down component) and an experience-based component (e.g. stimuli-driven beliefs processing or bottom-up component) for belief-based reasoning. By applying the model to the behaviour observed in false-belief tasks, it would be possible to determine which of the two components mostly affects belief reasoning and the conditions under which the behaviour is influenced, e.g. infant age, implicit or explicit false-beliefs, social component of task, environmental volatility (see Diaconescu et al. (2014) for a similar approach applied to study social advice-driven decision-making in adults). These types of investigations are difficult to conduct using standard lab-based research. Therefore, computational modelling could inform future false-belief studies and help determine if ToM-related behaviour can be observed at a certain age, given certain conditions, and when it may be impaired.

2.3 Neuroimaging

The advancements in brain imaging technology in the last decades have contributed immensely to the understanding of human brain structure and function. Neuroimaging has become a primary tool in neuroscience research, involving various techniques including positron emission tomography, functional magnetic resonance imaging, fNIRS, EEG / event-related potentials (ERPs), and magnetoencephalography (MEG) (Dick et al., 2014). Neuroimaging has also become an invaluable tool for increasing our knowledge of infant cognitive development and abilities. Specifically, neuroimaging has made possible the study of the relationship between the structural growth of the brain and the emergence of new behavioural abilities during development (Johnson, 2001). In addition, the role that genetic and environmental factors have on learning and development has been investigated (Munakata et al., 2004). Neuroimaging allowed the assessment of the large-scale dynamic changes in the interactions between brain regions and their functional specialisation throughout development, as well as the characterisation of typical and atypical development (Johnson et al., 2002). Furthermore, it can aid the study of classic developmental and cognitive issues (Munakata et al., 2004) in a way that was not possible until a few decades ago. For an extensive review on the neuroimaging methods used in infant psychology studies, I refer the reader to Azhari et al. (2020).

Compared to behavioural measures alone, neuroimaging methods have some advantages that are critical in the context of developmental psychology (see Spelke (2002) for a more detailed analysis of this comparison). To name a few, neuroimaging methods allow the direct study of brain activity related to a cognitive ability, which can only be done indirectly through behavioural studies. Therefore, they can narrow down the space of plausible interpretations compared to behavioural data. Furthermore, neuroimaging often permits the utilisation of the same procedures and measures with children of different ages to reliably study developmental neural changes. Overall, in a similar fashion to computational modelling, neuroimaging provides information about tentative mechanisms underlying development that is not obtainable from solely behavioural studies. However, neuroimaging relies on brain activity to provide insights into the computations and biological processes behind an observed behaviour. Following Spelke's observation that "developmental neuroimaging is likely to offer new insights into questions that have been central to developmental psychology for centuries" (Spelke, 2002, p. 392), in this section I apply this view to aid the interpretation of false-belief task results in infants.

As mentioned in the introduction, some uncertainties remain regarding whether looking behaviour is an accurate and reliable measure of infant ToM in false-belief tasks. Not relying on overt behaviour, neuroimaging methods potentially represent a great additional tool to determine which cognitive process(es) infants engage in during false-belief tasks, offering researchers access to infants' brain processing throughout the task. To my knowledge, three studies have employed neuroimaging methods (EEG/ERPs and fNIRS) to investigate infants' understanding of others' minds using false-belief tasks. Kampis et al. (2015) investigated the ability of 8-month-old infants to encode another person's perspective, using gamma-band electroencephalographic activity over the temporal lobes (i.e. neural signature for sustained object representation after occlusion). The authors observed gamma-band activity when an object was occluded from the infant's perspective and from the other person's perspective only, as well as when the person falsely believed the object to be present. The authors concluded that infants have a metarepresentational understanding of others' minds (i.e. the ability to represent others' mental states) even before the onset of language and that it is possible that some basic ToM mechanisms have an innate basis. Southgate and Vernetti (2014) measured sensorimotor alpha suppression (i.e. a neural indicator of action prediction) to determine whether 6-month-olds would generate action predictions that are appropriate given an agent's current belief in a false-belief task. The authors observed infants' motor cortex activation only when the agent had a false belief on the presence of a ball placed in a box, as opposed to a false belief that the ball was not in the box, indicating that they would not predict the agent to act in the latter condition. Importantly, infants based their predictions on what the agent, rather than themselves, believed on the state of the environment, suggesting that infants can exploit their sensitivity to others' minds for action prediction from a young age. Forgács et al. (2019) tested 14-month-old infants in a false-belief task and measured ERPs associated with semantic violations (N400 component). Infants were shown to exhibit an N400-like response when observing an agent having a false-belief about a labelled object, therefore suggesting semantic violation processing, even though labels were congruent from the infant perspective. The findings from this study indicate that infants were tracking the observed agent's comprehension during the interaction.

Altogether these studies suggest that infants in their first year of life are able to represent others' beliefs and use such beliefs for predicting others' behaviour, thus supporting early ToM interpretations. With the help of neuroimaging, it was possible for the authors of these papers to identify the mechanisms underlying infants' overt behaviour in such tasks by providing neural evidence of infants' processing of others' representations of the world (in particular when contrasting their own). Furthermore, as opposed to behavioural studies which observed this ability in +18-month-olds only, by using neuroimaging measures and not relying on overt behaviour, these authors reported ToM at 6 months of age. As highlighted through these studies, neuroimaging represents a great candidate tool to determine infants' covert competence vs overt performance in false-belief tasks and whether ToM is implicated.

Similarly to computational modelling, neuroimaging can also be a means to test contrasting theories, as different brain mechanisms have been hypothesised to underlie infant early ToM ability. For example, as mentioned earlier, the simulation theory requires the matching of an observed action onto one own's motor system for it to be understood (Southgate et al., 2008) and a few studies have shown this ability in infants (e.g. Southgate et al., 2009, 2010). Therefore, neuroimaging could be utilised to determine which brain areas, as suggested by any of the alternative accounts (e.g. simulation), are active during false-belief tasks in infants. This will aid the debate on whether ToM or simpler intention recognition better describes infants' behaviour seen in such tasks. Furthermore, neuroimaging applied to false-belief studies could also guide the identification of the mechanisms underlying ToM. For example, Saxe et al. (2004) concluded in their review that reasoning about others' minds (ToM) represents a separate domain of cognition to reasoning about others' goals and actions (e.g. association, simulation and teleological theories) by taking into consideration the activation of different brain areas during the two processes.

One last great advantage of employing neuroimaging methods in false-belief infant studies is the potential to investigate the activity of brain areas associated with adult ToM in the infant brain during these tasks. To my knowledge, the study by Hyde et al. (2018) is the only one to have used this approach to provide some insights into whether infants may be engaging in ToM during false-belief tasks. Hyde and colleagues (2018) used fNIRS to assess the activity of 7-month-old infants' temporoparietal junction (TPJ), a brain area suggested to be recruited in adults during ToM. The authors found infants' TPJ activity to reflect the tracking of others' beliefs during their false-belief task, with higher activity seen when the person's belief about the location of an object was false as opposed to true. This developmental correlation strengthens the authors' result interpretation of the presence of an adult-like functional organization relevant to ToM by 7 months of age. Ultimately, the use of neuroimaging methodologies alongside behavioural paradigms can shed some light on the mechanisms underlying ToM in young infants during false-belief tasks.

2.4 Predictive Coding

“ Whether infants engage in belief-based predictive coding during implicit false-belief tasks could be key to more reliably support interpretations of early ToM ability”

The human brain has been suggested to act as a “statistical organ” (Friston, 2018) or a “prediction machine” (Clark, 2013). In particular, for successful interactions to unfold, it has been suggested that the brain makes predictions about the world, using hierarchical generative models, which are then updated based on the evidence obtained through incoming sensory input. This mechanism is referred to as predictive coding (Clark, 2013, 2015; Friston, 2005). Predictive coding is believed to be useful for the accurate interpretation of the world we live in. Indeed, given the volatile and uncertain nature of the environment, predictions need to be continuously updated based on incoming information. When there is a discrepancy between incoming information and original predictions, a prediction error is formed, which is used to interpret the incoming sensory input and inform future predictions (den Ouden et al., 2012). The importance of predictive coding in the discussion on ToM emergence is reflected from its implication in ToM ability advanced by Van de Cruys et al. (2014). In their paper, the authors suggest predictive coding as a mechanism to understand ToM deficits in people with ASD. Specifically, the authors propose ToM impairment in ASD to be driven by an inability to flexibly process VOE in social contexts. From a predictive coding perspective, this would mean that people with ASD may find it difficult to give the correct weight to prediction errors and to choose whether to use such prediction errors to update their model of the world. While this proposal remains to be confirmed through empirical data, it provides an interesting view on the potential involvement of predictive coding for typical and atypical ToM.

Other studies have suggested a role for predictive coding in ToM ability in the typical population as well (Gordon, 2021; Koster-Hale & Saxe, 2013; Richardson & Saxe, 2020). Koster-Hale and Saxe (2013) reviewed findings from several papers on attribution of goal-directed actions, beliefs and desires, as well as preferences and personalities. Based on these studies, they advanced an interpretation of the original results in terms of predictive coding proposing to extend such framework to ToM. In particular, the authors identified a decreased neural activity in the TPJ in response to others' beliefs that are predictable, which they define as a key signature of predictive coding. In other words, considering that the TPJ has been previously indicated to show a robust response while thinking about an individual's beliefs and desires (Saxe & Kanwisher, 2003), the pattern of activation identified in Koster-Hale and Saxe (2013) (i.e. decreased response to expected beliefs-driven behaviours of others) highlights a predictive code framework for ToM. This can thus be used in future ToM studies (including false-belief tasks) to interpret infant results and to indicate more reliably whether ToM is implicated in their behaviour. Specifically, if such a decreased response in the TPJ is seen during false-belief tasks, researchers could more strongly conclude that infants do indeed understand and predict others' beliefs-driven behaviours. Predictive coding was also proposed by Gordon (2021) to be at the core of ToM ability, specifically indicating that simulation implemented through predictive coding may be the process behind ToM ability (see Part 1, section 1.3 - "Computational processes underlying ToM" for an explanation of the simulation theory). In particular, the author suggested that individuals represent and update beliefs (as well as other mental states) in an agent-neutral manner, meaning that mental representations are undifferentiated between the self and other. Others' representations are then corrected using predictive coding, i.e. in response to prediction errors deriving from the testing of hypothetical modifications of such representations, until a good enough projection is achieved. While Gordon (2021)'s proposal has not been validated with human studies, it provides a compelling role of predictive coding for ToM. Overall, the relevance of predictive coding for understanding brain mechanisms is clear, as is its potential involvement in ToM ability, thus becoming a great candidate measure for ToM assessment and false-belief tasks' result interpretation.

My interest in suggesting the use of predictive coding for corroborating interpretations of infant false-belief task results further stems from its central role in the context of implicit false-belief paradigms. Indeed, such paradigms are highly reliant on predictive coding by requiring infants to *predict* others' behaviours based on their changing beliefs. More specifically, most of the studies here reviewed assessed false-belief understanding using either the VOE or AL implicit paradigms, whereby infants make predictions of agents' belief-based behaviours, which need to be updated in light of incoming sensory input (see Part 1, section 1.1 - "Experimental paradigms" for more details on these paradigms). Briefly, VOE paradigms rely on the natural tendency of infants to look longer at events that violate, as opposed to confirm, their expectations, which are built through predictive coding. Consequently, in a VOE false-belief paradigm depicting an agent holding a false belief about a situation, infants' looking times increase when the agent behaves in a way that is inconsistent with his false belief (Scott & Baillargeon, 2017). In contrast, AL paradigms rely on the fact that infants will predict others' actions by visually anticipating the expected events, which again is possible through predictive coding. Therefore, in an AL false-belief task, where for example an agent holds a false belief about the location of an object, infants' successful performance entails visually anticipating that the agent will look for the object in the wrong location – according to their false belief (Scott & Baillargeon, 2017). It is therefore clear how both these paradigms are dependent on infants' ability to engage in belief-based predictive coding during false-belief tasks. Therefore, to support an interpretation of ToM presence in infancy, a study should provide clear evidence of belief-based predictive coding being the reason behind infants' success in such tasks.

While young infants have been shown to update their internal models of the world using predictive coding to represent a volatile environment (e.g. de Klerk et al., 2016; Kayhan et al., 2019; Poli et al., 2020), whether they engage in *belief-based predictive coding* during implicit false-belief tasks has not yet been fully evidenced. In fact, most studies assessing belief-based predictive coding in young infants have solely relied on their looking behaviour during the task, using AL and VOE as a proxy of predictive behaviour. However, while infants can show a looking behaviour that is consistent with predictive coding with respect to others' beliefs, these studies cannot rule out whether the predictive behaviour displayed could be driven by other processes (i.e. not belief-based). As contested by the behavioural rule account (Perner & Ruffman, 2005), infants' longer looking in false-belief tasks may represent a VOE of e.g. object location, which does not require belief understanding. While follow-up studies have directly addressed this assumption in specific paradigms using experimental variations which could exclude behavioural rules to guide infants' performance (e.g. Scott et al., 2015; Yott & Poulin-Dubois, 2012), belief-based predictive coding remains to be conclusively shown in these tasks. Both the computational modelling and neuroimaging complementary measures introduced earlier can help investigate the types of predictive coding that infants may be engaging with during implicit false-belief tasks and rule out whether other factors (e.g. different types of VOE) may be influencing study results, as opposed to belief-based predictive coding.

In the literature there exist only three studies (e.g. Hamlin et al., 2013; Kampis et al., 2015; Southgate & Vernetti, 2014) which have provided compelling evidence of infant belief-based predictive coding during false-belief tasks and which should inspire future research designs. Southgate and Vernetti (2014) and Kampis et al. (2015) used neuroimaging to assess infants' ability to predict others' actions according to their beliefs (belief-based action prediction) and sustained object representation after occlusion (belief-based sustained object representation), respectively. Hamlin et al. (2013) combined the looking behaviour measure with computational modelling to assess infants' ability to infer mental states for social evaluation (belief-based social evaluation). Overall, although using different methodologies, these studies all evidence belief-based predicting coding in infancy. In more detail, Southgate and Vernetti (2014)'s study evidenced belief-based action prediction through motor cortex differing activation based on the agent's beliefs. Kampis et al. (2015) reported belief-based sustained object representation after occlusion through differing gamma-band activity over the temporal lobes based on the observed agent's beliefs. Ultimately, these findings provide neural evidence of infants' prediction of the agent's behaviour based on the agent's beliefs, as well as behavioural evidence of belief-based AL and VOE, respectively. Finally, a study by Hamlin et al. (2013) employed Bayesian computational modelling, in addition to the classical looking behaviour measure, to determine whether infants were engaging in mentalistic inference for social evaluation in their task. Bayesian models are mostly used to formulate predictive coding; thus, they provide strong evidence of its presence (or absence). In their study, Hamlin et al. (2013) indicate that infants had priors on agent's behaviour, which were updated taking into consideration the agent's beliefs-guided behaviours to make social evaluation; thus following a predictive coding framework. A strength of this work is that the authors corroborated their looking behaviour findings with the additional computational modelling measure. This is a great example showing how (1) the looking behavioural measure has been and continues to be greatly useful to provide initial insights into infant predictive coding in false-belief tasks, and ToM presence as a consequence, and (2) more sensitive measures to belief-based predictive coding should be combined in experimental designs to more reliably interpret ToM-related findings in implicit false-belief studies.

2.5 Multimodal integration

"[…] a full picture of cognitive development will only emerge once we consider the fact that all of these processes depend crucially on multisensory interactions"
(Bremner et al., 2012)

It is now widely accepted that the brain receives several different types of information provided by multiple sensory modalities at any given time (Parker & Robinson, 2018). To interpret such information, the brain uses both bottom-up and top-down processes to automatically integrate signals from different sensory channels. This results in a "cross-modal" stimulus, a phenomenon referred to as multisensory integration (Dionne-Dostie et al., 2015; Parker & Robinson, 2018). A cross-modal stimulus produces a different neural response product to that obtained by the individual component stimuli (Stein & Rowland, 2011). For example, during food tasting, individuals can taste and smell their food, which are inputs from different sensory channels that are perceived as an integrated cross-modal stimulus, thus contributing to the same experience. If such inputs were perceived separately by individuals, a big change in overall perception and experience would occur. A fitting example of the impact of olfactory loss on food tasting has been recently provided by (Elkholi et al., 2021), who assessed the effect of Coronavirus Disease 2019 (COVID-19) on individuals' quality of life. Specifically, the authors indicated that 84.6% of patients who lost their sense of smell due to COVID-19 reported a decreased ability to taste and enjoy food, while 66.5% mentioned reduced appetite. Ultimately, real-world events are very rarely unimodal; multisensory integration allows a complete and coherent representation of what is being perceived (Dionne-Dostie et al., 2015). Given the impact of multisensory integration in our daily life, determining the underlying mechanisms, as well as the contributions of each sensory modality to multisensory integration is warranted (Parker & Robinson, 2018).

Developmental studies suggest that the ability of processing multisensory signals develops early in life. For example, Meltzoff and Borton (1979) showed that 29-day-old infants were able to recognise an observed shape with which they had previously had tactile experience. However, the age at which multisensory integration ability develops remains debated (Dionne-Dostie et al., 2015). Furthermore, several studies in the literature have evidenced the role of multimodal integration on cognition, including its effect on emotion (Lin et al., 2020), sense of body ownership (Botvinick & Cohen, 1998; Tsakiris, 2008) and embodiment (Baumard & Osiurak, 2019). The effect on the latter two cognitive abilities has been made possible by experimental illusions that manipulate individuals' sensory experiences, thus allowing the investigation of the brain's tendency to undergo multisensory integration of inputs from different modalities (Parker & Robinson, 2018). For example, in their widely known and replicated rubber hand illusion study, Botvinick and Cohen (1998) showed that synchronous tactile stimulation of an individual’s hand while watching the same touch in a rubber hand produces a sense of body ownership over the artificial hand (i.e. resulting from visuo-tactile multisensory integration). Illusions of body-ownership, such as the rubber hand illusion, have been shown to increase the perceived physical similarity between self- and other-, which in turn changes mental representations of self and others, such as attitudes and beliefs (see Tsakiris, 2017). As a result, multisensory experiences have been suggested to drive the experience of self-other distinction and embodiment (e.g. Tsakiris, 2017). Given the relevance of self-other distinction for ToM ability, multisensory-driven embodiment has been hypothesised to be important for ToM (e.g. Steinbeis, 2016; Chasiotis et al., 2006).

Therefore, multisensory integration may be a key factor in the discussion on ToM emergence. Yet, multisensory integration for ToM development has been under-researched, especially in infant populations. However, introducing this approach to ToM research may help shed light into the mechanisms behind ToM and contribute to the interpretation of infants’ behaviour in false-belief tasks. Both the computational modelling and neuroimaging complementary measures can help investigate multisensory integration during implicit false-belief tasks. I will now highlight the reasons why I think that this approach could shed some light into infants’ flexibility towards representing others’ beliefs, which may ultimately be influencing study results.

First, ToM tasks, including implicit infant false-belief tasks, generally rely on a single modality, i.e. (mainly) vision and (at times) audition (Beaumont & Sofronoff, 2008). While such unisensory studies assessing false belief understanding in infants have provided compelling findings into ToM development, the investigation of this cognitive ability is arguably incomplete unless multiple modalities are included in the picture. There are at least two main reasons as to why this might be the case.

On the one hand, real-life scenarios usually comprise multisensory stimuli which need to be integrated to allow for correct behavioural responses. Therefore, while infants may be able to pass a false-belief task using single-modality stimuli, their performance might not translate to real-world scenarios where the integration across sensory channels is required. Furthermore, infants' performance at multisensory false-belief tasks may be more consistent and indicative of their ToM ability, compared to unisensory false-belief tasks which have resulted in contrasting findings in the literature. Therefore, a multisensory approach to ToM may be able to uncover infants' ToM ability in naturalistic scenarios, as well as inform on infants' success at false-belief tasks.

On the other hand, the ability to represent beliefs may vary based on the sensory systems through which stimuli are delivered. In this sense, uncovering which - if any - sensory systems contribute the most towards ToM development and false-belief success is an open question. For example, a potential involvement of visual and auditory systems for ToM development has been suggested by studies showing atypical ToM in older children with severe visual impairments or total blindness (Peterson et al., 2000) or with hearing impairments (Meristo et al., 2007, 2016) compared to controls. However, the extent to which such sensory impairments contribute to delayed ToM has not been directly compared, nor their potential interaction for ToM ability assessed. The introduction of multisensory infant false-belief tasks would provide a starting point to address these questions related to the representation of beliefs from information obtained through multiple sensory modalities. Multisensory integration in false-belief tasks could be implemented by manipulating the nature of the sensory information (unisensory vs multisensory) presented to the observed agent to induce false beliefs during the task. Specifically, the introduction of multisensory information for false belief induction would make it possible to investigate whether infants are able to understand and predict others' behaviours driven by beliefs induced through multisensory integration.

Second, whether similarity between self- and other- is required for engaging in ToM is debated. Indeed, embodied theories of cognition suggest that sensorimotor experiences drive embodiment and that sensorimotor-driven embodiment in turn affects higher-order cognition (e.g. Pezzulo et al., 2011, 2013). However, it may be challenging to take someone else's perspective when the representations of self- and other-bodies differ. This is in line with the simulation mechanism, which has been previously indicated as a candidate mechanism underlying ToM (e.g. Keysers & Gazzola, 2007), but whose flexibility to understand "different others" has been questioned (Frith & Frith, 2006b). By introducing multisensory false-belief tasks involving the tactile sensory system, and by coupling it with, for example, neuroimaging or computational modelling, the mechanisms of sensorimotor-driven embodiment and simulation may be further investigated. This would in turn contribute to determining whether babies, who have bodies that necessarily differ to the adult agents usually involved in false-belief tasks, can indeed represent their beliefs, and thus engage in ToM. Ultimately, results using this approach may help the interpretation of false-belief studies, providing stronger support of (or evidence against) infants' engagement in ToM ability during false-belief tasks.

To my knowledge, only a few infant studies exist in the literature utilising false-belief tasks involving multiple modalities, although the authors did not interpret their findings in view of multisensory integration for ToM ability in infants (Forgács et al., 2019; Scott, 2017, Scott et al., 2010, 2015; Träuble et al., 2010). All these studies relied on infants' representation of beliefs formed through an integration of either visual and auditory or tactile sensory information, and evidenced infants' ability to pass false-belief tasks involving such multisensory stimulation.

For example, Scott et al. (2010; but see also Scott, 2017) familiarised infants with an agent shaking an object that produced a rattling sound. When the agent left the room, infants were introduced to a new object, identical to that in the familiarisation trial, which however did not rattle when shaken. Next, infants were tested on their ability to attribute a false belief to the agent returning to play with the object and showing surprise when the object did not rattle. The authors found that infants were able to infer the agent's false belief, as assessed through their looking behaviour. However, these results may also evidence infants' ability to engage in multisensory integration for ToM ability and succeeding in false-belief tasks.

Forgács et al. (2019) tested 14-month-old infants in a false-belief task measuring ERPs associated with semantic violations (N400 component). The authors suggested that infants exhibited an N400-like response when observing an agent having a false-belief about a labelled object, therefore suggesting semantic violation processing, even though labels were congruent from the infant perspective. Findings from this study indicate that infants were tracking the observed agent's comprehension during the interaction. In addition, I suggest that these results also demonstrate successful multisensory integration of visual and auditory information for representing and tracking others' beliefs. The advantage of using neuroimaging in this paradigm lies in the fact that the authors were able to show differing neural activity in response to beliefs induced through multisensory inputs, thus reducing the possibility of multiple interpretations.

Finally, in an attempt to challenge the view that infants' success at false-belief tasks may be a result of the "search where you last saw" behaviour rule, Träuble et al. (2010), included a "manual-control" condition to the classical false-belief task. This extra condition was reportedly similar to the false-belief condition, as its aim was to cause the agent in the scene to have a false-belief about the location of an object. In contrast to classical false-belief paradigms however, the "manual-control" condition was created by the agent herself, who used her hands to change the location of the ball. Crucially, the agent did not have visual access to the scene (i.e. she turned her back away from the scene). Given that the agent knew the location of the ball, one would expect that she held a true belief about where the ball was. Indeed, infants' looking time indicated an attribution of true belief to the agent, even though she had no visual access during the change of location. Based on these results, the authors conclude that infants could flexibly use different types of information, i.e. visual and tactile, to determine someone's beliefs. The results of this paper are compelling as they evidence infants' ability to maintain others' belief representations across sensory modalities.

Overall, although the authors of these studies did not interpret their findings in light of multisensory integration for ToM, I believe they provide compelling evidence of infants' ability to represent and track others' beliefs based on information obtained from multiple sensory systems, even when contrasting their own representations.

Furthermore, the study by Forgács et al. (2019) showed the advantage of relying on complementary methodologies such as neuroimaging, to measure infants' ToM ability during false-belief tasks.

Future studies directly addressing the mechanisms behind infants' behaviour in multisensory false-belief tasks are warranted, as well as the application of the neuroimaging and computational modelling methodologies to such tasks. Nonetheless, I believe that I have here delineated a viable approach to study ToM emergence and infants' performance in false-belief tasks from a new perspective.

3. Experimental Contributions

3.1 New multisensory false-belief paradigm for infant ToM assessment

Introduction

In the present study I introduce an innovative multisensory false-belief task, which was specifically developed for this research. This study was part of a bigger project aimed at investigating multisensory integration for ToM in infants using a laboratory setting, neuroimaging and different infant stimulation (see COVID-19 statement); however, given the event of the pandemic, this study was implemented online and slightly revisited. This newly developed false-belief task varied from most typical false-belief tasks according to several factors. I will now briefly describe the novelty of this task and outline my motivation to conduct this study, as well as my research questions.

Online study

The present false-belief task was the first of this sort to be conducted online using Lookit, i.e. an online platform for child research participation created by MIT Early Childhood Cognition Lab (https://lookit.mit.edu/). Given the online nature of this study, additional experimental procedures were implemented to replicate as faithfully as possible the procedures normally used in a laboratory setting. Briefly, this study used an infant-controlled procedure in an attempt to guide trial presentation based on infants' attention at each trial, following previous research (e.g. Phillips et al., 2002). However, this was the first study to implement this procedure online and with the help of parents who live-coded their infant's looking behaviours (see methods section for more details). Ultimately, while previous research (e.g. Scott & Schulz, 2017;

Semmelmann et al., 2017; Tran et al., 2017) supported the feasibility of online developmental studies, it was essential to determine the feasibility of this procedure.

Second, most VOE false-belief studies conducted in a laboratory setting involve live stimuli presentation to infants, e.g. a physical agent in the same room as the infant engaging in different scenarios (e.g. Onishi & Baillargeon, 2005; Song et al., 2008; Song & Baillargeon, 2008; Träuble et al., 2010). However, given the online nature of this study, the stimuli were recorded and presented to infants through a computer screen. While this approach has been previously used for AL false-belief studies, this was only implemented in two studies in the literature relying on VOE which did not evidence, or only partially evidenced, infants' early false-belief understanding (see Barone and Gomila (2021) for a summary of live vs videotaped false-belief studies). Considering that this factor may have an impact on infants' attention and engagement throughout the task, a feasibility test was needed.

Third, this study used webcam recording of infants' looking behaviour for offline coding, which has only recently become a more widespread and successful tool to measure infants' looking behaviour from the comfort of their homes (e.g. Scott & Schulz, 2017; Semmelmann et al., 2017; Tran et al., 2017). Considering the novelty of webcam recording as a tool for developmental studies, some limitations remain for accurate coding of infants' behaviour. For example, it remains challenging to determine when the infant is looking away from the screen, considering that e.g. there are no indications of the size of participants' screens. I sought to determine whether this online procedure would be appropriate for an accurate analysis of infants' looking behaviour in this specific task by adding an initial calibration, as one would do in a laboratory setting. Briefly, the calibration consisted in presenting infants with stimuli displayed at the left, centre, and right sides of the computer screen, while recording their looking behaviour. The recorded video of infants' looking behaviour during calibration was then used to more accurately determine whether infants were looking away from the screen during offline coding.

In the next sections, I will present more specifically the main differences between this study and the classical false-belief studies present in the literature. These will in turn highlight the features of this study that needed assessing for feasibility.

Multisensory integration

The newly introduced false-belief task involves a new scenario in which the agent depicted in the scene acquires a true- or false-belief about the unfolding scene by means of multisensory integration of information (i.e. both visual and tactile events). In other words, VOE in the agent depicted in this task is driven by the integration of multisensory information. Critically, this requires infants to integrate multisensory information from the agent's perspective, in order to successfully complete the task.

This multisensory scenario is in contrast with classical false-belief tasks, and most of such studies in the literature, which generally only rely on a single modality (i.e. vision) for belief induction in the agent observed in a scenario (Beaumont & Sofronoff, 2008). My choice to introduce a multisensory component to this false-belief task is inspired by the discussions in Part 1, chapter 2.5 of this thesis. To briefly summarise these, unisensory false-belief tasks do not fully relate to real-world scenarios, where information is very rarely unimodal and where multisensory integration is required for a coherent and complete representation of what is being perceived (Parker & Robinson, 2018). Therefore, introducing multisensory false-belief tasks may contribute to shed light on early ToM development which can be generalised to the real-world. Furthermore, it has been previously indicated by Dionne-

Dostie et al. (2015) that multisensory integration drives human behaviour in several situations and is beneficial towards understanding and predicting others' behaviours. Given this evidence, it seems clear that the introduction of multisensory integration in false-belief tasks would warrant a better understanding of infants' ability to understand others' beliefs-driven behaviours and the development of this ability in more naturalistic scenarios.

Only a few infant false-belief studies involving multisensory information for the manipulation of the beliefs of the agents depicted in the false-belief scenarios have been previously developed (e.g. Forgács et al., 2019; Scott et al., 2010, 2015; Träuble et al., 2010). While these studies did not directly interpret their results in view of multisensory integration for ToM, I have outlined in Part 1, chapter 2.5 of this thesis the contribution of their results towards this topic. Briefly, these studies involved the integration of either visual and auditory or visual and tactile sensory information for the representation of others' beliefs, and evidenced infants' ability to pass false-belief tasks involving such multisensory integration of sensory information. To contribute to this literature, I aimed to determine the feasibility of my study in achieving "belief induction" for the assessment of infants' multisensory integration ability for others' beliefs representation.

Stimuli

New stimuli were specifically created for this false-belief study. Therefore, their feasibility in generating a true- and false-belief condition, as well as their ability to provide access to infants' false-belief understanding, required addressing. As mentioned above, this study involved both visual and tactile stimulation provided to the agent depicted in the false-belief scenarios, which is in contrast with classical false-belief studies. However, the main novelty of this false-belief task, which distinguished this study also from that by Träuble et al. (2010) (which involved visual and tactile events as well), concerned the critical event for measuring infants' looking behaviour in response to the agent's false-belief. Specifically, the present false-belief study measured infants' ability to track and predict others' beliefs formed through multisensory integration of visual and tactile events, rather than infants' response to the agent's belief-driven behaviour (as in classical false-belief studies, including Träuble et al., 2010).

More in detail, classical false-belief studies relying on VOE as a measure of infants' belief understanding assess infants' looking in response to the agent behaving in a way that is inconsistent with her beliefs (e.g. performing an explicit inconsistent action, or reacting through a facial or verbal expression of shock) (e.g. Luo, 2011; Onishi & Baillargeon, 2005; Scott et al., 2010; Song & Baillargeon, 2008; Träuble et al., 2010). In contrast, this study did not measure infants' VOE in response to the agent's behavioural response event. Instead, it assessed infants' ability to take the agent's perspective by showing VOE in response to a tactile event, while the agent's visual access to the unfolding scene was occluded. In other words, the agent in this false-belief study acquired expectations based on visual information, which were then at times met through tactile information and at times violated when tactile stimulation was not provided. Therefore, this study measured infants' looking in response to such tactile event.

Finally, the present false-belief study differed from most studies of this sort with respect to the manipulation of the visual access of the agent in the scene, which caused a false-belief in the agent. Specifically, in most false-belief studies, agent's visual access is temporarily occluded with the agent (a) turning around and away from the scene (e.g. Senju, 2012; Träuble et al., 2010), (b) leaving the room (e.g. Scott, 2017; Scott et al., 2010), (c) or introducing a curtain (or another object) between the agent and the unfolding scene (e.g. He et al., 2011; Senju et al., 2011). In this false-belief study, I instead resorted to the agent looking up to the ceiling (moving up her head) for manipulating her visual access to the unfolding scene. Therefore, I deemed it necessary to determine whether this manipulation would indeed be associated by infants with a change in the agent's visual access to the scene, meaning whether infants would realise that the agent could not see the unfolding scene when looking up.

Paradigm

Classical VOE false-belief studies present a habituation phase, during which infants are repeatedly exposed to the same sequence of events to induce context-dependent expectations (Semmelmann et al., 2017). While this habituation technique has been used in previous false-belief studies (e.g. Onishi & Baillargeon, 2005), this was the first study of this sort to implement this paradigm online. Indeed, considering that habituation is generally driven by infants' decline in looking time in response to repeated habituation trials (and related stimuli), infants' looking needs to be coded real-life and habituation is terminated once the desired decline in looking time is reached. Given the online nature of this study, this real-life coding and control of trials was not possible. However, to mould the repetition of trials on a participant-basis and avoid disengaging participants' interest following the repetition of too many habituation trials (Phillips et al., 2002), I implemented the infant-controlled habituation procedure by asking parents to help live-code their infants' looking behaviour by pressing keys on their keyboard when their infant was not looking, which would guide the repetition of trials.

Furthermore, this false-belief study paradigm had an additional distinguishing feature that is not often used in VOE false-belief tasks involving habituation. Specifically, classical VOE false-belief studies present a habituation phase followed by a single test event which is either consistent or inconsistent with the habituation condition (e.g. Onishi & Baillargeon, 2005). In contrast, this newly introduced false-belief task relied on habituating infants to a certain condition and then comparing infants' looking to novel vs familiarised stimuli. Thus, following habituation, two novel conditions were presented, one consistent (familiar) and the other inconsistent (novel) with the habituation condition. This approach has been previously used in other studies supporting infants' ability to understand intentional actions (e.g. Phillips et al., 2002; Sodian & Thoermer, 2004; Thoermer et al., 2012; Wellman et al., 2004).

Research questions

In the present study, I aimed to investigate 18-month-old infants' false-belief reasoning using a new false-belief paradigm for online testing to further our knowledge of the development of human ToM. Specifically, the research questions included: (1) Do infants show habituation to an observed touch event in this new false-belief paradigm? (2) Do infants show differing looking times between test conditions (consistent true vs inconsistent false belief), thus supporting early belief understanding and multisensory integration for ToM? (3) Do infants show dishabituation when presented with the test trials and does the degree of dishabituation vary between consistent and inconsistent test conditions?

Methods

Participants

Eighteen infants between 17.5 and 18.5 months of age (10 F, M_{age} = 543.3 days, SD_{age} = 9.7 days) were recruited from the Essex Babylab database, the Lookit database, as well as social media. An a-posteriori power analysis on G*Power 3.1 identified a power of 0.85 with a large effect size. Additional 26 infants were tested but excluded due to one or more of the following reasons: (a) Participating infants were premature (< 37 months of gestational age; see (Emberson et al., 2017) (N = 1); (b) Data from the session could not be successfully retrieved in full from the Lookit platform (N = 2); (c) Participating infants were too fussy or distracted (N = 1); (d) Participating infants did not complete the habituation (4 trials) and at least 3 test trials for each belief condition (3 consistent- and 3 inconsistent-belief trials) (N = 23). Reasons for considering a trial not complete included poor video quality, unreliable looking behaviour, infants looking away from the screen for longer than 2 consecutive seconds and/or missing the critical touch event, and parents terminating the trial prematurely using infant-controlled live coding (more details below). Ethical approval was granted by the University of Essex Ethics Sub-Committee (ETH2021-0078).

Stimuli and design

This task was adapted from the classical false-belief task created by Wimmer and Perner (1983) to enable the study of implicit ToM through looking behavioural measures. In this version of the task, infants were shown videos of an agent with either a true or a false belief with regards to a touch event. Specifically, the experiment consisted of 4 habituation trials and 10 test trials, which were presented in a quasirandomised counterbalanced order. The test trials presented two novel conditions (i.e. change in agent’s visual access with respect to habituation) which differed regarding the independent variable, i.e. *belief type*. In half of the test trials, the agent had a true belief about the situation. Specifically, although the agent was not able to see the unfolding scene, the touch event was consistent with that observed during the habituation. In the remaining half of test trials, the agent had a false belief about the situation. Here, the touch event was inconsistent with that observed during the habituation and the agent was not able to see this change (see Figure 2 below). Infants’ looking behaviour (i.e. fixation duration) was measured for each given condition. This dependent variable was collected through webcam recording (using the participant’s computer webcam) on the Lookit platform. This design is in line with a previous study utilising the same revised visual habituation, preference-for-novelty technique to investigate 14-month-olds’ looking times in response to consistent vs inconsistent test conditions (Phillips et al., 2002). The study design was implemented on the Lookit platform through the Lookit experiment runner, which uses a JSON object specifying all stimuli and presentation parameters (Scott & Schulz, 2017). A text file of the JSON for this experiment can be obtained upon request, while the experiment can be previewed by a logged-in user at Lookit at https://lookit.mit.edu/studies/be28f7e1-8b4e-4186-904c-905b3d963d76/.

Habituation event

As shown in Figure 2, infants were presented with a four-phase video which depicted a female agent sitting at a table with her hands placed on the table and palms facing upwards. In the first phase of the video, two identical paintbrushes held from two hands appeared from both sides of the screen and moved towards the agent’s hands, who watched the scene unfold. This first phase always took 3.5 s. In the second phase of the video, the hands holding the paintbrushes paused when they were mid-way to reaching the hands on the table, and the agent looked straight to the infant while saying "Look" to capture their attention. Successively, the paintbrushes continued their trajectory towards the agent's hands. This second phase always lasted 2.5 s. In the habituation trials (Figure 2A), the third phase of the video sees the paintbrushes touching the agent's hands for 2 s (critical touch event), who continued to watch the scene unfold. Finally, in the fourth phase of the habituation trials, the video froze at the critical touch event for up to 20 s. Looking times were recorded to compute infants' attention from the third phase onwards.

Test events

Following habituation, infants were shown a total of 10 test trials, which consisted in two types of events: consistent and inconsistent events (5 each, respectively). The test events were presented in a quasirandomised order, meaning that half the participants viewed order A and the other half viewed the same events but in reversed order (B). The test event presented first was counterbalanced across infants, and the presence of an effect of order presentation on infants' looking times at the test trials was investigated. While each test trial followed the same four-phase structure of the habituation trials, the content of the second phase varied (as shown in Figures 1B and 1C), although its original duration was maintained across conditions. Specifically, both the consistent and inconsistent trials differed from the habituation trials in that the agent in the second phase looked away from the unfolding scene and stared at the ceiling after saying "Look". Furthermore, while in the consistent trial the paintbrushes moved again to touch the agent's hands, in the inconsistent trial they moved again to touch the surface nearby the agent's hands instead. As a result, the consistent events showed a touch that was consistent with that observed during the habituation trials, which lead to the agent having a true belief about the unfolding scene, even though she had no visual access to the scene. In contrast, the inconsistent events showed a touch that was inconsistent to that observed during the habituation trials, thus resulting in the agent having a false belief about the unfolding scene as she was not able to observe the change of direction of the paintbrushes. Therefore, inconsistent events represented a VOE in the agent, whose expectation of being touched by the paintbrushes was violated. This should have in turn resulted in longer infants' looking if they took the agent's perspective (as opposed to their own). As in the habituation, infants' attention was measured through their looking time recorded in the third phase of the video onwards.

Procedure

For this online study, parents and infants participated from their homes and there was no experimenter present. To start the study, participants accessed the Lookit platform, where the study was hosted. Given the online nature of this study, parents were asked to help live coding their baby's looking behaviour by pressing the 'Space' key on their keyboard while their baby looked away from the screen. Parents were specifically asked not to influence their babies' behaviour. This live coding procedure was included in this study to guide the repetition of trials during habituation and testing based on infants' attention at each trial (as one would do in a laboratory setting). Furthermore, additional checks were made in an attempt to control the experimental setting. Prior to the start of the study, parents completed 3 practice coding trials to ensure that they understood the coding procedure. Although parents were instructed not to have their infants present during their practice trials in order not to affect this habituation paradigm, 7 out of 18 infants were present during the practice trials. Therefore, the potential effect of infants' presence during practice trials on their looking times in the test trials was investigated through statistical analysis. Once parents completed their practice trials, they were asked to position their baby on their laps in front of the computer screen and to allow their baby to see the stimuli on the screen. Parents conducted a video quality check, to ensure that (a) their baby would be at the centre of the screen and visible for future offline coding, (b) there would be good lighting during video recording for seeing the baby's eyes clearly, (c) the webcam was working correctly, and (d) distractions in the environment (e.g. dogs, other monitors, siblings) would be minimised. Once parents and babies were ready to start, the calibration was performed, which was followed by the presentation of the habituation and testing stimuli on the computer screen.

Parents were able to pause the study if their baby became too fussy or if they needed to attend to something else momentarily, during which an infant-friendly video was played. When parents pressed the 'Space' key on their keyboard to code their baby's looking, an infant friendly sound was also played to catch back the baby's attention. A trial (either habituation or test) started with the start of the trial's video. A given trial would be considered invalid if infants missed or if parents terminated the trial before the critical touch event. Each trial terminated if (a) the infant looked away from the screen for more than 2 consecutive seconds, (b) the parent pressed the 'Space' key on their keyboard for more than 2 consecutive seconds, or (c) the video itself terminated after a period of 28 seconds (maximum length of trials). This infant-controlled procedure was implemented following previous studies (e.g. Phillips et al., 2002). Infants' looking behaviour was also coded offline by two independent coders (one of which was blind to the experiment hypotheses) to ensure that trials were only included in the analyses if they met the inclusion criteria outlined above. Audio consent was obtained from all included participants prior to the start of the study through the Lookit platform.

Reliability

Reliability for observation of infants' looking behaviour was calculated using the inter-rated reliability script for the Datavyu software (Datavyu Team, 2014), which yielded a percentage agreement of 81%.

Statistical analysis

To determine the existence of an effect of (a) infants' presence during practice and (b) order of presentation of the test event on infants' looking times during test trials, as well as a (c) difference in looking times across the two event conditions, a 2 x 2 x 2 mixed model analysis of variance (ANOVA) was conducted. Specifically, Practice (baby present or absent) and Order of presentation (hand or surface trial first) were included as the between-subject factor and Event (consistent or inconsistent) as the within-subject factor. Furthermore, paired-samples *t* tests were conducted to determine whether (a) significant habituation was achieved and (b) infants' looking behaviour differed significantly with respect to the independent variable, i.e. belief type, when comparing only the first trial of any given condition. This analysis was conducted given the adaptation of the "visual habituation, preference for consistent vs inconsistent condition" paradigm to false-belief tasks. Specifically, although collapsing looking times across trials is usually performed in this paradigm to stabilise variances (e.g. Phillips et al., 2002), false-belief studies are generally interested in the primary reaction to stimuli, rather than a stabilised reaction to repeated stimuli. Finally, an exploratory analysis on infants' differences in recovery from habituation between the consistent vs inconsistent test conditions was conducted using an independent-samples *t* tests. Furthermore, specific dishabituation in both conditions was calculated through paired-samples *t* tests.

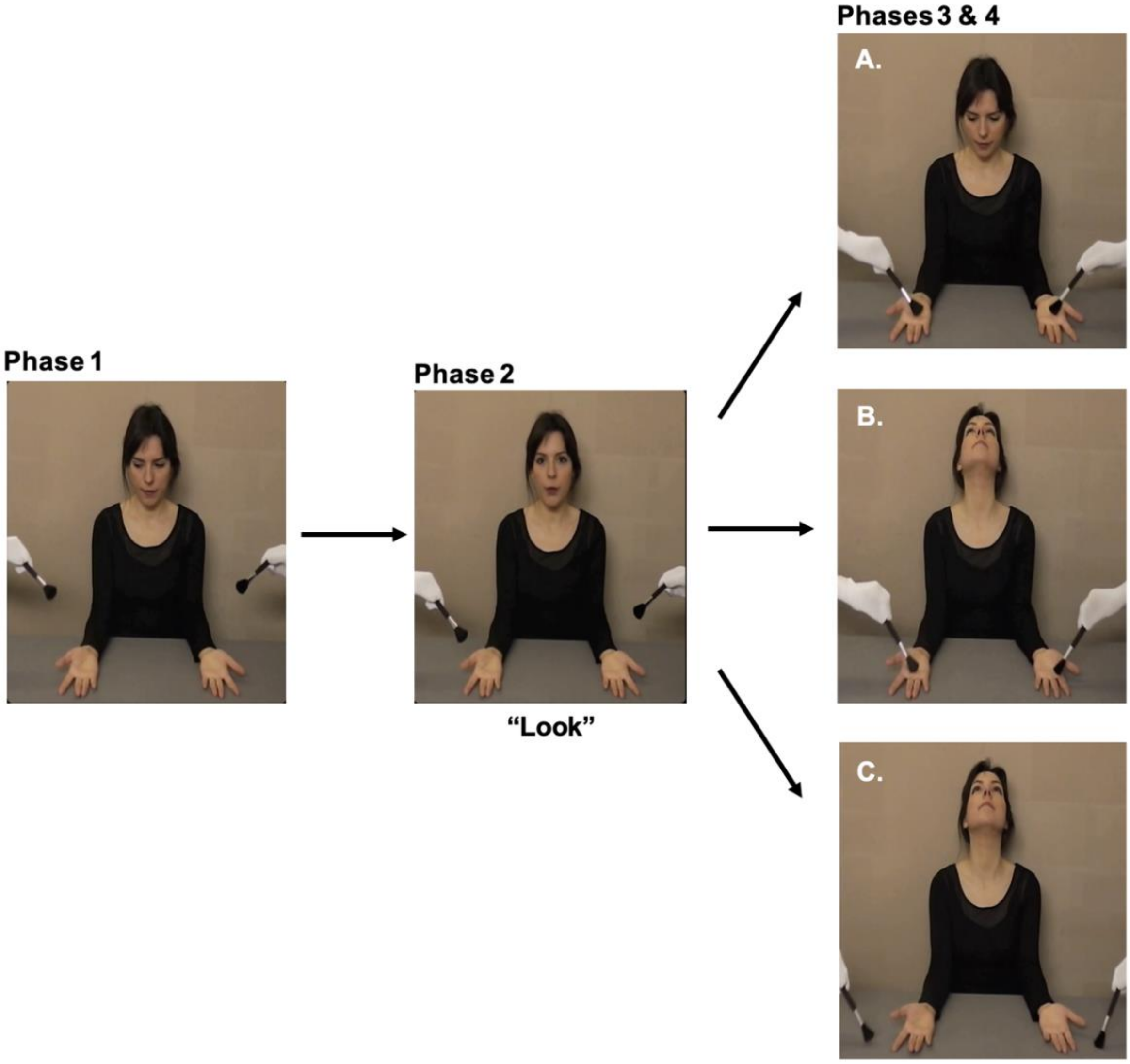


Figure 2. Screenshots of video trials from false-belief task showing the trial phases and the differing conditions (A, B, C), as seen from the infant's perspective. (A) Habituation: hand touch, visual access; (B) True-Belief: consistent hand touch, no visual access; and (C) False-Belief: inconsistent surface touch, no visual access.

Results

Habituation

Overall, infants included in this study successfully habituated to the condition presented in the habituation trials. Specifically, infants looked significantly longer at the first two habituation trials (M = 10464 ms, SD = 2471) compared to the last 2 habituation trials (M = 6334 ms, SD = 3023), t(17) = 5.186, *p* < .001.

Test (consistent vs inconsistent)

My main analysis focused on the comparison of infants' looking between consistent and inconsistent test events. Infants' looking behaviour in this study did not significantly differ between Events, F(1, 14) = .018, *p* = .896, ηp^2 = .001, with similar looking times observed between the consistent hand (M = 6602 ms, SD = 785) and inconsistent surface (M = 6705 ms, SD = 506) conditions. See Figure 3 below for a visualisation of this data. Furthermore, there was no interaction between Event (consistent vs inconsistent) and infants' presence during Practice, F(1, 14) = 2.111, *p* = .168, ηp^2 = .131, or Order of presentation of test trials, F(1, 14) = 2.019, *p* = .177, ηp^2 = .126. A three-way interaction was also not found, F(1, 14) = .049, *p* = .828, ηp^2 = .004. Overall, these results indicate that both being present at the practice trials or which test trial (whether consistent or inconsistent) was presented first to the infant did not affect the study results. For this reason, I maintained for the analysis both participants present and absent at the practice trials and collapsed data from the two orders of presentation of test trials.

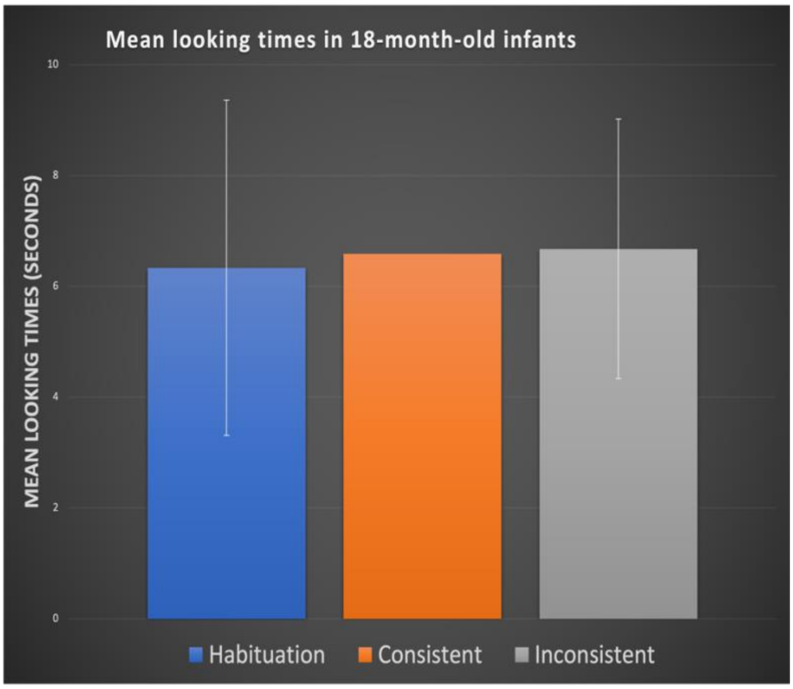


Figure 3. Mean looking times at habituation vs consistent vs inconsistent test trials in 18-month-old infants. Error bars indicate standard deviation.

Inspection of a trial-by-trial basis showed that infants looked longer (although this did not reach significance) at the hand consistent condition vs surface inconsistent condition in the first trial, while this relationship was inverted in the following two test trials. Overall, looking times for the consistent condition tended to decline from the first to the third test trial, presumably as infants gained familiarity with the test events themselves. In contrast, the same decline was not that evident in the inconsistent condition, possibly indicating that this condition was considered more novel and required more trials for familiarisation (see Figure 4 for a visualisation of the looking time decline among infants with respect to the consistent vs inconsistent condition). A paired samples *t* tests assessing infants' looking times at the first consistent (M = 8764 ms, SD = 4003) vs first inconsistent (M = 6885 ms, SD = 4271) test trial found no significant difference between the two conditions, t(17) = 1.249; *p* = .229.

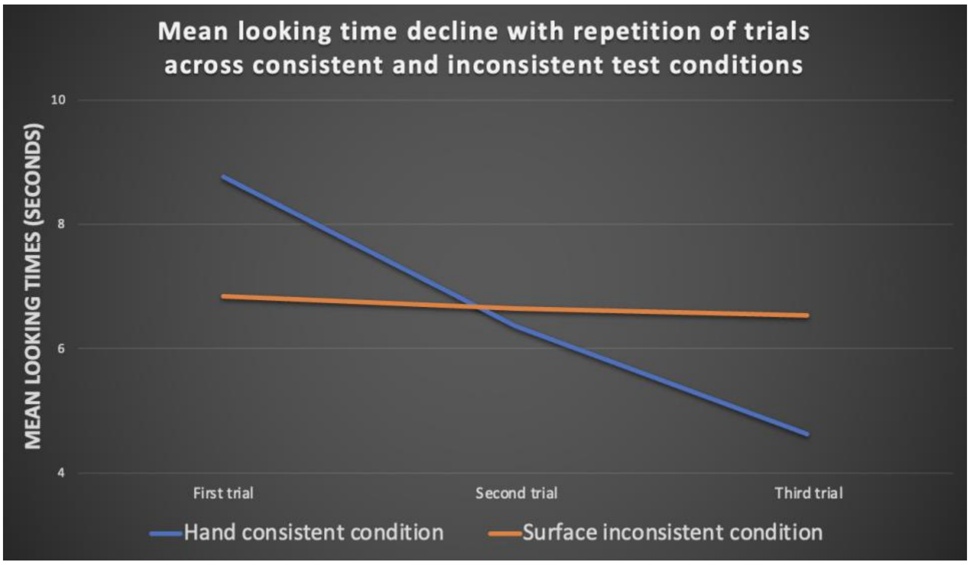


Figure 4. Mean looking time decline across three test trials in consistent vs inconsistent conditions in 18-month-old infants.

Exploratory analysis: Dishabituation

Dishabituation looking times were calculated as the mean looking time in the test phase minus the mean looking time in the last two habituation trials (Sodian & Thoermer, 2004). An independent-samples *t* tests performed over the dishabituation times showed that the consistent (M = 567 ms, SD = 3031) and inconsistent (M = 1147 ms, SD = 3129) test groups' looking times were similar, t(16) = .391, *p* = .701. Furthermore, a paired-sample *t* tests showed an absence of significant dishabituation for the consistent hand condition group. Specifically, looking times were similar between the last two habituation trials of participants who were exposed to the hand test condition first (M = 6090 ms, SD = 2948) and the first trial of the consistent hand condition (M = 6657 ms, SD = 2794), t(10) = .621, *p* = .549. Similarly, a paired-sample *t* tests showed an absence of significant dishabituation for the inconsistent surface condition group, although a trend to significance was observed. Specifically, looking times were similar between the last two habituation trials of participants who were exposed to the surface test condition first (M = 6719 ms, SD = 3334) and the first trial of the inconsistent surface condition (M = 10029 ms, SD = 4846), t(6) = 2.055, *p* = .086. See Figure 5 below for a visualisation of this data.

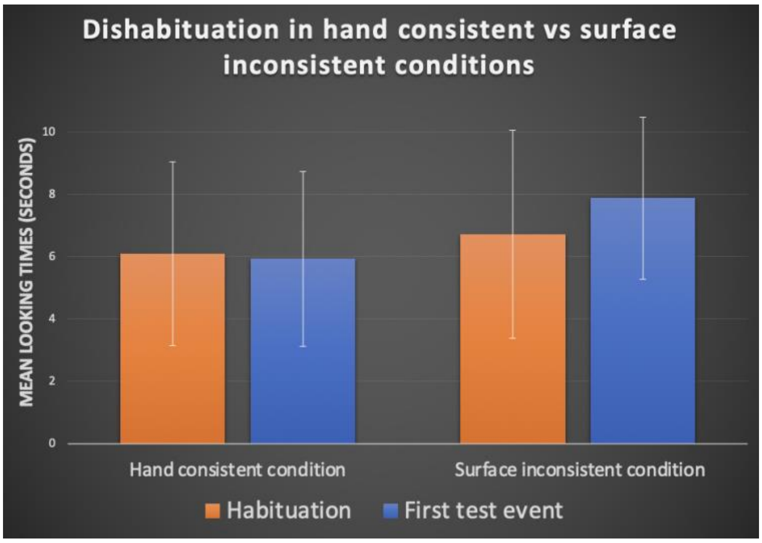


Figure 5. Failed dishabituation in the consistent vs inconsistent test conditions in 18-month-old infants. Mean looking time decline from habituation to first consistent vs inconsistent test trial. Error bars indicate standard deviation.

Discussion

In this online study, 18-month-old infants were assessed in a new multisensory false-belief task to examine ToM emergence and the role of multisensory integration for its development. The feasibility of the newly developed false-belief task was also assessed for follow-up research. Overall, results from this study show that infants did not display a differential looking time to consistent vs inconsistent test events. Thus, our study does not evidence an early ToM ability in 18-month-old infants. However, the online implementation and/or the newly introduced paradigm may have influenced the results. In the following paragraphs, I will discuss the results in light of the feasibility of the newly introduced false-belief task.

My results suggest that infants successfully habituated to the stimuli in the task. This was evidenced by a significant decline in looking time between the first two and last two habituation trials. Therefore, my stimuli and procedures seemed to be effective in familiarising infants with the touch event depicted. These results also support a previous study suggesting that 4 habituation trials are sufficient to induce habituation in infants older than 14 months of age (Phillips et al., 2002), as opposed to 6 habituation trials usually used in younger infants (e.g. Sodian & Thoermer, 2004; Thoermer et al., 2012; Wellman et al., 2004). Nonetheless, a 50% decline in looking time, which is considered as the “industry standard” (Aslin, 2007) for habituation paradigms, was not observed in this study. While I consider the significant decline observed in this study to be sufficient to determine habituation in this population sample, future studies are warranted to address whether an additional number of habituation trials would indeed be necessary to reach such a decline and whether this would impact results. Finally, a total of 18 participants (~95%) completed all the habituation trials using the infant-controlled habituation trialled in this online study, suggesting that this approach can be implemented online without impacting recruitment or data quality.

In this study, I did not find that 18-month-old infants were able to engage in ToM. These results seem to contrast previous findings showing infants’ ToM ability from an earlier age (as young as 6 months) (e.g. Hyde et al., 2018; Onishi & Baillargeon, 2005; Southgate & Vernetti, 2014). However, some of the characteristics of my newly developed task may have impacted the results. I will now discuss these characteristics in more detail, in an attempt to inspire future studies to validate my findings and shed light onto the feasibility of the new task here introduced.

First, as mentioned earlier, I did not observe a 50% decline in looking times following habituation. Therefore, I cannot exclude that this may have impacted infants' looking times during test trials and contributed to the lack of statistical significance between conditions. Future studies addressing this point are warranted to shed light on the importance of achieving such "industry standard" (Aslin, 2007) in this type of habituation paradigm and how the percentage of decline may change throughout development.

Second, this false-belief task involved multisensory integration for belief induction in the observed agent, which may have represented a too complex set of stimuli to fully process by infants during habituation. More specifically, according to Aslin (2007) infants may meet the habituation criterion, nonetheless they may not fully encode the stimuli when these are too complex or diverse. Such "less than full encoding" may in turn result in (a) infants seeking familiar, as opposed to novel, stimuli in the test trials, or (b) trigger a recognition response for familiarity that is stronger than the attentional response to novelty. This may be in part supported by the fact that ~67% of the infants included in this study looked longer at the consistent vs inconsistent condition. Specifically, infants on average only looked longer at the consistent vs inconsistent condition in the first trial, while this relationship was inverted in the following two trials. The higher looking times in the first consistent, as opposed to inconsistent, test trial may be interpreted in line with the above-mentioned suggestion and indicate the occurrence of a "less than full encoding" phenomenon in this study. Similarly, the higher looking times in the second and third inconsistent, as opposed to consistent, test trials may also be explained in light of this encoding phenomenon. Specifically, infants may have been able to appreciate the novelty of the test trials once they had fully encoded the stimuli presented in such inconsistent test trials. Indeed, infants had to fully encode that the stimuli differed on *two* factors (agent's visual access and touch event) from the habituation trials, as opposed to the consistent test trials which differed only on *one* factor (agent's visual access) from the habituation trials. As opposed to classical false-belief paradigms, this task involved a multisensory component, which may have increased the complexity of the stimuli. However, a few studies exist in the literature implementing multisensory integration in infant false-belief task and evidencing early ToM ability (e.g. Forgács et al., 2019; Scott et al., 2010, 2015; Träuble et al., 2010). Furthermore, previous research has described multisensory integration as beneficial for understanding and predicting others' behaviours (Dionne-Dostie et al., 2015). The present study thus highlights further the need of researching multisensory integration for ToM development and for the interpretation of false-belief task results.

Third, infants' looking time in response to both consistent and inconsistent events suggests that the two conditions were considered similar by infants. At a first glance, this may indicate that infants were not able to distinguish the variations in the touch event between conditions (i.e. touch hand or touch surface), thus consequently not differentiating the true- and false-belief conditions. However, Aslin (2007) argued that infants may not show any spontaneous preference for one of two test conditions when these are similar to the habituation condition, despite they may be able to discriminate between the two. This is attributed to a loss of saliency following repetition. Furthermore, the trial-by-trial analysis indicated only the looking times associated with the consistent test condition to decline across trials. This result is not in accordance with previous studies, e.g. Phillips et al. (2002), who saw a decline in looking time for both consistent and inconsistent conditions. This result may therefore suggest that infants indeed could distinguish between the two test conditions but that such an effect could not be captured with the small sample size included in this study. Overall, I therefore cannot conclusively determine whether infants were able to distinguish the two conditions, and thus whether false belief was induced by the stimuli developed for this study. Future research utilising this task with complementary measures of looking behaviour, such as neuroimaging, could help interpret the results by narrowing the range of possible interpretations. For example, assessing ERPs associated with saliency of stimuli (e.g. Simons et al., 2001) in this task would help e.g. in determining whether infants distinguished the two test conditions and if they found them novel compared to the habituation condition. Furthermore, it would be interesting to conduct experiments using this task and addressing multisensory integration from a first person (infant) perspective, e.g. by providing such multisensory stimulation to the agent and to the infant as well. In particular, this would help determine whether the failure of false belief induction in this study was a result of the task itself or whether it was driven by infants' inability to process multisensory integration-driven beliefs from another person's perspective. In addition, given the small sample size of this study, future investigations with a bigger population are warranted to validate these results.

Fourth, there are other reasons that may have impacted my findings, which are closely related to the online nature of the study. Specifically, the fact that my stimuli were recorded, rather than shown live as in classical VOE false-belief tasks, may have resulted in infants' lower engagement and attention. This is in concordance with previous studies using recorded stimuli for VOE which did not evidence (or only partly evidenced) ToM ability in infants (Surian et al., 2007; Yott & Poulin-Dubois, 2012).

Furthermore, the paradigm used in this task (i.e. visual habituation, preference-for-novelty technique with consistent and inconsistent test conditions) is not usually employed in false-belief studies, which instead tend to rely on only one new test condition. Follow-up studies controlling some of these factors or utilising different methodologies (e.g. neuroimaging or computational modelling) may shed light on the matter. Furthermore, my study focused on a critical event which differed from classical false-belief tasks. Specifically, this study did not measure infants' VOE in response to the agent's behavioural response event but required instead infants to take the agent's perspective. Therefore, I investigated whether infants showed VOE in response to the observed tactile event, while the agent's visual access to the unfolding scene was impaired. This difference from classical or previous multisensory false-belief studies may suggest that this type of VOE may have more complex processing requirements or underlie cognitive mechanisms developing at a later stage. Future neuroimaging studies can address this question by investigating infants' expectations of a tactile event. Lastly, in this false-belief study, I resorted to a new way of manipulating the agent's visual access to the unfolding scene, i.e. by looking up to the ceiling. Previous studies have shown that, by 18 months of age, infants are able to comprehend others' eye and head movement signals (e.g. Butterworth & Grover, 1989). Based on this, I expected my implemented manipulation to be successful. However, results indicate that this form of manipulation may not be strong enough to generate a change in the agent's visual access as seen from the infant's perspective. Future studies addressing this question by directly contrasting this occlusion with the ones typically used in false-belief tasks are warranted to shed light onto the feasibility of this manipulation.

With regards to the exploratory analysis on dishabituation, the results support my previous findings and discussions. Indeed, they indicate no significant difference in dishabituation between the consistent and inconsistent conditions, as well as between habituation looking times and those obtained during the two test conditions separately. Specifically, infants seem (a) not to attribute novelty to the stimuli presented in either of the two test conditions following habituation (thus not recognising the change in the agent's visual access); and (b) not to find the test conditions different from each other (thus not recognising the change in agent's belief type). Nonetheless, a trend towards significance was observed for the inconsistent surface condition vs habituation, possibly highlighting that this condition may indeed have been considered by infants more novel than the consistent condition. Future studies with a bigger sample size are needed to validate findings.

Finally, some last remarks with regards to the feasibility of this study will be outlined. Neither order of presentation of test events nor infants' presence during practice influenced infants' looking behaviour during the test trials. These results are compelling as they show that the effect evidenced in this study is sufficiently robust to survive the introduction of such variability. Nonetheless, a trend towards significance was observed for order of presentation of test events. Given the small sample size, further studies with a bigger population are warranted to elucidate the matter. With regards to the absence of an effect of practice on infants' looking time, I would like to highlight that this result supports the feasibility of conducting this study online. Indeed, some of the limitations associated with such an online approach identified in the introduction, e.g. parents not following the instructions and having their baby present during the practice trials, can be overcome. Webcam recordings and online calibration did not present a major limitation for coding infants' behaviour offline, considering that an inter-rater reliability of 81% was found in this study. Nevertheless, the online setting of this study may have negatively influenced inter-rater reliability, given that previous false-belief studies conducted with 18-month-olds in a laboratory setting reported inter-rater reliabilities of 99% (Scott et al., 2010) and 100% (Yott and Poulin-Dubois, 2012). In addition, while some participants were excluded due to technical issues, infant fussiness or distractions in the surrounding environment, the online setting of this study did not impact data quality. However, it did render participant recruitment challenging and its online implementation (e.g. recorded stimuli, online infant-controlled habituation) may have impacted my study results. Finally, the benefit of the newly introduced infant-controlled procedure for the repetition of trials guided by parents live-coding their babies' looking behaviour was mixed. Specifically, out of 18 parents who participated in this study, 16 live-coded their babies' looking behaviour, which helped the repetition of trials when their babies missed the critical touch event (12 parents) or the termination of the trial once the baby had looked away from the screen for more than 2 consecutive seconds. While parents' live-coding was not accurate enough to be considered as the sole measure of looking time in a given trial, their efforts contributed to a quite successful implementation of an online infant-controlled procedure of the task. Future studies are warranted to find ways to further improve parents' live coding and infant-controlled procedures for online habituation paradigms.

Conclusion

To conclude, the online nature of this study did not seem to impact the implementation of the paradigm and procedures, or the quality of my data. However, it may have influenced the results. Some questions remain regarding the interpretation of my findings. Specifically, whether infants have the ability to understand others' beliefs induced by the multisensory integration of information could not be concluded in this study. Indeed, whether this study’s findings were driven by its online implementation and/or the newly introduced paradigm or by an absence of infants’ ToM ability at this young age remains undetermined at this stage. However, I believe to have shown ways to address this in future research and contribute to the psychology debate on ToM emergence and multisensory integration for ToM development in infants.

Part 2

On Theory of Mind: individuals with limb difference

1. Background

As discussed previously in this thesis, the mechanisms behind ToM and its development remain to be determined. This is valid both from a biological point of view, as well as a computational point of view. Furthermore, the factors likely to be important for the development of ToM ability are yet to be outlined. Decades of research have focused on infants' success in false-belief (or other ToM-related) tasks to tackle these unanswered questions. I propose in this thesis that the pioneering exploration of ToM in the limb difference population can shed light into the mechanisms underlying ToM emergence in a way that has not been done before. I will now briefly provide a definition of the limb difference population and a summary of the literature related to the effect of having a limb difference on brain development and function, as well as on ToM ability.

1.1 Limb difference defined

The term "limb difference" refers to the partial or complete absence, loss, or malformation of a limb. Specifically, the limb difference community includes an overarching group of people, both children and adults, who were born with a limb deficiency or reduction defect and/or who lost a limb at different stages during their life. This distinction in type of limb difference is also referred to as congenital and acquired limb differences, respectively. Limb differences are often further classified by the level at which they occur, i.e. upper- and/or lower- body extremities. Congenital limb difference can result from a variety of causes, including vascular disruption or malformation, genetic factors, environmental teratogen exposure, as well as unknown causes. In contrast, acquired limb differences are generally associated with trauma and diseases (e.g. vascular diseases, infections, diabetes mellitus) (Ephraim et al., 2003; Le & Scott-Wyard, 2015).

Exact estimates of the number of individuals with limb differences worldwide, both by type and level, remain unknown, although some studies and registries have attempted to investigate these and provided some figures (e.g. Ephraim et al., 2003; Heikkinen et al., 2007; Mai et al., 2019; Vuillermin et al., 2021). For example, following the assessment of congenital limb differences in the USA between 2010 and 2014, Mai et al. (2019) identified a prevalence of 5.27 (95% CI: 5.07, 5.47) per 10,000 births, in a ratio of 2:1 upper to lower limb differences. In contrast, an amputation incidence between 100 and 500 per million people annually was identified in Western countries (Heikkinen et al., 2007) and generally a 1:9 upper to lower ratio is associated with amputations (Ziegler-Graham et al., 2008).

Some of the psychological and functional impact that congenital and acquired limb differences may have on these individuals and their families have been investigated, such as (a) coping strategies following lower limb amputation (Oaksford et al., 2005), (b) relationships between children with congenital upper limb difference and their families (Murray et al., 2007), (c) functional and emotional impact of congenital upper limb differences (Bae et al., 2018), (d) needs and preferences related to prostheses (Stephens-Fripp et al., 2020). For example, Bae et al. (2018) indicated decreased limb function but better peer relationships and more positive emotional states in children with congenital upper limb differences compared to the general population. Stephens-Fripp et al. (2020) reported issues and concerns regarding prostheses among individuals with limb differences, including weight, manipulation and dexterity, aesthetics, sensory feedback and financial cost.

However, relatively few studies have investigated the effect that being born or acquiring, as well as living with a limb difference has on the mechanisms underlying brain development and function. Investigations on such topics include changes in body representation following amputation and its association with pain (Bekrater-Bodmann et al., 2015; Mayer et al., 2008), brain mechanisms underlying others' action comprehension and imitation (Aziz-Zadeh et al., 2012; Cusack et al., 2012), brain organisation (Liu et al., 2020; Striem-Amit et al., 2018; Vannuscorps et al., 2019), brain activity in response to visual presentation of lost limbs following amputation and its relation to phantom limb experience and prostheses use (Chan et al., 2019; Guo et al., 2017; Lyu et al., 2017). For example, Bekrater-Bodmann et al. (2015) identified an association between phantom and residual limb pain and recall of an impaired body representation in dreams. Mayer et al. (2008) instead highlighted a functional adaptation of prostheses to body schemas following amputation, which prevented telescoping (i.e. shortening of the phantom limb). Furthermore, Lyu et al. (2017) conducted a neuroimaging study investigating brain activity related to action performance in individuals with congenital upper limb difference (who perform actions with their feet) vs controls (who use their hands). The authors suggested that activity in frontoparietal association motor areas shows a preference for action type (e.g. reaching or grasping) regardless of the effector used to complete that action.

Overall, the literature suggests that we still know relatively little with regards to the effect that having a limb difference has on cognition and brain functions. There are at least three reasons why researching limb difference is fundamental. First, the numbers of amputation are expected to double by 2050, as a result of the aging population and the high rates of dysvascular conditions (mainly diabetes) seen among older adults (Ziegler-Graham et al., 2008). Second, limb loss has been acknowledged across the globe as a significant public health issue, requiring research-based strategies to improve social outcomes. Third, limb differences affect the health and well-being of people worldwide (Ephraim et al., 2003). As such, research is warranted to bring awareness to these conditions, increase our knowledge of the implications of limb differences on brain development and function, as well as inform interventions and therapies to help and improve the quality of life of people affected and their families.

1.2 Evidence of ToM in People with Limb Difference

To my knowledge, previous research aimed at directly assessing ToM and its development in individuals with limb difference has not yet been conducted. However, two studies (Aziz-Zadeh et al., 2012; Cusack et al., 2012) have provided some insight into the ability of people with limb differences to engage in ToM-related abilities in different situations. Specifically, individuals with limb difference were assessed in tasks involving neuroimaging and either the observation of others' actions (Aziz-Zadeh et al., 2012) or their imitation (Cusack et al., 2012). As a result, both studies provided evidence of neural activation of areas associated with ToM, although such activations were interpreted in relation to processes other than understanding others' mental states (i.e. action observation and imitation).

More in detail, Cusack et al. (2012) investigated the mechanisms behind imitation of people with bodies different from the self. Specifically, prosthesis users with acquired limb difference and controls without limb difference were asked to imitate actions of either other prosthesis users or control agents, while their brain activity was recorded with EEG. Interestingly, prosthesis users with acquired limb difference showed typical brain activation associated with imitation planning (i.e. left parietofrontal activation) after watching the action performed by another prosthesis user. However, additional brain areas associated with mentalising (i.e. right parietal and occipital regions) were recruited after watching actions executed by agents without limb difference. This was in contrast to what was observed in individuals without limb difference, who showed typical neural activity associated with imitation planning in both cases. These results suggest that a mentalising mechanism may be necessary when planning to imitate actions performed by someone with a body different from the self.

Aziz-Zadeh et al. (2012) using neuroimaging investigated instead the ability and the mechanisms by which an individual with congenital limb difference understands the actions of an agent with a radically different body from the self. Specifically, the participant observed an agent without limb difference completing some actions, which either could or could not be executed by the participant with congenital limb difference, while her brain activity was recorded with fMRI. Interestingly, the participant engaged her sensorimotor representations (i.e. activity in mirror regions, including the premotor cortex and the inferior parietal lobe) when observing actions that she could execute herself (even though at times using different effectors). This therefore suggests a simulation approach to goal understanding. In contrast, neural activation of areas associated with mentalizing mechanisms (i.e. precuneus, right TPJ, medial prefrontal cortex (mPFC)) were additionally observed when the participant observed actions that she could not execute herself. This indicated that when trying to understand the goal of a person that has a different body to her own, neural areas beyond the simulation areas needed to be recruited.

Altogether, these two studies suggest mentalising for understanding or imitating goal-directed actions of people with different bodies from the self. However, evidence is yet to be provided to determine whether a similar approach is also transferred to the *inference of others' mental states*. Furthermore, these studies provide some promising initial insights into mentalising ability in people with different bodies from the self. However, further research is required to validate these findings given the small sample sizes of the two above-described studies (6 and 1 individuals with acquired and congenital limb difference, respectively). Furthermore, research comparing different types of limb differences (e.g. congenital vs acquired, or upper- vs lower- limb differences) is also warranted to fully understand the role of embodiment and sensorimotor experience in ToM ability and development, as well as to identify the underlying candidate mechanisms.

2. Theoretical and Methodological Contributions

Why researching people with limb difference for understanding ToM emergence and underlying mechanisms? In the next sections, I will briefly highlight the invaluable role that this population can play for the study of ToM emergence and development, as well as for the assessment of the underlying mechanisms and factors important for successful ToM.

2.1 ToM development and underlying mechanisms

While ToM presence in adulthood is widely accepted and this ability is generally evidenced from 4 years of age, its emergence during the first months of life is currently debated. While studies with infants have provided great insights into ToM development and its underlying mechanisms, there remains some fundamental questions which are challenging to tackle solely relying on infant research, as well as unresolved controversies (see Part 1 of this thesis for a detailed discussion). To complement and advance the findings so far observed, I suggest two main reasons as to why studying the limb difference population can provide insights into ToM emergence and development.

Firstly, given that people with limb difference have different body characteristics compared to the general population, they could help the scientific research identify whether (a) mechanisms (e.g. simulation) that were shown to underly the understanding of others' goals and actions also underly others' mental state inference; (b) ToM requires similarity between the observer (the self) and the person observed (the other); (c) alternative mechanisms exist to successfully engage in ToM towards people different from the self; and (d) different mechanisms are utilised to perform ToM in different situations.

In more detail, previous studies with infants and the general population have proposed different mechanisms, e.g. simulation vs teleological vs separate mentalising theories to underly ToM development and ability; however, it is unclear which (if any) of these mechanisms is critical to ToM. Interesting insights on the matter have already been provided by the two studies described above by Aziz-Zadeh et al. (2012) and Cusack et al. (2012). Specifically, they identified the recruitment of additional brain areas associated with mentalising in individuals with limb difference during the imitation and understanding of goal-directed actions of others who have bodies that differ from the self. These therefore suggested an involvement of differing mechanisms driven by differing embodiment and sensorimotor experiences for understanding others' observable behaviour. Since additional brain areas that go beyond simulation were recruited by participants of these studies, this evidence would (at least partly) contrast proposals of simulation being the (sole) mechanism behind ToM development and ability. This is also in light of simulation requiring self-other similarity. However, it remains debated in the literature whether the simulation mechanism is recruited when interacting with individuals whose bodies differ from the self.

Indeed, previous findings from other studies with the limb difference population suggest motor simulation not to be the mechanism behind various abilities, including visual speech interpretation (Vannuscorps et al., 2021), efficient recognition of facial expression (Vannuscorps et al., 2020), conceptual processing of action verbs (Vannuscorps & Caramazza, 2019), action perception and interpretation (Vannuscorps & Caramazza, 2015, 2016). Nonetheless, it cannot be excluded that simulation may indeed be the mechanism underlying such abilities in the general population. People with limb difference may instead rely on some compensatory mechanisms (e.g. neural systems matching action observation, action execution, and motor imagery (Funk et al., 2005)). As a consequence of this yet open question, the extent to which the simulation mechanism is important for understanding others' mental states (thus for ToM development and ability) remains to be determined.

The existence of compensatory mechanisms in the limb difference population remains under investigation. Aziz-Zadeh et al. (2012) suggested that the activation of the simulation network when observing actions out of someone's motor repertoire may be resulting from observation learning. In line with this view, Price (2006) and Brugger et al. (2000) both proposed simulation to be possible also when interacting with bodies different from the self, through the same observation learning mechanism. More in detail, Price (2006) proposed the acquisition of body images through observation of others, which may in turn allow simulation of "different others" and may be associated with phantom limbs. Similarly, Brugger et al. (2000) suggested habitual observation of other people moving their limbs to contribute to the development of sensorimotor representations of absent and phantom limbs. Finally, Corradi-Dell'Acqua and Tessari (2010) assessed individuals with anomalous anatomical and sensorimotor bodily features in three different visual tasks and indicated a key role of visual experience in building a model of bodies different from the self. Furthermore, the authors suggested this model to mediate the processing of biological stimuli and to operate in parallel, or as an alternative, to the representation of one's own body. The authors also argue that pure embodied accounts, such as the simulation theory in our discussion, should be reconsidered when processing biological stimuli.

Overall, it remains unclear whether people with limb difference use mentalising and/or simulation mechanisms to understand beliefs (which go beyond action observation) of others with bodies different from the self. Assessing this in future studies with this population would be an interesting way to determine whether theories describing goal-directed actions can also explain ToM development or whether a separate mentalising mechanism underlying this ability exists. Alternatively, this research may also show that these mechanisms coexist and are used in different situations (e.g. mentalising may mainly be utilised when inferring mental states of others different from us).

Furthermore, it has not yet been studied whether these mechanisms are specific to the limb difference population or whether they extend to the general population when interacting with people with different bodies from their own, e.g. with people with limb difference. Therefore, involving the limb difference population in scientific research could represent a way to tackle these questions and to study the existence of compensatory mechanisms.

Secondly, given that limb difference can be classified into congenital or acquired (through amputation), the limb difference population allows the investigation of the impact of varying body characteristics throughout development on ToM. Therefore, the above open questions can also be studied from a developmental point of view in a new insightful way. In other words, while assessing an adult population, it is possible to indirectly investigate the role and flexibility of processes and mechanisms important during development for ToM. For example, it would allow the assessment of the role of simulation for ToM development and ability. This approach would be particularly insightful as it would enable the collection of explicit (as opposed to solely implicit) and self-reported (as opposed to solely task-related) measures, which is not possible e.g. in the infant population. Overall, studying this population can further allow us to assess 1) at which point during development we form an understanding of others' minds as a result of differing sensory experiences, and 2) how flexible this ability is.

2.2 Sensorimotor-driven embodiment for ToM

Conducting research with individuals with limb difference would be beneficial also for investigating the role of sensorimotor-driven embodiment for ToM development in an innovative way. Indeed, as previously discussed in this thesis, the mechanisms underlying ToM emergence and ability remain unknown and it has been debated whether a sensory and bodily component of ToM exists.

Embodied cognition implicates an involvement of sensorimotor experiences towards the creation of mental representations, which are thus grounded and processed at this sensorimotor level, rather than being "represented and processed abstractly in an amodal conceptual system" (Pezzulo et al., 2011, p. 1). Previously, sensorimotor-driven embodiment has been suggested to underlie higher cognitive capabilities, including ToM (e.g. Chasiotis et al., 2006; Dyck et al., 2006). Specifically, the embodied theory for ToM suggests that others' minds are understood and predicted via embodied representations which can influence such higher-level cognition, even in abstract domains such as inference of mental states. It has been previously indicated that sensorimotor experiences, thus also the physical characteristics and constraints of an individual's body, shape such embodied representations (Pezzulo et al., 2011).

Only a few studies have investigated the influence of sensory impairment on ToM, mainly through the assessment of individuals with visual (e.g. Anghel, 2012;

Bedny et al., 2009; Koster-Hale et al., 2014; Peterson et al., 2000; Sak-Wernicka, 2016) or hearing (e.g. Figueras-Costa & Harris, 2001; Marschark et al., 2019) impairments. However, similarly to the general debate on ToM emergence during infancy, evidence from studies with these populations contrast each other, preventing conclusive remarks. Nevertheless, these studies also provide interesting insights into the role of sensory experience and embodiment for ToM development and ability.

For example, Sak-Wernicka (2016) did not find differences in ToM ability between blind and sighted individuals, which supported previous work suggesting that ToM development may not depend on visual experience. However, this study identified an impairment in the recognition of mental states in the group with visual impairment and suggested that there is (at least partly) a role for visual cues in the understanding of others' minds and predicting others' behaviour. Koster-Hale et al. (2014) investigated the representation of others' mental states in sighted and congenitally blind adults by assessing activation of the right TPJ (rTPJ) in response to stories representing mental states. This study reported comparable neural representation of mental states in both groups, indicating that these representations can emerge also with no first-person perceptual experience with sight. Similar findings were provided by Bedny et al. (2009), who reported a group of congenitally blind individuals to have typical ToM ability through neuroimaging assessment during ToM-related tasks. Furthermore, this study concluded that the neural mechanisms underlying ToM develop from innate factors and experience, regardless of the modality of experience.

While these results are compelling and suggest that embodiment might not be necessary to represent others' mental states, developmental evidence demonstrated the presence of atypical ToM in children with severe visual impairments or total blindness (Peterson et al., 2000) or with hearing impairments (Meristo et al., 2007, 2016; Richardson et al., 2020), thus possibly supporting the embodied theory for ToM ability and development. Briefly, Peterson et al. (2000) investigated the development of ToM in children of three age groups (6, 8 and 12 years) and with severe visual impairments or total blindness. They identified a significant improvement in performance in false-belief tasks with increasing age, which was independent of the level of visual impairment and that was ultimately comparable to typical belief understanding at 12 years of age. Richardson et al. (2020) investigated instead ToM development in 4-12-year-old children and adults with hearing impairment through neuroimaging and behavioural tasks. Results from this study indicated a reduced selectivity of the rTPJ for mental states in children with delayed exposure to sign language, while the same was not valid for adults. Therefore, the authors concluded that language exposure facilitates the development of brain regions specialised for ToM. Similar findings on the importance of conversational exchanges for the promotion of expression of ToM were reported by Meristo et al. (2007, 2016) who also investigated delays in understanding others' mental states in deaf children.

In support of the importance of an embodied theory for ToM, Hughes and Leekam (2004) suggested that intact sensory experiences provide access to crucial information about other people's mental states, which are indicated to possibly explain ToM deficits in individuals missing e.g. visual and auditory cues during social interaction due to sensory impairment or loss. A role for motor experience for higher-level cognition, including ToM, was also evidenced by Dyck et al. (2006) who found significant correlations between scores at motor coordination and ToM tasks, as well as at other cognitive tasks assessing emotion recognition and understanding. Finally, the influence of sensorimotor experience on ToM has also been highlighted by

Chasiotis et al. (2006). Specifically, by comparing the performance at conflict inhibition and ToM tasks of pre-schoolers diagnosed with a sensory integration disorder vs controls, the authors found a worst performance in the former group. This was attributed to differences in sensory motor inhibition and the authors suggested motor inhibitory abilities as a prerequisite for ToM.

The above studies, and some others (e.g. see review by Leonard & Hill (2014)), point to a rich and complex relationship between sensorimotor experiences and ToM, as well as other cognitive domains, which may be developmental in nature. Indeed, embodied theories of cognition generally predispose that the body shapes cognition during development and at all later stages.

Nevertheless, whether and how sensorimotor experiences (and lack thereof) impact ToM development remains to be determined, as well as the plausibility of an embodied theory for ToM.

Introducing the limb difference population in ToM research represents a unique opportunity to investigate the influence on ToM of varying sensorimotor experiences, and thus embodiment, compared to the general population. In other words, such research will contribute to shed light into the relationship between embodiment and ToM development, as well as its flexibility, which remain open questions in the literature. Furthermore, by assessing and comparing the two limb difference subgroups (i.e. congenital and acquired), such mechanisms can be studied from a developmental perspective and could further inform our knowledge on ToM emergence acquired through infant studies. Indeed, it can be determined through this population whether there exists a critical time in development during which sensorimotor-driven embodiment has an impact on ToM ability, as well as whether compensatory mechanisms (e.g. neural systems matching action observation, action execution, and motor imagery (Funk et al., 2005)) may take place to achieve typical ToM in individuals with sensory impairments.

2.3 Perspective taking for ToM

Amongst other factors, perspective taking ability has often been suggested as being critical for ToM. Indeed, engaging in ToM is often referred to as putting yourself in someone else's shoes or taking their point of view (e.g. Hynes et al., 2006; Jauniaux et al., 2019). Perspective taking has been previously categorised into (1) level-1 vs level-2 perspective taking (e.g. Kessler and Rutherford, 2010) or (2) visuo-spatial vs psychological perspective taking (e.g. Erle and Topolinski, 2015). In contrast to level-1 and visuo-spatial perspective taking, level-2 and psychological perspective taking refer to the ability to mentally adopt someone else's point of view. This ability has indeed previously been implicated in success in ToM-related tasks (e.g. false-belief tasks) and poor performance is instead associated with the inability to differentiate one own's perspective (e.g. beliefs) from that of others (e.g. Brandt et al., 2016). Erle and Topolinski (2015) and Kessler and Rutherford (2010) suggest an embodied component to psychological perspective taking, considered by the former as a deliberate simulation movement and as an embodied self-rotation by the latter. Investigating perspective taking in the limb difference population would represent a great tool to investigate (a) the role of perspective taking for ToM development and ability; (b) the embodiment component of psychological perspective taking; and (c) factors that may lead to impaired perspective taking (e.g. self-other dissimilarity) and the resulting effect on ToM ability. In turn, this would further inform the current debate on ToM emergence and the mechanisms underlying this cognitive ability.

Furthermore, given the studies of developmental nature possible with the limb difference population, the factors (e.g. sensorimotor-driven embodiment, experience, innate and automatic mechanisms, etc.) influencing the development of perspective taking and self-other distinction (also when conflicting) with respect to ToM can be further assessed.

2.4 Mental rotation for ToM

Another factor considered important for ToM is mental rotation ability, which refers to “spatial visualization” and involves the “ability to imagine the movements of objects and spatial forms” (Hegarty & Waller, 2004, p. 175). Most of the evidence in the literature supports the differentiation of mental rotation ability from perspective taking ability (e.g. De Beni et al., 2006; Hegarty & Waller, 2004; Hirai et al., 2013; Inagaki et al., 2002; Kozhevnikov et al., 2006; Kozhevnikov & Hegarty, 2001), with perspective taking rather referring to “spatial orientation” (Hegarty & Waller, 2004).

By assessing healthy participants in several perspective taking tasks, Erle and Topolinski (2017) concluded that “visuospatial perspective-taking involves a mental transformation of one's body schema into the physical location of another person” (p. 683). Furthermore, Xie et al. (2018) assessed participants in a false-belief task involving conditions varying both in self- and other-perspective disparity, as well as the angle of disparity. The results from this study suggested a role for mental rotation in false belief understanding, thus for ToM. Overall, these studies support the involvement of mental rotation for ToM development and ability, suggesting embodied mental transformation for better performance in perspective taking and false belief tasks, respectively. However, additional studies are warranted to further evidence the involvement, flexibility and extent of this effect of mental rotation ability on ToM. For example, studies should address (a) whether this mental transformation also happens when the observer has a different body schema to the person observed; (b) how necessary is the embodied component for achieving correct mental rotation and whether additional or compensatory mechanisms exist; and (c) the effects of impaired mental rotation for ToM.

Once again, individuals with limb difference represent an exemplary population to address these questions and to further understand the role of mental rotation for ToM. There are three main reasons underlying my statement, that I will now briefly describe. First, individuals with limb difference have different body representations and characteristics compared to the general population (Guo et al., 2017); therefore, these characteristics allow the assessment of both the flexibility of mental transformation and the embodiment component. Second, individuals with acquired limb difference were shown to have impaired mental rotation ability (Guo et al., 2017; Lyu et al., 2017). Specifically, Lyu et al. (2017) indicated, using neuroimaging in a mental rotation task, a decreased perceptual salience of hand pictures and an increased significance of the intact hand in upper limb amputees. In addition, Guo et al. (2017) suggested that this effect could be reversed or refrained when using prostheses, thus that mental rotation ability could be preserved. Notwithstanding this evidence, no study has directly addressed the impact of mental rotation impairment seen in people with limb difference on their ToM ability. Third and last, the developmental stages for the above can also be addressed through the limb difference population, ultimately enabling a more complete understanding of ToM development and ability.

3. Experimental Contributions

3.1 Enhanced Theory of Mind in Individuals with Limb Difference: Embodiment for Theory of Mind Development & Ability

Introduction

A critical question that remains unanswered in the literature concerns the developmental mechanisms underlying ToM. Specifically, while different theories have been put forward suggesting different causal factors and mechanisms involved in ToM ability and development, little empirical evidence has provided direct support to these theories. While I discussed these theories in detail in the previous chapters of this thesis, I direct in this study the conversation on the embodied theory for ToM ability and development, focusing on the plausibility of the simulation mechanism underlying ToM and the associated factors to consider when taking this stance. This study addresses discussions included in Part 2, chapter 2 of this thesis (especially subchapters 2.1 and 2.2). Following, these will be summarised and my motivations for conducting this study will be highlighted.

Embodied theories of cognition have recently become increasingly widespread as an alternative approach to more traditional views of cognition. The embodied theory for ToM suggests that others' minds are understood and predicted via embodied representations which can influence such higher-level cognition, even in abstract domains such as inference of mental states. Evidence in the literature of this embodied theory for ToM is mixed. Indeed, studies assessing ToM ability in people with differing sensory abilities to the general population, i.e. with visual (e.g. Anghel, 2012; Koster-Hale et al., 2014; Peterson et al., 2000) and hearing (e.g. Figueras-Costa & Harris, 2001; Marschark et al., 2019) impairments, provided contrasting results (see Part 2, chapter 2.2 of this thesis for more details). Nevertheless, findings from such studies point to a rich and complex relationship between sensorimotor experiences and ToM, as well as other cognitive domains, which may be developmental in nature. Indeed, embodied theories of cognition generally predispose that the body shapes cognition during development and at all later stages. Nonetheless, questions and doubts on the plausibility of an embodied theory for ToM remain. This is especially true considering the still debated computational mechanisms which may be supporting embodied cognition. Furthermore, its flexibility remains challenged as to whether it can account for a variety of sensorimotor experiences and cognitive representations within and between individuals. Similarly, its flexibility for novel situations (thus with no first-person experience) that humans are faced with on a regular basis is questioned.

The computational mechanism indicated to support embodied cognition is simulation, which is also one of the main candidate mechanisms suggested to underlie (e.g. Asakura & Inui, 2016) or be the precursor of (e.g. Keysers & Gazzola, 2007) ToM. Other mechanisms, such as association and teleological, have also been previously proposed as the mechanisms behind ToM. While a detailed discussion on all these alternative mechanisms is provided in Part 1, chapter 1 of this thesis, I here focus on the simulation mechanism for ToM given its close relationship to embodiment. Specifically, the simulation theory proposes that actions are understood when the observer directly matches, or mirrors, the observed action onto their own motor system (Rizzolatti et al., 2001); thus fully supporting an embodiment approach to ToM. Some evidence from developmental research (Southgate et al., 2009; Southgate & Vernetti, 2014) suggests that infants might engage in a simulation mechanism to understand others' actions (see Part 1 of this thesis). However, infant studies have provided contrasting results with respect to which mechanism may underlie ToM; thus, whether simulation underlies ToM ability and development remains undetermined.

The suitability of the simulation mechanism for ToM ability and development has been contested (e.g. Frith & Frith, 2006b) by researchers who believe the simulation mechanism to be limited by factors which cannot justify a full understanding of others' minds in our day-to-day life, and who instead suggest a mentalising mechanism, that goes beyond simulation, for engaging in ToM. One of these main factors is that simulation requires similarity between the observer (the self) and the person observed (the other) for the former to be able to share representations and understand others' minds. This similarity is valid on different levels, from bodily and sensorimotor, to perceptual, attentional and more in general cognitive similarities. However, this requirement for self-other similarity does not always translate to real life situations, in which the observer may differ from the observed person, a phenomenon that is referred to in the literature as Correspondence or Embodiment Mismatch problem (Brass & Heyes, 2005; Nehaniv & Dautenhahn, 2002). Nonetheless, given that it is yet to be determined the flexibility of this simulation mechanism and the extent to which self-other similarity may be a prerequisite of ToM, whether a simulation approach may be a good candidate mechanism behind ToM development and ability remains unclear (see Part 1, chapters 1-2, Part 2, chapter 1 of this thesis).

While the contribution of simulation to high-level cognition (e.g. inferring actors' intentions) has been previously challenged (e.g. Heyes & Catmur, 2022), Pezzulo et al. (2013) suggest that embodied simulation may have a key role in higher cognition, but that "further research is necessary to assess how, and how much, sensorimotor and simulative processes are reused for cognitive tasks" (p. 10) and their flexibility. Furthermore, in another paper, Pezzulo et al. (2011) highlight that "grounded cognitive processes have the same power, but also the same constraints, as bodily actions" (p. 7). From these starting statements, in this study I investigated ToM ability in a population new to ToM research, i.e. the limb difference population. This was done in an attempt to shed some light into the role that embodiment may play towards this cognitive ability, as well as into the plausibility of the simulation mechanism for ToM and its flexibility.

Introducing this population with varying sensorimotor, and thus simulative experiences compared to the general population to ToM research provides an opportunity to investigate the relationship between embodiment, simulation and ToM development from a new perspective. Furthermore, this population makes available the comparison of individuals with congenital limb difference vs limb loss acquired at different stages throughout development, as well as varying degrees of limb differences between individuals. Thus, it represents a great chance to study the flexibility of the simulation mechanism, which represents an open question in the literature surrounding embodiment and simulation for cognitive abilities (Pezzulo et al., 2011).

I refer to this approach as an innovative perspective on the study of ToM because it allows the investigation of the mechanisms behind the development of this cognitive ability from the perspective of people whose sensorimotor experiences differ the most from the majority of the population. This study also distinguishes itself from previous literature investigating ToM development in populations with other sensory impairments, as the present research aims to directly address the debate concerning embodiment and simulation for ToM. While studies on individuals with visual and hearing impairments have provided the grounds for assessing the role of embodiment for ToM, I believe that the sensorimotor differences seen in the limb difference population provide the closest condition to achieve an ideal experimental approach to embodiment. Longo et al. (2008) describe such ideal experimental approach as "involving the comparison of one condition in which a participant has a body and another in which they do not" (p. 980). The condition the authors mention is not feasible to implement in human studies, thus their choice to resource to the rubber hand illusion to assess participants' embodiment. In contrast, I believe that studying individuals with limb difference allows us to achieve a condition closer to such an ideal experimental approach by not requiring the manipulation of participants' sensory experiences affecting embodiment and yet allowing the assessment of a different embodiment (compared to the general population) on ToM.

Previous studies (e.g. Aziz-Zadeh et al., 2012; Cusack et al., 2012) proved this population to be promising for tackling the open questions on the relationship between embodiment, simulation and ToM, thus for shedding further light on ToM development and ability. Specifically, they identified the recruitment of brain areas associated with mentalising, in addition to simulation, in individuals with limb difference during the imitation and understanding of goal-directed actions of others who have bodies that differ from the self. The same was not valid for control individuals from the general population, who only relied on simulation for completing the tasks (see Part 2, chapter 1 for more details). Overall, these studies suggest a role for embodiment in imitation and understanding of others' goal-directed actions, as neuroimaging findings vary between people with limb difference and controls. However, they also seem to suggest that a mentalising, rather than a sole simulation, mechanism may better explain these abilities in people with bodies radically different from the self; perhaps highlighting the lack of flexibility of the simulation mechanism. Nonetheless, evidence is yet to be provided as to whether a similar mechanism and flexibility is also transferred to the inference of others' mental states, thus ToM, and as to understanding the involvement of embodiment in ToM development.

To answer these questions, in the next studies I investigated ToM ability in individuals with limb difference vs controls through different assessment tools, including the Strange Stories Film Task (SSFt) (Murray et al., 2017), as well as the Empathy Quotient (EQ) (Baron-Cohen & Wheelwright, 2004) and Interpersonal Reactivity Index (IRI) (Davis, 1980) questionnaires (i.e. self-reported measures).

I chose a mixed experimental approach to this research, including both self-reported and behavioural measures of ToM, in an attempt to also assess whether there is consistency in the extent to which people think they engage in ToM (self-reported ToM) and their actual ToM ability (behavioural data), as well as whether this varies between the limb difference and control groups. For example, Spek et al. (2010) found correlations between self-reported ToM (as assessed through the EQ) and ToM measured through other neuropsychological tasks (i.e. the Strange Stories Task (SSt) (Happé, 1994) and Faux-Pas task (Stone et al., 1998)) in their study comparing individuals with high-functioning autism (HFA) and Asperger syndrome to controls. The self-report questionnaire showed the highest power in discriminating between groups based on their ToM ability. However, conflicting results exist in the literature, such as Melchers et al. (2015) who found the IRI and EQ self-report questionnaires to barely correlate with the Reading the Mind in the Eyes behavioural task (Baron-Cohen et al., 1997). Similarly, Murray et al. (2017) found partial correlations only between some of the SSFt (behavioural) and IRI and EC (self-reported), which varied between the groups compared in their study, i.e. individuals with ASD vs controls.

Furthermore, this approach allows the investigation of whether having a limb difference alters specifically implicit (behavioural) or explicit (self-report) measures of

ToM. This question is driven by results from developmental research debating that implicit and explicit measures may have a variable ability to access ToM (Low & Perner, 2012). For example, implicit tasks were suggested to be more ideal for assessing ToM in children younger than 4 years of age, as they overcome limitations seen in the explicit tasks, such as requirement for syntactic and executive functions (Grosse Wiesmann et al., 2017). Nonetheless, it would be interesting to measure whether there is also an influence of sensorimotor experiences and embodiment on ToM measures, and if this varies between implicit and explicit measures.

The Strange Stories Film Task (SSFt)

The SSFt was first introduced by Murray et al. (2017), as an adaptation of the original SSt, and it was developed to study the social-cognitive difficulties in adults with ASD. Given that this task was found to successfully and sensitively differentiate ToM ability in adults with ASD vs controls (Murray et al., 2017), I used this task in the present study in an attempt to compare ToM ability in people with limb difference vs controls. Specifically, this task allows the assessment of participants' ability to attribute mental states to others, to interact in a socially acceptable manner based on others' mental states, as well as to use language associated with mental states when describing others' behaviours.

While the SSFt has only been used in the above-mentioned study, its original vignette-based version (the SSt) has been often employed in ToM research, e.g. in relation to ASD (Happé, 1994; Spek et al., 2010) and cross-cultural differences in ToM in children (Wang et al., 2021), as a measure of individual differences in ToM across middle childhood (Devine & Hughes, 2016), to assess the relationship between executive functions and/or social competence and ToM (Devine et al., 2016; Lecce et al., 2017), as well as the relation between ToM and internal-state language (Meins et al., 2006). Furthermore, the SSFt is of particular interest for the present study and our discussion on embodiment for ToM. Indeed, scores at the original SSt were previously found to correlate with scores at a motor coordination task (Dyck et al., 2006). It therefore seems that this task provides ground to investigate how varying sensorimotor abilities influence ToM development.

The Empathy Quotient (EQ) questionnaire

The EQ self-report questionnaire was first introduced by Baron-Cohen and Wheelwright (2004) and was originally developed to study empathising difficulties in adults with Asperger Syndrome and HFA, as well as sex differences in empathy from an adult sample from the general population. This self-report questionnaire was created to assess empathy, defined as follows by the authors in their original paper: “the drive or ability to attribute mental states to another person/animal, and entails an appropriate affective response in the observer to the other person’s mental state” (Baron-Cohen & Wheelwright, 2004, p. 168).

Given that Baron-Cohen and Wheelwright themselves considered several items of the EQ to be representative of ToM, this task has previously been used in research assessing ToM ability, e.g. in relation to ASD (Baron-Cohen & Wheelwright, 2004; Spek et al., 2010) and patients with schizophrenia (Pino et al., 2015), or to cross-cultural studies (Groen et al., 2015).

A following study by Lawrence et al. (2004) investigating the validity and reliability of this assessment tool, managed to distinguish the cognitive and affective components of empathy as assessed through the EQ. Specifically, through a principal component analysis, the authors identified three factors in the EQ, i.e. (1) cognitive empathy, effectively representing ToM; (2) emotional reactivity, linked to affective ToM or empathy; and (3) social skills. This allows a further investigation of the specific cognitive and affective components of the EQ for a more accurate interpretation of study results. The findings by Lawrence et al. (2004) were supported by Muncer and Ling (2006) who also created a short version of the EQ taking into account the three factors.

In addition to its relevance for the assessment of ToM, this questionnaire was included in this study as it may also provide some insights into embodiment for ToM. Indeed, Seiryte and Rusconi (2015) evidenced the EQ score to be predictive of subjective ownership feelings and phenomenological self–other merging in a rubber hand illusion study. Pino et al. (2015) found that action observation and imitation training improves the scores in the EQ in patients with schizophrenia vs controls. Using the EQ, these studies highlight a possible role for sensorimotor-driven embodiment and simulation towards empathy and ToM which I further explored in my studies with the limb difference population.

The Interpersonal Reactivity Index (IRI) questionnaire

The IRI self-report questionnaire, which was first introduced by Davis (1980), allows the assessment of the sub-components of empathy, particularly including 4 different subscales, that are: (1) perspective taking (PT: “the tendency to spontaneously adopt the psychological point of view of others”), (2) empathic concern (EC: “assesses ‘other-oriented’ feelings of sympathy and concern for unfortunate others”), (3) fantasy (FT: “taps respondents' tendencies to transpose themselves imaginatively into the feelings and actions of fictitious characters in books, movies, and plays”), and (4) personal distress (PT: “measures ‘self-oriented’ feelings of personal anxiety and unease in tense interpersonal settings”).

In this study, only the PT and EC subscales were employed. This choice was driven by the fact that these two subscales were determined in previous studies to be assessing more robust and representative components of empathy (Alterman et al., 2003), while the validity of the other two subscales has been debated (Cliffordson, 2001). The inclusion of this questionnaire in this research allows us to further assess the association between empathy and/or perspective taking and ToM, considering that previous studies found both subscales of the IRI to correlate to EQ scores as well as the “emotional reactivity” factor of the EQ (e.g. Lawrence et al., 2004).

Furthermore, the choice of additionally administering this questionnaire to participants was driven by the fact that the IRI includes a measure of perspective taking, i.e. the PT subscale, which is of great interest for my research. Indeed, ToM has often been described in the literature as the ability to take someone else’s perspective (e.g. Hynes et al., 2006). Therefore, this questionnaire has been previously used in research on ToM as a self-report measure of ToM, e.g. in relation to both neuroimaging and behavioural studies on ToM deficits in people with schizophrenia and their relatives (Hooker et al., 2011; Montag et al., 2012; Schiffer et al., 2017).

Finally, perspective taking has been previously suggested in the literature to underlie the simulation mechanism (Conson et al., 2015; M. R. Johnson & Demiris, 2005; R. Langdon & Coltheart, 2001). The scores at the PT subscale of the IRI have in turn been previously reported to be associated with increased prefrontal cortex and premotor activity, as well as delayed emotion attribution decisions by Haas et al. (2015); thus showing the ability of the IRI-PT subscale to provide access to discussions on the simulation mechanism. For this reason, the PT subscale of the IRI in particular will enable the assessment of embodiment, as well as simulation in my sample population.

In the next sections, results from all the studies will be presented and discussed separately. Then, a general discussion will be provided to summarise the implication of all the findings for ToM development and ability, as well as the limitations of these studies. Finally, some conclusions will be drawn.

Study 1: The Strange Stories Film Task

In Study 1, I used the SSFt to assess and compare ToM ability in individuals with and without limb difference, in an attempt to determine the involvement of embodiment in ToM ability and development, as well as to investigate the plausibility of the simulation mechanism underlying this cognitive ability and its flexibility.

Methods

Participants

A total of 27 adults with limb difference (N congenital limb difference = 12, N acquired limb difference = 14, N congenital and acquired limb difference = 1) and 26 adult controls were recruited from the "Research for Limb Difference" database (University of Essex), as well as internally at University of Essex or through social media. (See Part 2, chapter 3.3 of this thesis for more details on the Research for Limb Difference database and individuals with limb difference recruited). Additional 40 participants (N limb difference = 22, N controls = 18) were tested but excluded due to one or more of the following reasons: (a) Participant did not complete the study attempt (N = 16; N limb difference = 13, N controls = 3); (b) Participant did not complete at least 60% of the total experimental trials and 67% of the control trials (N = 8; N limb difference = 2, N controls = 6); (c) Audio recordings from the session were not successfully retrieved from the Gorilla platform, an online platform for behavioural studies (N = 4; N controls = 4); (d) Audio recordings were incomplete or unclear (N = 7; N limb difference = 2, N controls = 5); (e) Participants could not be located in the database based on the IDs they inserted upon starting a given online study (N = 5; N limb difference = 5). Reasons for considering the audio recordings not clear included poor audio quality, inaudible voice, background noise interfering with the audio recording, breaks in the recording, unclear sentences. All this was assessed on an individual basis. Upon completion of the study, participants were compensated with a £5 Amazon Gift Card. Ethical approval was granted by the University of Essex Ethics Sub-Committee (ETH2021-0065).

The Strange Stories Film Task

A total of 12 experimental and 3 control clips were presented to participants, where a female and a male actor socially interacted in different situations. The experimental clips were created to test the ability of a person to attribute intentions to others in scenarios including lie, irony, double bluff, pretence, joke, appearance/reality, white-lie, persuasion, misunderstanding, forgetting, contrary emotions, and idioms. The control clips mirrored the experimental clips, except that, rather than requiring the attribution of mental states to the actors in the clips, they tapped into logical reasoning instead. These abilities were assessed through three questions on (1) actor’s intentions, (2) social interaction, and (3) memory, which were presented to the participants following each clip (including the three practice clips). Clips were presented in a quasirandomised order, meaning that half the participants viewed order

A and the other half viewed the same clips but in reversed order (B). Clips lasted no longer than 27 seconds each (M = 17.5, SD = 5.83) and the total running time was 6 min and 21 sec. The scoring system for the SSFt was based on Murray et al. (2017). For the Intention question, the score given reflected how accurately the participant recognised the relevant mental states. Mental state language was also scored to identify whether participants used mental state words (e.g. he wants or she thinks) to describe the actors' intentions. For the Interaction question, scoring reflected the appropriateness of the participant's suggested response to the speaker. For the Memory question, all scores were based on correctly identifying the factual information in the relevant clip. Possible scores ranged from 0 to 2 for the Intention, Mental State Language and Interaction questions and 0–1 for the Memory question for each clip; maximum total scores were therefore 24, 24, and 12, respectively. Please see Box C below for an example and scoring of the Intention, Interaction and Memory questions related to a white lie scene in the SSFt.

Box C. Example Intention, Interaction, and Memory questions in a white lie scene from the SSFt (from Murray et al., 2017).

Intention Question: **"Why did Max say that?":**

2 points - reference to white lie or making her feel good or not wanting to hurt Alice's feelings

1 point - response that states simple traits (e.g. he is nice, being supportive, polite) or is simply relational (e.g. he likes her). Incomplete response (e.g. offering fake praise) or solely motivational (e.g. so she won't be annoyed, avoid an argument, reassure her)

0 points - incorrect for example, "he thought it was good" or only "he didn't like it," or irrelevant responses

Mental State Language **(scored together with above question):**

0 points - no mental state words

1 point - simple mental state words regarding one character or another character's actions OR words that imply psychological states in social context

2 points - meta-cognitive statements for example, beliefs about beliefs OR intentions to affect another person's mental state for example, he did not want to hurt her feelings OR complex collection of mental states

Interaction question: **"if you were in Alice's situation, what would you say next?":**

2 points - statement that acknowledges that Max's comment might not have been completely honest and either asks for additional clarification or additional feedback in socially appropriate manner (e.g. "do you really mean that?"); sarcastic agreement with his opinion that implies it could be improved.

1 point - Incomplete response for example, "thank you," that does not reflect white lie.

Procedure

Given the online nature of the study, participants participated from their homes and there was no experimenter present. The study was hosted on the Gorilla platform. Participants were asked to read the instructions and complete the SSFt. The online experimental procedure of the SSFt was adapted from the original study by Murray et al. (2017). Specifically, clips were presented through the online platform, and the face-to-face interview was replaced with questions being displayed on participants' computer screens with responses recorded on the online platform for offline scoring by two independent raters (Percentage agreement: 80.38%; Cohen's Kappa: .71). Furthermore, participants were asked to conduct some additional checks in an attempt to control the experimental setting and ensure good data quality. Specifically, prior to conducting the SSFt, participants were provided with the audio recording instructions and were asked to complete 3 practice trials (without feedback) to ensure good audio recording quality and understanding of the procedure. Only after completing the practice trials, participants were able to start the SSFt study. Online written informed consent was obtained from all included participants prior to the start of the study.

Statistical analysis

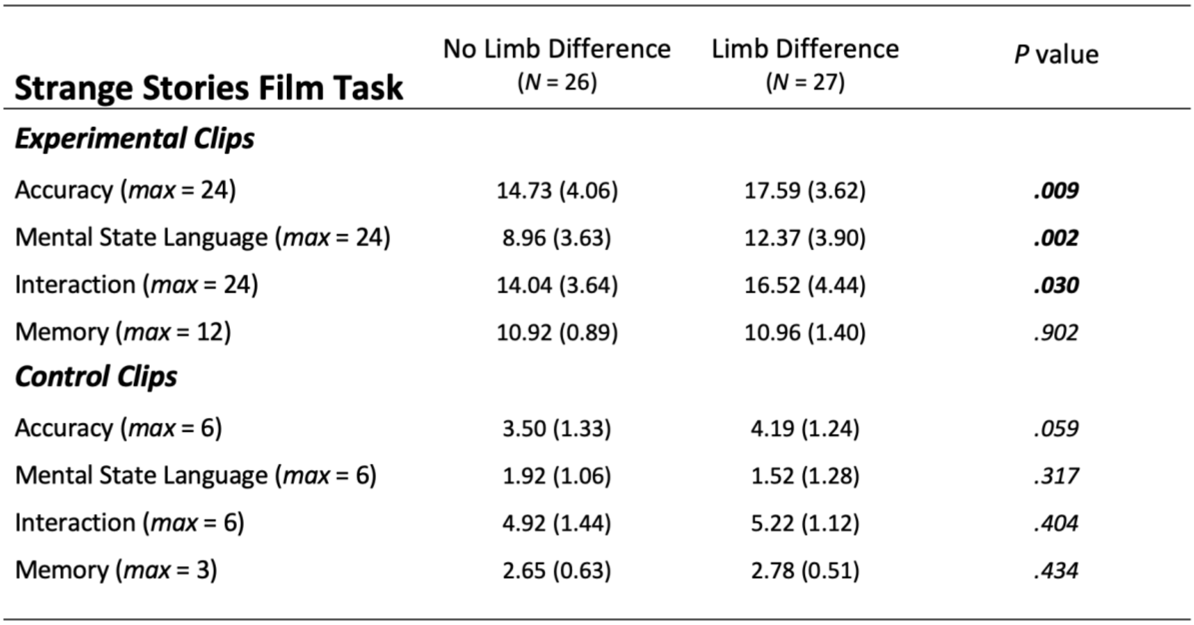
A series of two-tailed independent sample *t* tests were conducted to separately compare the scores in the SSFt between the two groups (with respect to accuracy, mental state language, interaction, and memory questions from *Experimental* and *Control Clips*). Furthermore, a series of two-tailed independent sample *t* tests were conducted to determine the existence of potential differences in performance between the limb difference subgroups (congenital vs acquired) and controls with respect to these variables.

Results

The Strange Stories Film Task

Scores in the SSFt for the limb difference vs control groups are reported, before presenting an analysis of performance by limb difference subgroups (i.e. congenital vs acquired) and their relation to the main results.

Table 4. Participants' performance in the SSFt by group.



Strange Stories Film Task	No Limb Difference (*N* = 26)	Limb Difference (*N* = 27)	*P* value
Experimental Clips			
Accuracy (*max* = 24)	14.73 (4.06)	17.59 (3.62)	***.009***
Mental State Language (*max* = 24)	8.96 (3.63)	12.37 (3.90)	***.002***
Interaction (*max* = 24)	14.04 (3.64)	16.52 (4.44)	***.030***
Memory (*max* = 12)	10.92 (0.89)	10.96 (1.40)	*.902*
Control Clips			
Accuracy (*max* = 6)	3.50 (1.33)	4.19 (1.24)	*.059*
Mental State Language (*max* = 6)	1.92 (1.06)	1.52 (1.28)	*.317*
Interaction (*max* = 6)	4.92 (1.44)	5.22 (1.12)	*.404*
Memory (*max* = 3)	2.65 (0.63)	2.78 (0.51)	*.434*

Table 4 above shows participants' performance in the SSFt by group. Participants with limb difference scored significantly higher than controls in all *Experimental Clips* of the SSFt, except for the Memory question, which yielded similar results between the two groups. Both groups performed similarly well in all control clips and no significant group differences were observed. Please see Figure 6 below for a visualisation of the *Experimental Clips* data by group.

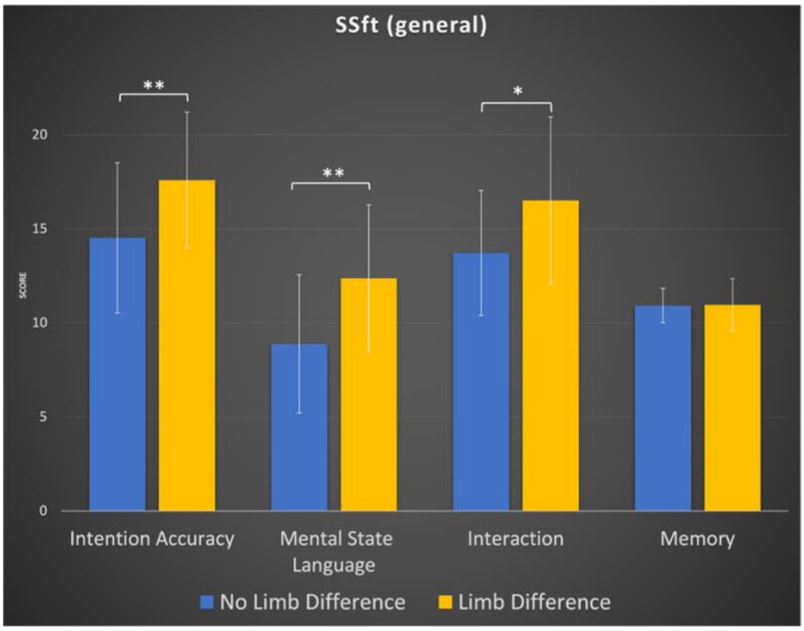
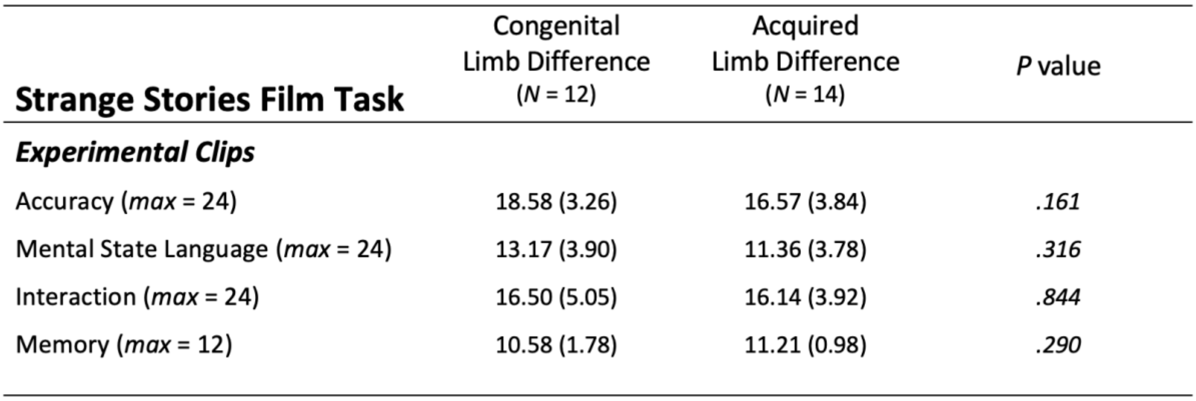


Figure 6. Comparisons of performance at the SSFt between the limb difference population and controls by subscales. Average points scored in each subscale of the task. Error bars indicate standard deviation. SSFt: Strange Stories Film Task; **: significant at the .01 level (2-tailed); *: significant at the .05 level (2-tailed).

A further analysis conducted by limb difference subgroups identified the scores of participants with congenital and acquired limb difference to be similar in all items of the SSFt (see Table 5).

Table 5. Participants' performance in the SSFt by subgroup.



Strange Stories Film Task	Congenital Limb Difference (*N* = 12)	Acquired Limb Difference (*N* = 14)	*P* value
Experimental Clips			
Accuracy (*max* = 24)	18.58 (3.26)	16.57 (3.84)	*.161*
Mental State Language (*max* = 24)	13.17 (3.90)	11.36 (3.78)	*.316*
Interaction (*max* = 24)	16.50 (5.05)	16.14 (3.92)	*.844*
Memory (*max* = 12)	10.58 (1.78)	11.21 (0.98)	*.290*

While these scores did not differ significantly between subgroups, the Accuracy and Mental State Language scores of participants with congenital limb difference were found to significantly differ from those of controls (Accuracy: M = 18.58 points, SD = 3.26 vs M = 14.73, SD = 4.06, respectively, t(26.396) = 3.127, *p* = .004; Mental State Language: M = 13.17, SD = 3.90 vs M = 8.96, SD = 3.63, respectively, t(20.107) = 3.155, *p* = .005. Please see Figure 7 below for a visualisation of the *Experimental Clips* data by subgroups vs controls.

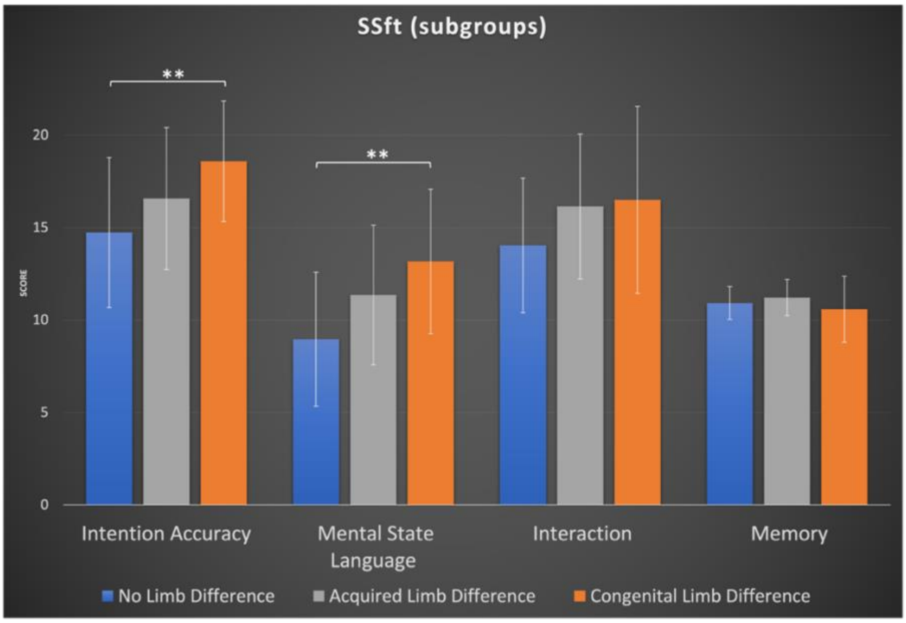


Figure 7. Comparisons of performance at the SSFt between the limb difference subgroups and controls by subscales. Average points scored in each subscale of the task. Error bars indicate standard deviation. SSFt: Strange Stories Film Task; **: significant at the .01 level (2-tailed).

Discussion

In this study, I assessed ToM ability in people with limb difference and compared it with controls, in an attempt to determine the involvement of embodiment for ToM ability and development, as well as to investigate the plausibility of the simulation mechanism underlying this cognitive ability and its flexibility. I used the SSFt to measure ToM ability in participants, as this task has been previously suggested to capture differences in ToM between people with ASD and the general population. Furthermore, given that scores at its original version have been previously shown to correlate with motor skills scores, it provided ground to investigate embodiment guided by sensorimotor influences on ToM development.

In summary, my study findings suggest people with limb difference have enhanced ToM ability compared to controls, as they were found to better understand others' intentions and mental states (as seen through their responses to the Accuracy and Mental State Language questions of the SSFt), as well as better social skills (as seen through their responses to the Social Interaction question of the SSFt). Furthermore, my results indicate that this effect may be developmental in nature and more strongly related to ToM rather than social cognition in general. Indeed, individuals with congenital limb difference (thus with no sensorimotor experience at all during development from at least one of their limbs) were found to perform significantly better than controls (who never experienced any sensorimotor impairment) in inferring intentions and referring to mental states, while the same was not valid for their social interaction abilities. In contrast, a significant difference in performance was not seen between individuals with acquired limb difference (who had varying sensorimotor experience prior to impairment) and controls in any of the SSFt questions, although a trend was observed with the former group scoring slightly higher. Overall, my results seem to indicate a role for embodiment driven by sensorimotor experience on ToM development and ability, with sensorimotor impairment playing a part towards improved ToM, possibly driven by its motor component. Finally, the results from my subgroup analysis suggest that a critical window during development may exist for achieving this relative strength in people with limb difference, an effect which however is flexible and persists throughout adulthood.

More in detail, this study represents the first successful attempt at utilising the SSFt to assess ToM in adults with limb difference, as opposed to adult controls. Specifically, while SSFt was originally developed to test differences in ToM among people with ASD and the general population (Murray et al., 2017), my results suggest that this task can also be informative for testing differences in ToM ability among people with and without limb difference. This result thus extends the utility of this tool to also test differences in ToM in people with differing sensorimotor abilities (in our case driven by limb loss or deficiency), in addition to social and cognitive impairments only. Furthermore, this result supports and extends findings of Dyck et al. (2006) who found a correlation between the scores at the original version of the SSFt and motor coordination. Specifically, although I did not measure any sensorimotor skill in particular, by investigating people with limb difference in the SSFt and observing improved performance in this group, my findings support the stance of an embodied component driven by sensorimotor experience for ToM, which is accessible through the SSFt.

My study found that participants with limb difference scored significantly higher than controls in all subsections of the SSFt, except for the memory question. These results indicate enhanced social cognition in participants with limb difference in the realms of intention recognition, mental state language utilisation, and appropriateness of social interaction relevant to everyday social communication and interaction. These results contribute to the debate surrounding the role for embodiment and sensorimotor experience towards ToM. Specifically, on the one hand, given the lack of impairment in ToM seen in the limb difference population, my findings may, at a first glance, seem to exclude a role for embodiment and first-person sensorimotor experiences towards ToM. Indeed, my findings may seem to support previous studies suggesting intact ToM in people with other sensory impairments. For example, Koster-Hale et al. (2014) indicated comparable neural representations of mental states in individuals with visual impairments and controls. On the other hand, however, my findings not only indicate that a ToM impairment is not present in people with limb difference, but they report an *enhanced* ToM ability driven by limb difference. These results make it thus clear that limb difference does affect ToM ability and that this impairment does lead to a relative strength, suggesting that individuals have an advantage with respect to ToM ability given their varying sensorimotor impairment. Therefore, overall my findings support previous studies indicating atypical ToM in individuals with sensory impairments, e.g. visual (Peterson et al., 2000) and hearing (Figueras-Costa & Harris, 2001) impairments. However, they suggest atypical ToM in people with limb difference in terms of enhanced, rather than impaired, ToM. While these papers showing atypical ToM in people with sensory impairments indicate a role for embodiment in ToM ability and development, the fact that only impairments in the sensorimotor realms lead to *enhanced*, rather than impaired, ToM may highlight that the motor component of such experiences specifically impacts ToM. The nature of this interaction can be twofold. One the one hand, motor experiences may have a positive impact on ToM ability and development. This potentially leads the way to the involvement of the simulation mechanism underlying such an effect, given that simulation involves the direct representation of others' actions onto own *motor* system (Rizzolatti et al., 2001). However, it remains at this stage unclear how simulation can be achieved in individuals who lack direct sensorimotor experience and whose self differs from others.

On the other hand, these results may indicate a supplementary role for motor experiences towards ToM, instead highlighting a crucial role of the non-motor, more deliberative component of ToM in the limb difference population. Nevertheless, I am not able through this study to determine whether enhanced ToM in individuals with limb difference is a result of their (a) enhanced motor simulation ability (through compensatory mechanisms) or (b) enhanced non-motor component of ToM. Nonetheless, I attempted to further investigate the plausibility of the simulation mechanism in individuals with limb difference in Studies 2 and 3 below. Ultimately, the interpretations here advanced will be further discussed in the general discussion of this subchapter, in light of the findings from the next studies as well.

It is worth noting that both groups reported the lowest scores in the Mental State Language subsection of the SSFt compared to the Accuracy and Interaction subsections. This result may suggest that reasoning about and referencing others' mental states may be separate to inferring their intentions and may be a more complex ability given the lower scores in both groups. However, the poorer performance seen in this subscale may also indicate that this type of behavioural task may not be the most sensitive tool to access this ability. Indeed, while this SSFt investigates the elaboration of others' mental states implicitly, more explicit tools, such as self-report questionnaires, may be more indicative of this ability. This possibility is further assessed in Studies 2 and 3 of this subchapter, which make use of two explicit measures previously employed to assess self-reported ToM, and the "Further Analyses" section, which examines the correlation between explicit and implicit measures for ToM. Nonetheless, the higher scores seen in the limb difference group again suggest this population has an advantage with respect to talking about others' mental states. One of the reasons behind the increased ability of people with limb difference to use mental state language may be increased exposure to conversations about the mind, improving their understanding of others' mental states and social situations. The *conversational account* of ToM development suggests that experience with language associated with mental states during development may have an impact on an individual's ability to understand and reference others' mental states. For example, richness of parent-child conversations on ToM, as well as the presence of siblings and other family members has been previously related to enhanced understanding and reference to mental states (Garfield et al., 2001). Similarly, Symons (2004) concluded, following a review on studies in the literature on ToM development, that there exists a relationship, although not deterministic, between children's exposure to discourse about others' mental states and their understanding of others' mental states. In addition, Taumoepeau and Ruffman (2008) evidenced that mothers' reference to others' mental states in their conversations with their 24-month-old infants predicted children's later mental state language use at 33 months. Ornaghi et al. (2011) conducted a training study in pre-school children and confirmed that use of mental state lexicon results in increased metacognitive vocabulary understanding and emotional understanding. Future studies with the limb difference population are warranted to validate this conversational account hypothesis for ToM in this population. Another possible explanation behind this result is the relation between use of others' mental states language and emotional understanding, as highlighted e.g. in Ornaghi et al. (2011). In concordance, Hughes & Leekam (2004) discuss the critical role of emotions in understanding others' minds and present evidence in the literature suggesting that the emotional context influences children's reflection on others' inner states. Whether people with limb difference also report higher emotional understanding and empathy, thus whether these are factors that may have affected such results on this SSFt task, will be assessed in the next studies.

Further insights on the relative strength that represents having a limb difference with respect to ToM can be provided by my subgroups analysis comparing different types of limb differences, i.e. congenital and acquired, vs controls. My results from this analysis suggest that the better performance seen in the limb difference group in the Accuracy and Mental State Language questions of the SSFt, compared to controls, is driven by individuals with congenital (as opposed to acquired) limb difference. Indeed, the congenital subgroup reported significantly higher scores in both elements of the SSFt than controls, while similar scores were achieved in the Interaction question. First, this indicates that, rather than affecting general social skills (such as social interaction), congenital limb difference (thus differing embodiment driven by sensorimotor impairment) may have a specific impact on intention recognition and mental state inference and expression (thus ToM-related skills). Second, this result may suggest that the embodiment driven by sensorimotor experience for ToM might be a phenomenon developmental in nature. Specifically, this result may indicate the presence of a critical period during development for acquisition of ToM as mediated by sensorimotor-driven embodiment. Indeed, individuals with congenital vs acquired limb difference vs controls all have differing embodiments given their varying (or absent) impairment in sensorimotor experiences throughout development. Specifically, individuals with congenital limb difference lack typical sensorimotor experience from at least one of their limbs since birth. Individuals with acquired limb difference report typical sensorimotor experience early in their life but at varying points during development this is impaired due to limb amputation. Controls report instead typical sensorimotor experience throughout their life. Overall, my results seem to suggest that embodiment driven by sensorimotor experiences is critical for ToM ability and development, with its impairment being most influential towards enhanced ToM when occurring since birth (in congenital limb difference). However, the influence of sensorimotor-driven embodiment does seem to continue to affect ToM throughout adulthood, given my results from the acquired limb difference population, showing a flexibility in its effect. However, given the small sample size of these subgroups (see limitations section), additional studies to validate these findings are warranted.

Study 2: The Empathy Quotient Questionnaire

In contrast to Study 1, which used an implicit measure of ToM, in Study 2 I assessed whether the different embodiments driven by varying sensorimotor experience also affect an explicit, self-report measure of ToM and other components of social cognition (e.g. emotional reactivity and social skills). This enabled the assessment of whether the results found in Study 1 were specific to the particular measure used, as well as whether such relative strength observed in people with limb difference could be extended to other components of social cognition.

Methods

Participants

The same recruiting procedures as Study 1 were here followed. A total of 37 adults with limb difference (N congenital limb difference = 13, N acquired limb difference = 23, N congenital and acquired limb difference = 1) and 31 adult controls took part in Study 2.

The Empathy Quotient Questionnaire

The EQ questionnaire consists of 40 empathy items and 20 filler/control items, which include statements needed to be scored on a 4-point Likert scale varying from “definitely agree” to “definitely disagree”. On each empathy item a person can score 2, 1, or 0, so the EQ has a maximum score of 80 points and a minimum of 0. While the experimental items aim at tapping into empathy, the filler items were created and included in the questionnaire to distract the participants from focusing their attention on empathy. Please see Box D below for some examples of experimental vs filler items of the EQ.

Box D. Example of experimental vs filler items in the EQ (from Baron-Cohen & Wheelwright, 2004).

***Example Experimental Items*:**

- I can easily tell if someone else wants to enter a conversation
- I find it easy to put myself in somebody else's shoes

Example Filler Items:

- I prefer animals to humans
- I try to keep up with the current trends and fashions

Procedure

Given the online nature of the study, participants took part from their homes and there was no experimenter present. To start the study, participants accessed the Gorilla platform where the study was hosted. Next, participants read the study instructions and completed the EQ questionnaire. Keyboard responses to the EQ questionnaire were recorded on the Gorilla online platform for offline scoring. Written consent was obtained from all included participants prior to the start of the study.

Statistical Analysis

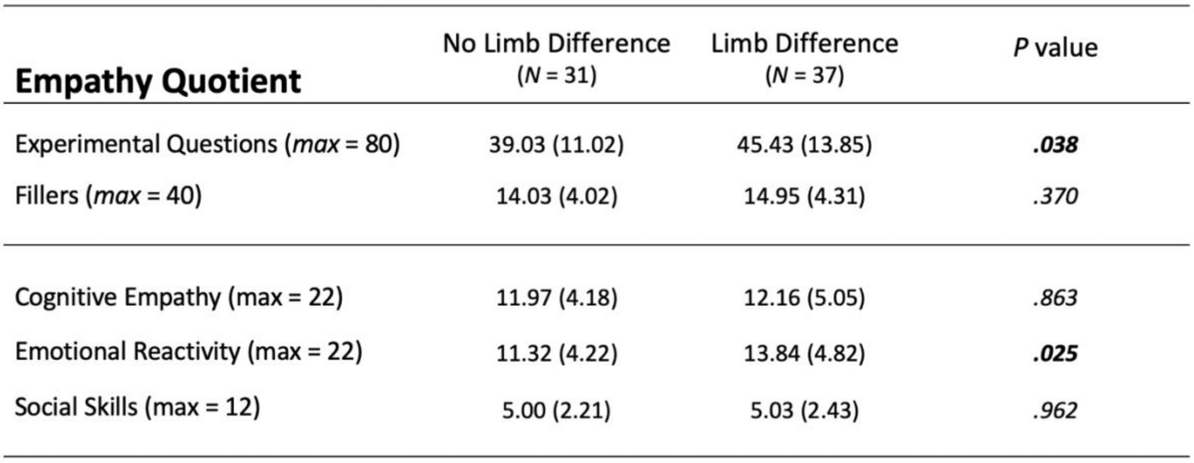
A series of two-tailed independent sample *t* tests were conducted to separately compare the scores between the limb difference vs controls obtained in the EQ self-report questionnaire (with respect to the *Experimental* and *Filler Questions*). Furthermore, a series of two-tailed independent sample *t* tests were conducted to determine the existence of potential differences in performance between the limb difference subgroups (congenital vs acquired) and controls with respect to these variables. The same analyses were then conducted by factors of the EQ, i.e. cognitive empathy, emotional reactivity, and social skills.

Results

The Empathy Quotient Questionnaire

Scores in the EQ questionnaire for the limb difference vs control groups will be reported, before presenting an analysis of performance by limb difference subgroups (i.e. congenital vs acquired) and their relation to the main results.

Table 6. Participants' performance in the EQ self-report questionnaire by group.



Empathy Quotient	No Limb Difference (*N* = 31)	Limb Difference (*N* = 37)	*P* value
Experimental Questions (*max* = 80)	39.03 (11.02)	45.43 (13.85)	***.038***
Fillers (*max* = 40)	14.03 (4.02)	14.95 (4.31)	*.370*
Cognitive Empathy (max = 22)	11.97 (4.18)	12.16 (5.05)	*.863*
Emotional Reactivity (max = 22)	11.32 (4.22)	13.84 (4.82)	***.025***
Social Skills (max = 12)	5.00 (2.21)	5.03 (2.43)	*.962*

Mean EQ scores by groups are presented in Table 6. On average, participants with limb difference scored significantly higher than controls in the *Experimental Questions* of the EQ. In contrast, the filler questions yielded similar results between the two groups. With regards to the EQ factors analysis, the only significant difference in scores between groups was in relation to the Emotional Reactivity factor, with the limb difference group outperforming controls. Furthermore, the limb difference group scored significantly higher in the Emotional Reactivity (M = 13.84, SD = 4.82) vs Cognitive Empathy (M = 12.16, SD = 5.05) component of the EQ, t(36) = 2.740, *p* = .009, while the same was not valid for the control group. Please see Figure 8 below for a visualisation of the *Experimental Questions* data by group.

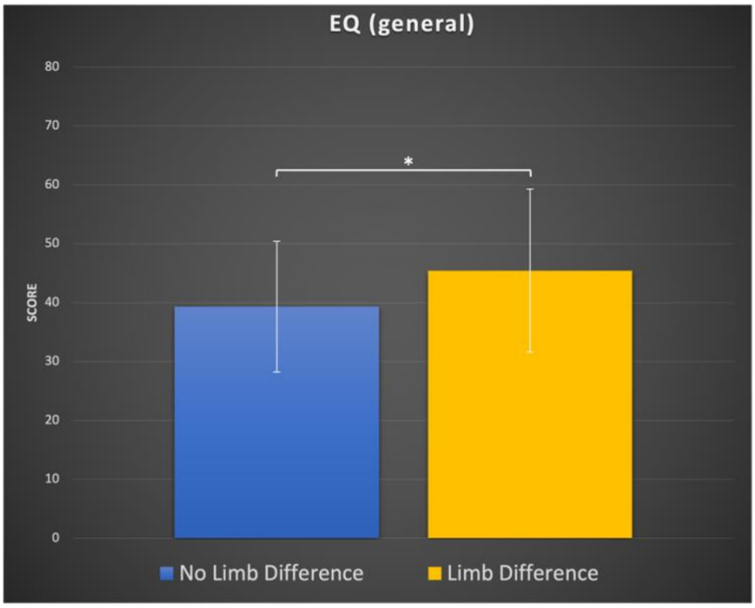
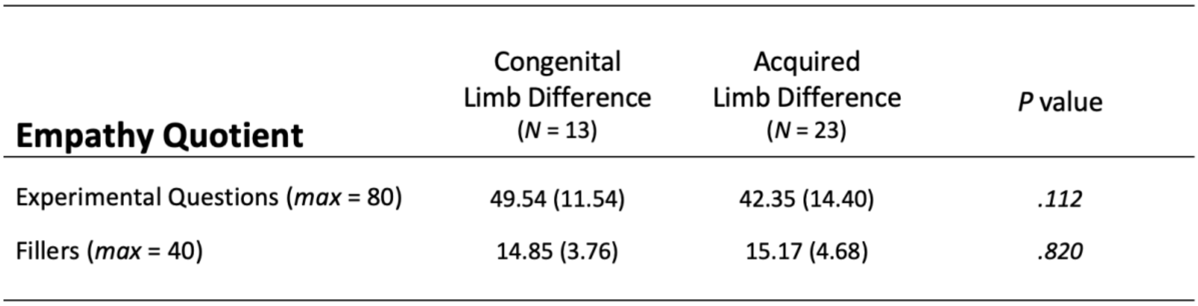


Figure 8. Comparisons of performance at the EQ questionnaire between the limb difference population and controls. Error bars indicate standard deviation. EQ: Empathy Quotient; *: significant at the .05 level (2-tailed).

A further analysis conducted by limb difference subgroups identified the scores of participants with congenital and acquired limb difference to be similar in both the *Experimental and Filler Questions* of the EQ questionnaire (see Table 7).

Table 7. Participants' performance in the EQ self-report questionnaire by subgroup.



Empathy Quotient	Congenital Limb Difference (*N* = 13)	Acquired Limb Difference (*N* = 23)	*P* value
Experimental Questions (*max* = 80)	49.54 (11.54)	42.35 (14.40)	*.112*
Fillers (*max* = 40)	14.85 (3.76)	15.17 (4.68)	*.820*

While these scores did not differ significantly between subgroups, the score for *Experimental Questions* of participants with congenital limb difference (M = 49.54, SD = 11.54) was found to significantly differ from that of controls (M = 39.03, SD = 11.02), t(21.653) = 2.791, *p* = .011. Please see Figure 9 below for a visualisation of the *Experimental Questions* data by subgroups vs controls.

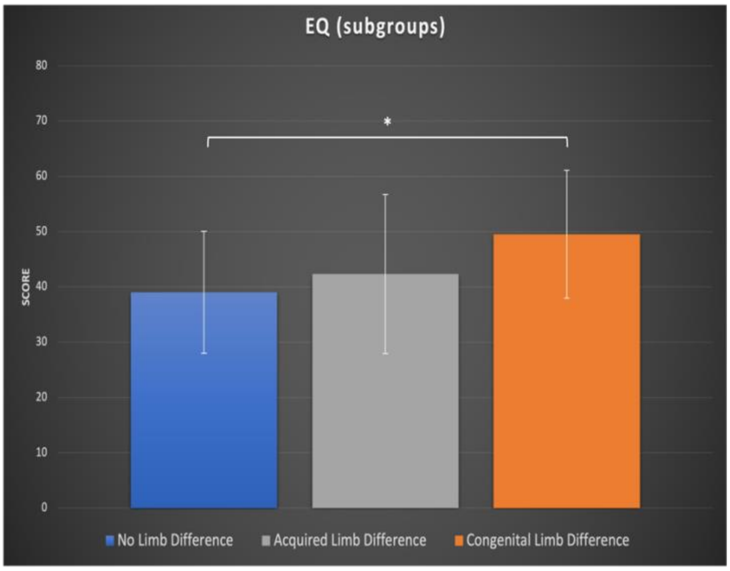


Figure 9. Comparisons of performance at the EQ questionnaire between the limb difference subgroups and controls. Error bars indicate standard deviation. EQ: Empathy Quotient; *: significant at the .05 level (2-tailed).

Similarly, the congenital limb difference group scored significantly higher (M = 15.46, SD = 3.89) than controls (M = 11.32, SD = 4.22) in the Emotional Reactivity factor of the EQ, t(24.426) = 3.141, *p* = .004. Furthermore, they scored significantly higher (M = 14.31, SD = 4.20) than the acquired limb difference group (M = 10.78, SD = 5.18) in the Cognitive Empathy factor of the EQ, t(29.616) = 2.222, *p* = .034. Please see Figure 10 below for a visualisation of the scores at the EQ factors by subgroups.

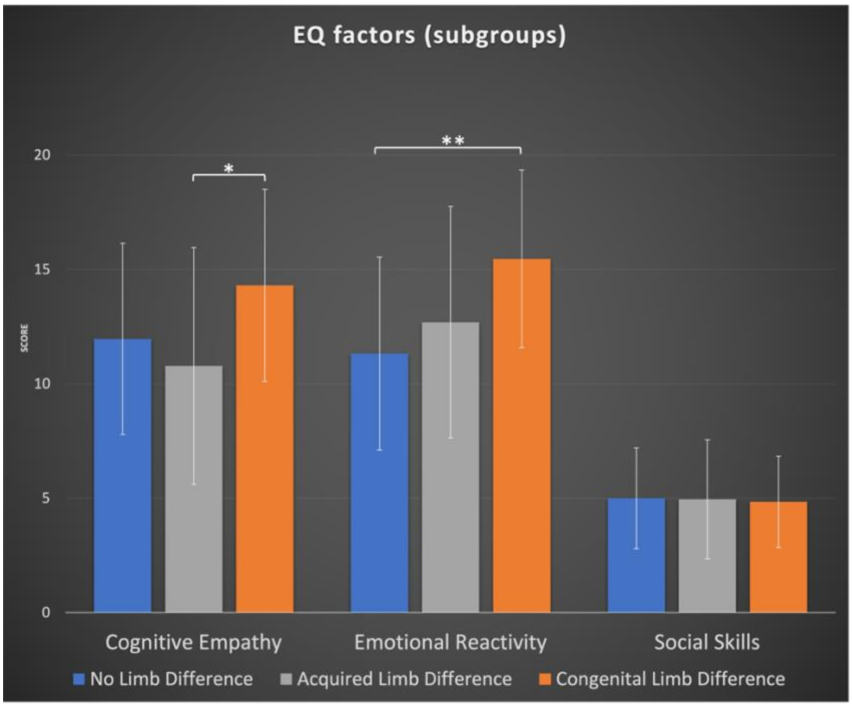


Figure 10. Comparisons of performance at the EQ factors between the limb difference subgroups and controls. Error bars indicate standard deviation. EQ: Empathy Quotient; **: significant at the .01 level (2-tailed); *: significant at the .05 level (2-tailed).

Discussion

In Study 2, I used the self-report EQ as an explicit measure of ToM in an attempt to determine whether the finding of different embodiments driven by varying sensorimotor experiences affecting ToM were specific to Study 1 and the implicit measure of ToM used. Furthermore, I conducted a secondary analysis on the EQ factors to determine whether such relative strength of people with limb difference observed in Study 1 could be extended to other components of social cognition.

The results of Study 2 show that individuals with limb difference have an enhanced ability to understand others' minds, as measured by the self-report EQ questionnaire, compared to controls. My subgroup analysis indicates this effect may be developmental in nature. Indeed, people with congenital limb difference were found to significantly outperform controls at the EQ, while the same was not true for the acquired limb difference subgroup. In addition, my analysis of the EQ factors furthered my results by highlighting that such better performance seen at the EQ for people with vs without limb difference is driven by Emotional Reactivity, rather than Cognitive Empathy or Social Skills. My subgroup analysis of the scores at the EQ factors indicates that this effect may also be developmental in nature, as people with congenital limb difference were found to score significantly better than controls, while the same was not valid for the acquired limb difference subgroup. Finally, a significantly better performance was also seen in the congenital vs acquired limb difference group with respect to the Cognitive Empathy factor. Overall, these results indicate that the role for embodiment driven by sensorimotor experience for ToM, which was previously suggested in Study 1 as assessed through an implicit measure of ToM, maintains also in Study 2 involving an explicit, self-reported measure of ToM. Furthermore, the developmental nature of this enhancing effect highlighted in Study 1 is further supported by my findings in this study. Finally, Study 2 also provides evidence for the extension of this relative strength to the affective component of ToM and provides possible links to simulation as the mechanism underlying such an effect.

Similarly to the SSFt, the EQ has been previously developed to test differences between people with ASD and the general population with respect to their empathic and ToM abilities. This study is the first to use the EQ self-report questionnaire as a tool to assess differences in ToM in people with limb difference vs the general population, thus extending its utility to investigate the influence of varying sensorimotor abilities (in this case driven by limb loss or deficiency) on ToM. Furthermore, this result supports and extends findings from Seiryte and Rusconi (2015) and Pino et al. (2015)

who reported relations between scores at the EQ and body ownership and action observation and imitation, respectively. While I did not measure embodiment through experimental manipulations of participants' body ownership or action observation and imitation, by assessing people with limb difference with the EQ and observing improved performance in this group, my findings support the stance of an embodiment component driven by sensorimotor experience for ToM.

My results suggest that people with limb difference scored significantly higher at the *Experimental Questions* of the EQ compared to controls, highlighting a higher understanding of others' minds, thus confirming results from Study 1. Furthermore, the developmental nature of this effect was also evidenced in Study 2 through my subgroup analysis identifying a significantly better performance in the EQ for the congenital vs control group. Therefore, this suggests that the effect seen in Study 1 is not specific to implicit measures of ToM and can be extended to explicit, self-reported measures. This consistency between implicit and explicit measures of ToM is supported by previous studies, e.g. Spek et al. (2010) who suggested the validation of self-reports for examining ToM in adults with HFA or Asperger syndrome. Nonetheless, this does not imply a correlation between the two assessment tools used in Studies 1 and 2, which is a topic that will be addressed in the "Further analyses" section.

Furthermore, my study replicates the findings obtained in the original paper by Baron-Cohen and Wheelwright (2004) with respect to the scores of the EQ of control participants, as they were reportedly similar to the ones observed in this study (42.1 (10.6) vs 39.30 (11.10), respectively). Interestingly, people with limb difference in this study scored higher than controls in the original study, reaching a score of 46.14 (13.40). Although the statistical significance of these differences cannot be reported, these figures seem to support my finding of an enhanced ability of people with limb difference to understand others' minds.

My secondary analysis on the EQ factors allowed a more specific investigation of the components of social cognition assessed with this self-report questionnaire, to determine which were the ones affected by having a limb difference. Indeed, ToM has been previously described as composed of a "cognitive" and an "affective" component, with the latter emphasising on emotional facets of ToM (e.g. Kalbe et al., 2007; Shamay-Tsoory & Aharon-Peretz, 2007). It is worth noting however that inconsistencies and overlaps exist between the terms ToM and empathy, with empathy also being often described as comprising a "cognitive" and "affective" component (e.g. Baron-Cohen & Wheelwright, 2004). Therefore, "cognitive and affective ToM" and "cognitive and affective empathy" have been used interchangeably; ultimately with the term "cognitive" implicating the cognitive understanding of another person's point of view and the term "affective" suggesting the sharing of another person's feelings (Kalbe et al., 2007). This is consistent with several neuroimaging studies which have identified the recruitment of different brain regions, although overlapping and interacting, when engaging in ToM and empathy (e.g. Abu-Akel & Shamay-Tsoory, 2011; Völlm et al., 2006). Similarly, while the EQ was created as a self-reported measure of empathy, its creators do recognise that this questionnaire taps into cognitive and affective components of empathy, with the former effectively being representative of ToM ability (Baron-Cohen & Wheelwright, 2004). Furthermore, the same distinction was evidenced by Lawrence et al. (2004) through a follow-up study on the EQ, which provided the tools to separately analyse three factors, i.e. cognitive empathy, emotional reactivity, and social skills, within the EQ. For these reasons, the EQ has been previously used as a self-report measure of both ToM and empathy interchangeably (e.g. Baron-Cohen & Wheelwright, 2004; Groen et al., 2015; Kraemer et al., 2013; Pino et al., 2015; Seiryte & Rusconi, 2015). In agreement with Kalbe et al. (2007), I favour to use ToM as an umbrella term for inferring others' mental states, differentiating its cognitive and affective components based on the type of mental states inferred. Following this statement, I continue to consider the EQ as a self-report measure of ToM and, through my secondary analysis of the EQ factors, I investigated which components of ToM are most affected by embodiment driven by sensorimotor experience, which was found to have an enhancing effect on ToM in Study 1.

Findings from my secondary analysis on the EQ factors highlighted that such better performance seen at the EQ for people with vs without limb difference may be driven by Emotional Reactivity, rather than Cognitive Empathy or Social Skills. This suggests that the relative strength resulting from sensorimotor impairment seen in people with limb difference may be more prominent in the affective component of ToM. This result may be supported by previous studies identifying a role for bodily experience and simulation for understanding others' emotions (e.g. Adolphs et al., 2000; Heims et al., 2004). Specifically, Adolphs et al. (2000) investigated emotion recognition and reference from visually presented facial expressions in subjects with focal brain lesions and found that this ability required intact activity of somatosensory-related cortices. Therefore, the authors conclude that their findings are consistent with a simulation mechanism for representing others' emotional states, implicating the representation of other emotional states by internal generation of somatosensory representations that simulate the feelings of the observed person given their facial expression. Helms et al. (2004) identified patients with pure autonomic failure to perform worse than controls in a test of emotional attribution, highlighting a role for autonomic bodily responses in predicting the subjective emotional feelings of others, a mechanism which was speculated by the authors to be supported by empathetic emulation. Therefore, while future neuroimaging studies directly investigating the involvement of sensorimotor cortices for understanding others' mental states in people with limb difference are warranted, the supporting papers suggest a role for bodily experience and, possibly for simulation, for engaging in affective ToM. Furthermore, my results indicate that the relative strength with respect to ToM ability driven by sensorimotor impairment identified in Study 1 may extend and be more prominent in affective ToM.

To expand on the previous paragraph, the limb difference group was found to score significantly better in the Emotional Reactivity vs the Cognitive Empathy components on the EQ, while the same was not valid for the control group. These results support previous findings on elevated emotional reactivity in association with affective rather than cognitive ToM, and their relation to sensorimotor experiences (Kalbe et al., 2007). Specifically, in their paper, Kalbe et al. (2007) observed increased emotional reactivity, as assessed through skin conductance responses, in healthy adults when listening to affective-, rather than cognitive- or non-ToM stories. The authors also speculated that such emotional reactivity may be dependent on a mechanism involving simulation, according to which participants represent other people's mental states by 'simulating' their states with resonant own mental states, thus putting oneself in someone else's shoes. They oppose this to cognitive ToM stories which, not having the emotional load of affective ToM stories, would rely more on a cognitive process supported by the teleological mechanism, involving the rational modelling or inference of others' states through a system which is independent of one's own mental states (non-motor, more deliberative component of ToM). According to this paper, my findings may therefore support this differentiation by highlighting an increasing role for sensorimotor experience towards affective ToM ability, as sensorimotor impairment positively influences affective ToM scores. Similarly, these results may suggest that people with limb difference may resort more compared to controls to a simulation mechanism to understand others' minds compared to the general population. This mechanism seems to represent an advantage towards the understanding of others' mental states, as also evidenced from the better performance of people with limb difference in the SSFt compared to controls. However, how simulation can be achieved in the limb difference population, who does not share similar embodiment with the general population, will be further discussed in the general discussion of this subchapter, in light of the findings from all of my studies.

The lack of an effect in the cognitive and social components of the EQ, as opposed to the significant differences seen in the ToM and social interaction questions of the SSFt, respectively, may indicate that such components may indeed assess different constructs compared to the SSFt. Specifically, while the EQ allows the separate assessment of cognitive and affective ToM, the SSFt possibly includes the assessment of both components, thus leading to differing results. Similarly, while the EQ investigates the spontaneous use of social skills and/or a lack of intuitive social understanding, the SSFt points at the appropriateness of social interaction, which may thus represent slightly different constructs. The relationship between task measures will however be further explored in the correlation analyses included in the "Further analyses" section.

Finally, it is worth mentioning that, similarly to my findings in Study 1, the enhancing effect of having a limb difference on Emotional Reactivity scores seems to be developmental in nature, as my subgroup analysis found people with congenital limb difference to score significantly better than controls in this component, while the same was not valid for the acquired limb difference subgroup. Therefore, these results support findings from Study 1 and extend them to affective ToM. Interestingly, a significantly better performance was however seen in the congenital vs acquired limb difference group with respect to the Cognitive Empathy factor. This may indicate that individuals with congenital limb difference may rely significantly more than individuals with acquired limb difference on the cognitive process of ToM supported by the teleological mechanism (non-motor, more deliberative ToM component), as described in (Kalbe et al., 2007). See the general discussion section for the proposed mechanisms underlying the effects here observed.

Study 3: The Interpersonal Reactivity Index Questionnaire

In Study 3, I investigated and compared Empathic Concern and Perspective Taking abilities of people with and without limb difference. Similarly to Study 2, I here used implicit, self-reported measures, as assessed through the IRI questionnaire. Study 3 was conducted in an attempt to further highlight the influence of embodiment driven by varying sensorimotor experiences for ToM, with a specific interest for the Perspective Taking subscale of the IRI to explore the role of simulation for ToM.

Methods

Participants

Participants recruited for Study 2 also completed this study.

The Interpersonal Reactivity Index Questionnaire

The perspective taking (PT) and empathic concern (EC) subscales of the Interpersonal Reactivity Index (IRI) self-report questionnaire were utilised in this study. Each subscale includes 7 items, scored on a 5-point Likert scale varying from “describes me well” to “does not describe me well”. On each item, a person can score between 0 and 4 points, so both subscales of the IRI have a maximum score of 28 points and a minimum of 0. Please see Box E below for some examples of questions from the PT vs EC subscales of the IRI.

Box E. Example of questions from the PT and EC subscales in the EQ (from Davis, 1980).

Perspective taking (PT)**:**

- I sometimes find it difficult to see things from the "other guy's" point of view
- I sometimes try to understand my friends better by imagining how things look from their perspective

Empathic Concern (EC):

- I often have tender, concerned feelings for people less fortunate than me
- Sometimes I don't feel very sorry for other people when they are having problems

Procedure

This study followed the same procedure described in Study 2. However, participants completed the IRI questionnaire at this time.

Statistical Analysis

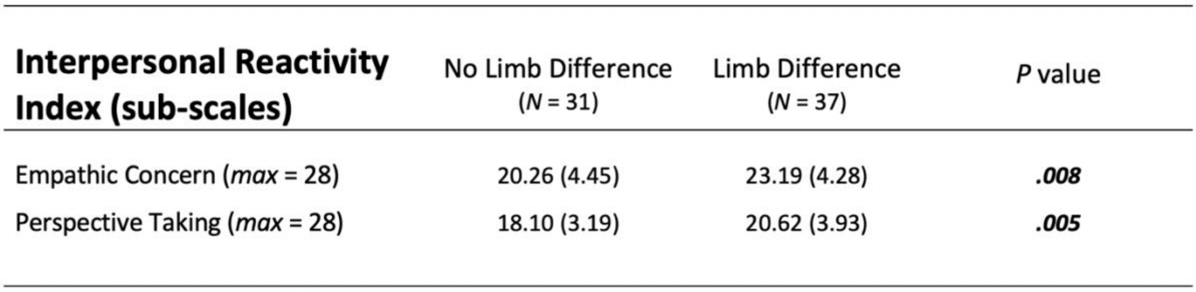
A series of two-tailed independent sample *t* tests were conducted to separately compare the scores of the IRI self-report questionnaire (with respect to the EC and IRI subscales) between the two groups of participants. Furthermore, a series of two-tailed independent sample *t* tests were conducted to determine the existence of potential differences in performance between the limb difference subgroups (congenital vs acquired) and controls with respect to these variables.

Results

The Interpersonal Reactivity Index Questionnaire

Scores in the Empathic Concern and Perspective Taking subscales of the IRI questionnaire for the limb difference vs control groups will be reported, before presenting an analysis of performance by limb difference subgroups (i.e. congenital vs acquired) and their relation to the main results.

Table 8. Participants' performance at the IRI self-report questionnaire by groups.



Interpersonal Reactivity Index (sub-scales)	No Limb Difference (*N* = 31)	Limb Difference (*N* = 37)	*P* value
Empathic Concern (*max* = 28)	20.26 (4.45)	23.19 (4.28)	***.008***
Perspective Taking (*max* = 28)	18.10 (3.19)	20.62 (3.93)	***.005***

Mean scores in the IRI subscales by groups are presented in Table 8 above. On average, participants with limb difference scored significantly higher than controls in both questions assessing the *Empathic Concern* and the *Perspective Taking* subscales of the IRI questionnaire. See Figure 11 below for a visualisation of the *Empathic Concern* and *Perspective Taking* data by group.

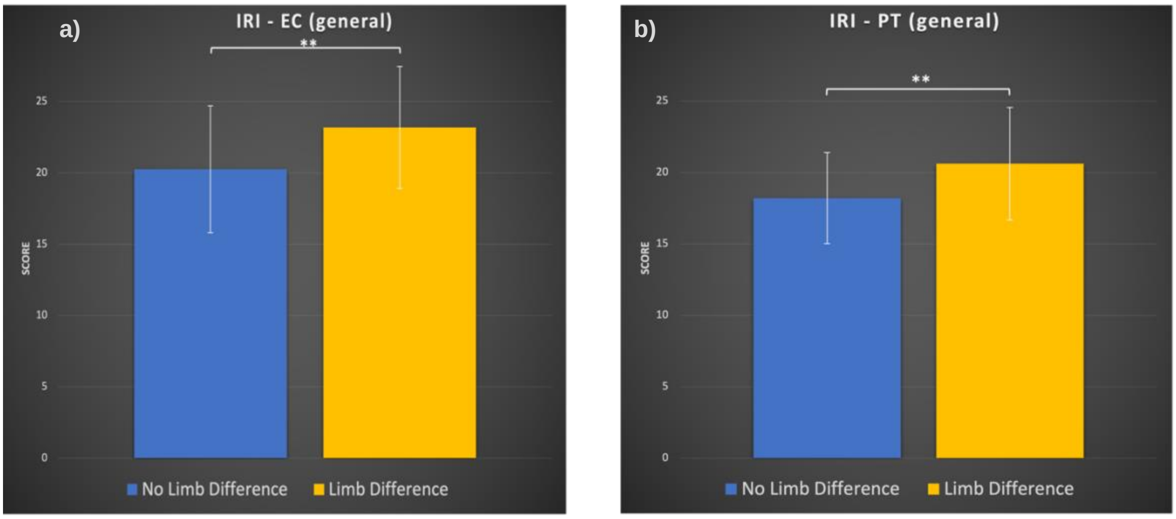
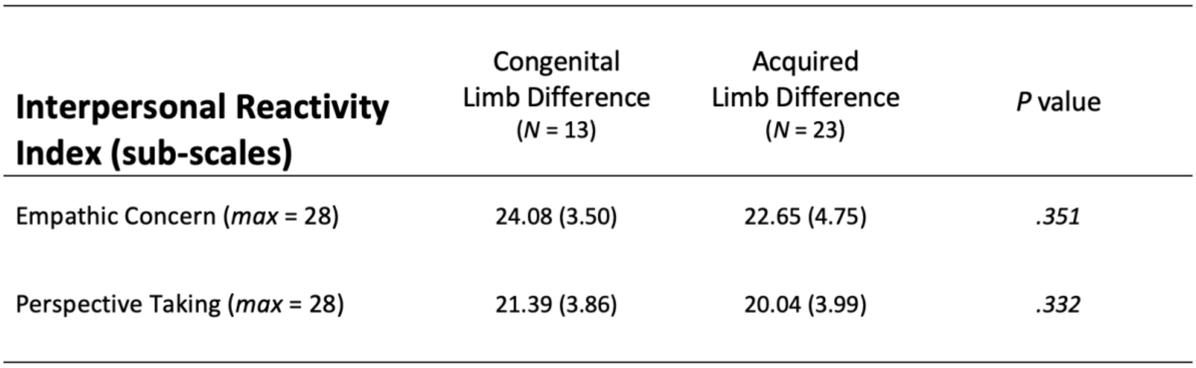


Figure 11. Comparisons of performance at the IRI-EC (a) and IRI-PT (b) questionnaire subscales between the limb difference population and controls. Error bars indicate standard deviation. IRI: Interpersonal Reactivity Index; EC: Empathic Concern; PT: Perspective Taking; **: significant at the .01 level (2-tailed).

A further analysis conducted by limb difference subgroups identified participants with congenital and acquired limb difference to score similarly in both the *Empathic Concern* and *Perspective Taking* subscales of the IRI (see Table 9).

Table 9. Participants' performance at the IRI self-report questionnaire by subgroups.



Interpersonal Reactivity Index (sub-scales)	Congenital Limb Difference (*N* = 13)	Acquired Limb Difference (*N* = 23)	*P* value
Empathic Concern (*max* = 28)	24.08 (3.50)	22.65 (4.75)	*.351*
Perspective Taking (*max* = 28)	21.39 (3.86)	20.04 (3.99)	*.332*

While these scores did not differ significantly between subgroups, the scores for the *Empathic Concern* subscale of participants with congenital limb difference (M = 24.08, SD = 3.50) were significantly higher than those of controls (M = 20.26, SD = 4.45), t(28.542) = 3.038, *p* = .005. The same was valid for scores at the *Perspective Taking* subscale (M = 21.39, SD = 3.86 vs M = 18.10, SD = 3.19, respectively), t(19.202) = 2.707, *p* = .014. See Figure 12 below for a visualisation of the *Empathic Concern* and *Perspective Taking* data by subgroups vs controls.

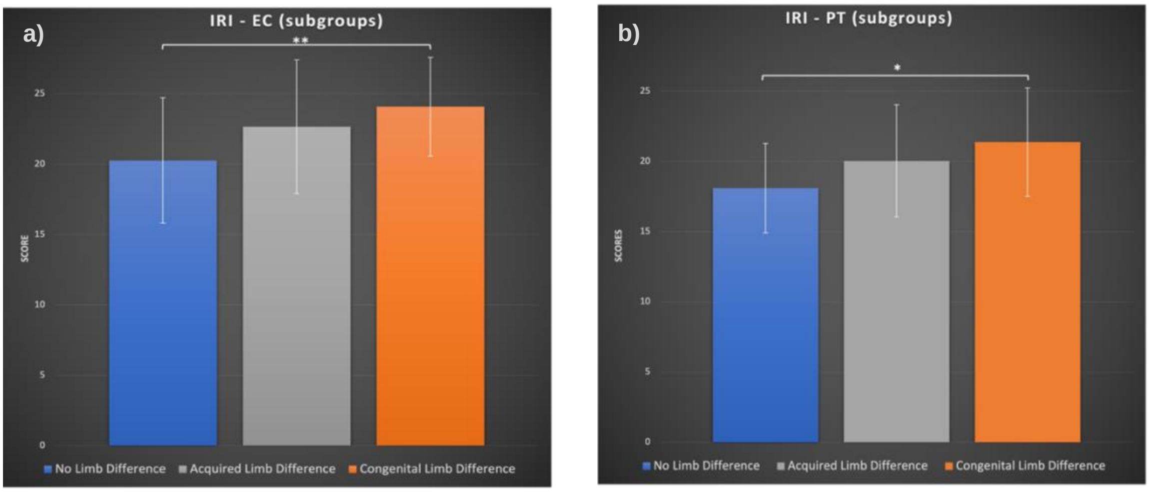


Figure 12. Comparisons of performance at the IRI-EC (a) and IRI-PT (b) questionnaire subscales between the limb difference subgroups and controls. Error bars indicate standard deviation. IRI: Interpersonal Reactivity Index; EC: Empathic Concern; PT: Perspective Taking; **: significant at the .01 level (2-tailed); *: significant at the .05 level (2-tailed).

Discussion

In Study 3, I assessed Empathic Concern and Perspective Taking abilities of people with limb difference and compared it with those of controls, in an attempt to support my findings with respect to a role for embodiment driven by sensorimotor experience for ToM. This was especially investigated by assessing the differences in perspective taking abilities between the two groups and the subgroups, which I believe would provide some insights into the plausibility of a simulation mechanism behind the enhancing effect observed resulting from having a limb difference.

Results from Study 3 suggest people with limb difference have enhanced empathy and perspective taking abilities compared to controls (linked with affective and cognitive ToM, respectively) as they were found to score higher in both subscales of the IRI questionnaire. Furthermore, once again my results indicate these effects to be developmental in nature, as the congenital limb difference subgroup outperformed the control group in both subscales, while the same was not true for the acquired limb difference subgroup. These results support my previous findings on overall enhanced ToM ability in the limb difference group, both in its affective and cognitive component, driven by a relative strength associated with impaired sensorimotor experience. Furthermore, they possibly provide support for embodied simulation as a mechanism (a) underlying ToM ability and development, and (b) driving this relative strength seen in people with limb difference.

Similarly to the SSFt and EQ, the IRI has been here successfully utilised for the first time to test differences in ToM among people with and without limb difference, thus extending its utility to investigate the influence of sensorimotor abilities (in our case driven by limb loss or deficiency) on ToM. Specifically, my results highlight a significantly better performance in people with limb difference vs controls in both the empathy and perspective taking subscales of the IRI. I will not digress on the significance of my results from the IRI-EC scale as they simply support findings and conclusions from Study 2; the correlation between measures from this study and Studies 1 and 2 will be analysed in the “Further analyses” section. In contrast, I will focus this discussion on the IRI-PT subscale. Specifically, my results are in line with Haas et al. (2015) who reported, through a neuroimaging study, higher IRI-PT subscale scores to be associated with increased prefrontal cortex and premotor activity, as well as delayed emotion attribution decisions. Specifically, by identifying a significantly different performance in the IRI-PT subscale between individuals with and without limb difference, my results support an involvement of sensorimotor experience for perspective taking and confirm that this self-report questionnaire can access this relationship.

Furthermore, findings in Haas et al. (2015) are particularly compelling for our discussion considering their implications for the simulation mechanism for understanding others' minds. Indeed, a transcranial magnetic stimulation study on healthy individuals (Balconi & Bortolotti, 2013) suggested premotor activity as responsible for embodied simulations for facial emotion recognition. Haas et al. (2015) identified the IRI-PT to be associated to prefrontal and premotor cortex activity, and in turn premotor cortex activity has been associated with the simulation mechanism for understanding others' minds (Balconi & Bortolotti, 2013). Therefore, I might speculate that the higher scores at the IRI-PT by people with limb difference vs controls may indicate an increased reliance on the simulation mechanism in the former group. Nonetheless, future studies utilising neuroimaging in the limb difference population are warranted to validate this hypothesis.

It is worth mentioning at this point that the debate on clear definitions surrounding ToM vs empathy presented above extends to perspective taking. Indeed, ToM has been previously used as an umbrella term also including perspective taking, and a cognitive and affective component for perspective taking have also been previously indicated (e.g. Hynes et al., 2006). ToM is often defined as the ability to "put yourself in someone else's shoes" or to take their perspective (e.g. Hynes et al., 2006; Jauniaux et al., 2019), thus once again making it challenging to determine the difference between ToM and perspective taking cognitive abilities. Nonetheless, perspective taking has been previously associated with cognitive ToM (e.g. Murray et al., 2017); therefore, I can conclude that the results from Study 3 extend the enhanced ToM ability in people with limb difference to the cognitive component of ToM.

Finally, my subgroup analysis identified a significantly better performance in both the IRI-EC and -PT subscales for people with congenital limb difference compared to controls. The same was not valid for the acquired limb difference group. These findings support the previously suggested developmental nature of sensorimotor-driven embodiment for ToM and extend it to the cognitive component of ToM as well, confirming results from Study 1.

Further analyses: Associations within and between groups

In this analysis, I correlated the measures used in the above studies to assess ToM and related components, in an attempt to determine (a) whether associations exist between explicit and implicit measures of ToM, as assessed through behavioural and self-report questionnaire in these studies; (b) which tools correlate with each other to identify which among the above tasks assess the same constructs, to help bring some clarity of which measures can be used to investigate similar or different components of ToM, e.g. cognitive vs affective ToM.

Methods

Participants

Participants who successfully completed all three previous studies were included in this analysis, i.e. a total of 52 participants (N = 27 with limb difference, N = 25 controls). An additional 10 and 5 participants from the limb difference and control groups, respectively, were excluded due to missing trials in the SSFt (N = 14) or incomplete attempts at the questionnaires (N = 1).

Tasks

Participants completed all three above-described tasks, which include the SSFt, as well as the EQ and IRI self-report questionnaires.

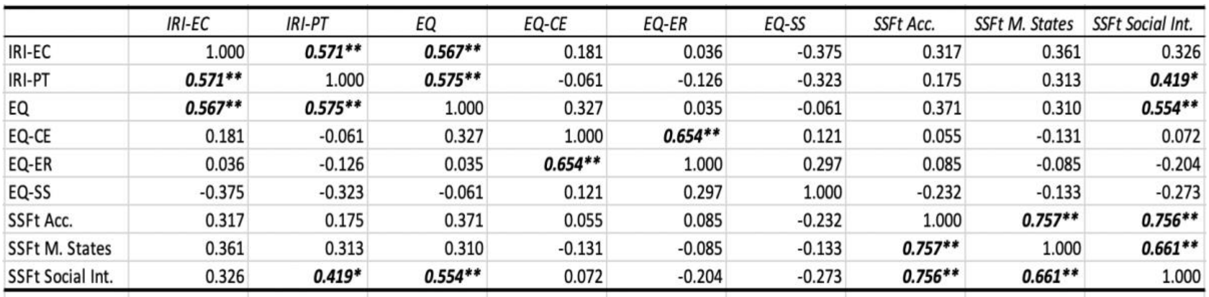
Statistical Analysis

A bivariate Pearson Correlation analysis was conducted to determine the correlation between measures in both groups, especially to identify whether a correlation between self-reported measures and behavioural measures exists.

Results

In the limb difference group, all the measures of the SSFt positively correlated with each other (Accuracy and Mental State Language: r = .757, N = 27, *p* < .001; Accuracy and Social Interaction: r = .756, N = 27, *p* < .001; Mental State Language and Social Interaction: r = .661, N = 27, *p* < .001). Furthermore, only the Social Interaction measure of the SSFt was found to positively correlate with self-reported measures, specifically with the Empathy Quotient and the Perspective Taking subscale of the IRI (SSFt Social Interaction and EQ: r = .554, N = 27, *p* = .003; SSFt Social Interaction and IRI-PT: r = .419, N = 27, *p* = .030). In addition, positive correlations were found between self-reported measures. The EQ measure was seen to positively correlate with both subscales of the IRI (EQ and IRI-EC: r = .567, N = 27, *p* = .002; EQ and IRI-PT: r = .575, N = 27, *p* = .002). Finally, no correlations between EQ and EQ factors were observed; however, a positive correlation between the Cognitive Empathy and Emotional Reactivity factors of the EQ was found, r = .654, N = 27, *p* < .001. The Social Skills factor of the EQ did not correlate with any measure. Please see Table 10 below for a visualisation of the correlation scores within the limb difference group.

Table 10. Associations between tasks within the limb difference group.



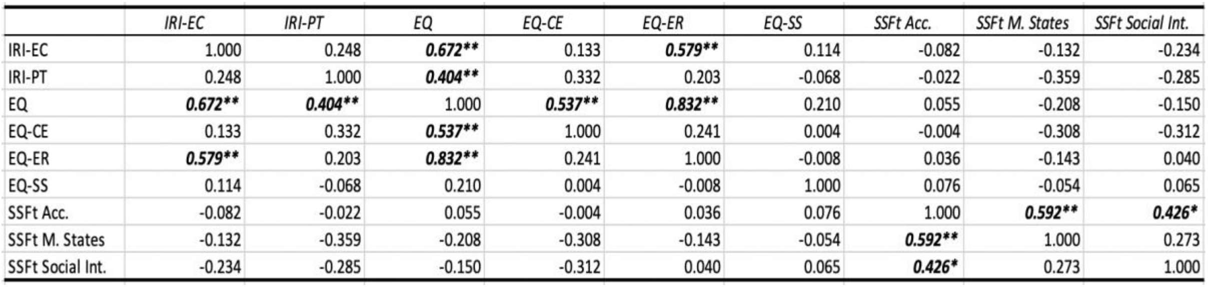
	IRI-EC	IRI-PT	EQ	EQ-CE	EQ-ER	EQ-SS	SSFt Acc.	SSFt M. States	SSFt Social Int.
IRI-EC	1.000	0.571**	0.567**	0.181	0.036	-0.375	0.317	0.361	0.326
IRI-PT	0.571**	1.000	0.575**	-0.061	-0.126	-0.323	0.175	0.313	0.419*
EQ	0.567**	0.575**	1.000	0.327	0.035	-0.061	0.371	0.310	0.554**
EQ-CE	0.181	-0.061	0.327	1.000	0.654**	0.121	0.055	-0.131	0.072
EQ-ER	0.036	-0.126	0.035	0.654**	1.000	0.297	0.085	-0.085	-0.204
EQ-SS	-0.375	-0.323	-0.061	0.121	0.297	1.000	-0.232	-0.133	-0.273
SSFt Acc.	0.317	0.175	0.371	0.055	0.085	-0.232	1.000	0.757**	0.756**
SSFt M. States	0.361	0.313	0.310	-0.131	-0.085	-0.133	0.757**	1.000	0.661**
SSFt Social Int.	0.326	0.419*	0.554**	0.072	-0.204	-0.273	0.756**	0.661**	1.000

IRI-EC: Interpersonal Reactivity Index - Empathic Concern; IRI-PT: Interpersonal Reactivity Index - Perspective Taking; EQ: Empathy Quotient; SSFt Acc.: Strange Stories Film Task - Accuracy; SSFt M. States: Strange Stories Film Task - Mental State Language; SSFt Social Int.: Strange Stories Film Task - Social Interaction; SSFt Mem.: Strange Stories Film Task - Memory; EQ-CE: Empathy Quotient - Cognitive Empathy; EQ-ER: Empathy Quotient - Emotional Reactivity; EQ-SS: Empathy Quotient - Social Skills; **: correlation is significant at the .01 level (2-tailed); *: correlation is significant at the .05 level (2-tailed).

With respect to the control group, only the Accuracy measure was positively correlated with both the other measures of the SSFt (Accuracy and Mental State Language: r = .592, N = 25, *p* = .002 and Accuracy and Social Interaction: r = .426, N = 25, *p* = .034). In contrast, no correlation was observed between the Mental State Language and Social Interaction measures of the SSFt. Furthermore, no correlation was found between the SSFt measures and any of the self-report measures. However, the EQ measure was seen to positively correlate with both subscales of the IRI (EQ and IRI-EC: r = .672, N = 25, *p* < .001; EQ and IRI-PT: r = .404, N = 25, *p* = .045), as well as both the Cognitive Empathy and Emotional Reactivity factors of the

EQ (EQ and EQ-CE: r = .537, N = 25, *p* = .006; EQ and EQ-ER: r = .832, N = 25, *p* < .001); an absence of correlation was found with the Social Skills factor of the EQ. Finally, a positive correlation was found between the Emotional Reactivity factor of the EQ and the Empathic Concern subscale of the IRI, r = .579, N = 25, *p* = .002, while the Cognitive Empathy factor of the EQ did not correlate with the Perspective Taking subscale of the IRI. Please see Table 11 below for a visualisation of the correlation scores within the control group.

Table 11. Associations between measures used to assess ToM in these studies within the control group.



	IRI-EC	*IRI-PT*	*EQ*	*EQ-CE*	*EQ-ER*	*EQ-SS*	*SSFt Acc.*	*SSFt M. States*	*SSFt Social Int.*
IRI-EC	1.000	0.248	***0.672*****	0.133	***0.579*****	0.114	-0.082	-0.132	-0.234
IRI-PT	0.248	1.000	***0.404*****	0.332	0.203	-0.068	-0.022	-0.359	-0.285
EQ	***0.672*****	***0.404*****	1.000	***0.537*****	***0.832*****	0.210	0.055	-0.208	-0.150
EQ-CE	0.133	0.332	***0.537*****	1.000	0.241	0.004	-0.004	-0.308	-0.312
EQ-ER	***0.579*****	0.203	***0.832*****	0.241	1.000	-0.008	0.036	-0.143	0.040
EQ-SS	0.114	-0.068	0.210	0.004	-0.008	1.000	0.076	-0.054	0.065
SSFt Acc.	-0.082	-0.022	0.055	-0.004	0.036	0.076	1.000	***0.592*****	***0.426****
SSFt M. States	-0.132	-0.359	-0.208	-0.308	-0.143	-0.054	***0.592*****	1.000	0.273
SSFt Social Int.	-0.234	-0.285	-0.150	-0.312	0.040	0.065	***0.426****	0.273	1.000

ToM: Theory of Mind; IRI-EC: Interpersonal Reactivity Index - Empathic Concern; IRI-PT: Interpersonal Reactivity Index - Perspective Taking; EQ: Empathy Quotient; SSFt Acc.: Strange Stories Film Task - Accuracy; SSFt M. States: Strange Stories Film Task - Mental State Language; SSFt Social Int.: Strange Stories Film Task - Social Interaction; SSFt Mem.: Strange Stories Film Task - Memory; EQ-CE: Empathy Quotient - Cognitive Empathy; EQ-ER: Empathy Quotient - Emotional Reactivity; EQ-SS: Empathy Quotient - Social Skills; **: correlation is significant at the .01 level (2-tailed); *: correlation is significant at the .05 level (2-tailed).

Discussion

In this section, I measured whether there exists a correlation between the tasks utilised in the above studies to assess ToM, in an attempt to determine (a) the relationship between implicit (behavioural) and explicit (self-reported) measures, and (b) which measures may be used interchangeably to access to the same constructs.

Overall, my findings suggest a very weak link between explicit and implicit measures, which was observed in the limb difference group only, not replicating previous findings in the literature. Furthermore, my results support an association between the EQ and both subscales of the IRI which is consistent between the two groups, suggesting that they indeed may assess the same constructs. In addition, my results made possible some further considerations with respect to the subscale measures from the self-report questionnaires and their ability to investigate the same constructs. Finally, while mixed results prevent us from establishing whether a cognitive and affective component of ToM should be considered separately, my findings do highlight an influence of differing embodiment driven by sensorimotor experience on the correlations between measures.

My findings seem to contrast previous literature suggesting a correlation between self-reported and neuropsychological measures of ToM (e.g. Spek et al., 2010). Indeed, an absence of correlations between the self-report questionnaires and the SSFt were observed in the control group, while only the Social Interaction measure of the SSFt was found to be positively correlated with both the Perspective Taking subscale of the IRI and the EQ in the limb difference group. Nonetheless, my results do not imply the invalidity of self-reported measures for assessing ToM, as all the self-reported measures led to results which converged with the behavioural task, i.e. an increased ability in people with limb difference to understand others' minds. Therefore, while I am unable to confirm that these explicit and implicit measures assess the same constructs, the similar trajectories in scores seen in all tasks of the studies included in this work make it clear that these are at least related. Furthermore, my results do partly replicate the findings from the original study in which the SSFt was first employed (Murray et al., 2017). Indeed, Murray and colleagues (2017) only evidenced a substantial partial correlation between the Accuracy measure of the SSFt and the PT subscale of the IRI in controls. Therefore, although I did not manage to replicate this finding in my control group, it does indicate an absence of strong correlations between all the SSFt subscales and self-reported measures, suggesting that this issue may be task specific. Overall, although an association between explicit and implicit measures was not evident in my results, I do believe that the present work shows the value of both types of measures in assessing ToM ability.

Nevertheless, in my study I did observe a positive correlation between the Social Interaction question of the SSFt and the PT subscale of the IRI, which was not present in Murray et al. (2017). Interestingly, this correlation was only found in the limb difference group. Considering that the Interaction measure of the SSFt was also found to be correlated in Murray et al. (2017), although with respect to the EC subscale of the IRI in individuals with ASD, it may suggest that this measure may be more sensitive to external factors influencing individuals' ability to understand others' minds. The fact that in my study the Social Interaction measures correlated with the EQ and, particularly, with the PT measure of the IRI only in the limb difference group may suggest an increased ability for people with limb difference in self-reflection and a stronger relation between perspective taking and social interaction compared to controls. While this remains a speculation, it possibly may highlight that people with limb difference weigh perspective taking differently to controls during social interaction.

Finally, my results make it clear that some of these measures may indeed point to the same constructs and contribute to the debate on the overlap between empathy, ToM and perspective taking. Specifically, in accordance with previous literature (e.g. Lawrence et al., 2004), positive correlations between the EQ and IRI subscales were found in both groups, providing evidence in support of the common underlying construct assessed. This is not surprising, as the EQ includes both questions associated with empathic concern and perspective taking. Furthermore, items in both the EQ and IRI have been previously differentiated into two groups representative of an affective and cognitive component of ToM. Specifically, the Cognitive Empathy factor of the EQ and the PT subscale of the IRI have been associated with the cognitive component, while the Emotional Reactivity factor of the EQ and the EC subscale of the IRI have been associated with the affective component (e.g. Lawrence et al., 2004; Murray et al., 2017). This differentiation is supported by my results in the control group reporting the absence of an association between the Cognitive Empathy vs Emotional Reactivity factors of the EQ, as well as the EC vs the PT subscales of the IRI. These results thus suggest that the two questionnaires may indeed be able to access an affective vs cognitive component of ToM through their subscales. In contrast, results from the limb difference group indicate a positive correlation between the affective and cognitive components of both the EQ and IRI questionnaires, suggesting that such a differentiation may be redundant. Furthermore, results from the control group report an association between the affective subscales of the two questionnaires (Emotional Reactivity factor of the EQ vs the EC subscale of the IRI), thus indicating that they may indeed assess the same affective component, separate from the cognitive one. In contrast, a correlation between the cognitive subscales of the two questionnaires (Cognitive Empathy factor of the EQ vs the Perspective Taking subscale of the IRI) was not found in this group. Furthermore, and interestingly, these associations did not maintain in the limb difference group, where correlations were not observed between neither the affective nor the cognitive subscales of the two questionnaires. Overall, these results lead us to two conclusions. First, findings from the control group may support the differentiation of an affective and cognitive component of ToM, although suggesting the need to revise the subscales currently used to assess such different components. Second, findings from the limb difference group suggest this distinction to be redundant. While the last remark seems to be in contrast with my previous statement, this different pattern of associations among individuals with limb difference compared to controls may suggest the presence of specific factors in this population resulting in an interaction between the affective and cognitive components of ToM. This ultimately seems to be advantageous given their higher scores in all tasks. Based on my previous studies' findings I speculate that this effect may be driven by their different embodiment with respect to the majority of the population due to their sensory impairment. I will discuss this point further in the general discussion section below.

General Discussion

Based on findings from the above series of studies, I propose that individuals whose sensorimotor bodily experiences differ substantially from the majority of the population in virtue of their body differences represent others' mental states in a significantly distinctive way.

Specifically, in Study 1, I identified enhanced ToM ability in people with limb difference vs controls as assessed through the behavioural SSFt in the realms of intention recognition, mental state language and social interaction. This effect was found to be developmental in nature. Indeed, while it seems to be critical at birth, it persists throughout development and adulthood, showing some flexibility. I further suggested this enhancing effect to specifically impact ToM-related components, i.e. intention recognition and mental state language, rather than social cognition in general, given the similar performance between subgroups in the Social Interaction question of the SSFt. Indeed, the lack of this effect was further supported by the similar scores between groups in the Social Skills factor of the EQ. Although the correlation analysis in the “Further analyses” section did not find an association between the Social Interaction question and the Social Skill factor of the EQ in either of the groups, thus indicating that they possibly assess different constructs, an absence of the effect on both these social components is evident. Finally, I indicated that the motor component of sensorimotor experiences may specifically impact ToM development and ability, either leading the way to the simulation mechanism supporting ToM or highlighting a supplementary role for motor experiences for ToM in the limb difference population.

This advantageous effect on ToM of having a limb difference persisted in Study 2, employing the EQ questionnaire, i.e. a self-reported explicit measure of ToM. The results from this study confirmed that this effect is therefore not specific to one type of measure (either implicit or explicit) and indicated consistency between self-awareness and actual ToM ability in both groups. Furthermore, Study 2 extended the results from Study 1 by identifying that the enhancing effect may have a specific impact on affective, rather than cognitive, ToM, and that it may involve the simulation mechanism. Finally, this study confirmed the developmental nature of this enhancing effect, which extends also to the affective component of ToM.

These results were replicated also in Study 3 using the IRI questionnaire (another self-reported, explicit measure of ToM), whereby I also found an enhanced affective ToM in people with limb difference vs controls. The effect reported in Study 3, however, extended to cognitive ToM. Indeed, higher scores at the PT subscale of the IRI were seen in the limb difference population, which is representative of cognitive ToM. Results from Study 3 also support the presence of a simulation mechanism behind such an enhancing effect. Finally, this study confirmed the developmental nature of this effect, which extends also to the cognitive component of ToM.

Finally, the "Further analyses" section indicated that my results do not support a link between explicit and implicit measures, suggesting that they may indeed assess different constructs, only partially replicating previous literature. Furthermore, these results seem to partially support the distinction between affective and cognitive components of ToM, while their assessment using the identified subscales of self-reported measures is less clear. Nonetheless, my results indicate that such associations between measures, and thus the distinction between affective and cognitive ToM, becomes blurred in people with limb difference, with the two components interacting with each other. This seems to be ultimately advantageous, given the higher scores, and I speculate may be driven by differing embodiment that people with limb difference experience given their sensorimotor impairment.

Overall, embodiment driven by sensorimotor experience seems to influence ToM ability and development, resulting in a relative strength in people with impaired sensorimotor abilities. Furthermore, the influence of sensorimotor-driven embodiment for ToM seems to be most crucial at birth, however it persists throughout adulthood, showing flexibility.

As highlighted in the introduction of this chapter, cognitive embodiment proposes that sensorimotor interactions, which also include the physical characteristics and constraints of an individual's body, shape embodied mental representations (Pezzulo et al., 2011). Here, I provided empirical evidence of an involvement of differing embodiment (in virtue of differing sensorimotor experiences) in ToM ability and development.

Interestingly, individuals with limb difference outperformed the general population in all ToM tasks, which may seem counterintuitive considering their sensorimotor impairment. Therefore, the rest of this section will discuss some candidate mechanisms by which embodiment driven by impaired sensorimotor experiences may actually result in increased ToM ability. Specifically, I speculate on (1) ”imaginary simulation”, (2) “teleological for mentalising”, (3) “self-other blurring” and (4) “increased self-other distinction” as possible mechanisms underlying the effect observed in this study. Crucially, while I will outline each mechanism separately, these may not be mutually exclusive and may instead interact with each other.

Imaginary simulation

The plausibility of the simulation mechanism behind ToM ability has been previously contested given its requirement for self-other similarity. The flexibility of this requirement is debated, and it is unclear the extent to which others’ representations which differ from self-representations (including mental states) can be understood through simulation.

In light of this, my studies’ findings would not seem to support a simulation mechanism behind the enhanced performance of people with limb difference in ToM tasks. Indeed, self-other similarity may not apply in the limb difference population given that their self-sensorimotor representations differ from the majority of the population or are missing altogether. According to a simulation perspective, this should lead to poor ToM. Specifically, we would expect ToM development to be impaired in people with limb difference if relying on a simulation approach lacking flexibility, as they would not be able to represent others' differing representations throughout development. This would in turn result in an inability to understand other people's minds and to perform well in ToM-related tasks.

Following this reasoning, I would expect in my results that having a limb difference negatively affects performance in the utilised ToM-related tasks, with the congenital limb difference group showing the worst performance. This is however in contrast with my results, that instead show that individuals with limb difference perform better in ToM tasks. This may imply that (a) the simulation account may be more flexible than previously thought and that it can account for self-other dissimilarities (possibly through compensatory mechanisms). Alternatively, my results may indicate (b) the presence of another mechanism underlying enhanced understanding of others' minds in the limb difference population. I will first discuss interpretation (a) given the other findings from my studies pointing to potential enhanced simulation in the limb difference population, in turn resulting in enhanced ToM. In the next section ("*Teleological for mentalising"*), I will discuss interpretation (b) as a candidate explanation, in alternative to or coexisting with simulation, for the better performance by the limb difference population here observed.

My previous results showing enhanced perspective taking and emotional reactivity in people with limb difference support an involvement of simulation towards the enhancement of ToM ability in people with limb difference. Indeed, perspective taking and emotional reactivity have both been previously associated with this mechanism (see discussions of Studies 1, 2 and 3 for more details). Furthermore, the simulation mechanism seems to be supported also by the general finding of the involvement of embodiment driven by sensorimotor experiences in ToM, considering that simulation has been previously suggested to be the mechanism underlying cognitive embodiment (see Introduction). Therefore, the question to tackle to provide plausibility to simulation as the mechanism underlying enhanced ToM ability in people with limb difference is the following;

"How can people with limb difference engage in embodied simulation for understanding mental states of others whose body substantially differs from theirs?"

Previous studies on the limb difference population have suggested that neural representations of the body, in addition to prior experience of sensorimotor-driven embodiment, rely on observation of full-bodied others through the simulation mechanism (Brugger et al., 2000; Price, 2006). Specifically, a review of the literature on phantom limbs (Price, 2006) indicated two stages in the development of body image: (a) in utero as a result of spontaneous muscular activity and proprioceptive feedback, and (b) during the first decade of life (and later when phantoms are induced by prostheses) through various modalities, including vision and touch. Price (2006) links the simulation mechanism to such body images acquired through the observation of others and suggests them to be linked to phantoms. Similarly, Brugger et al. (2000) reported that data in the literature indicate that sensorimotor representations in the brain can be developed without having such body parts. The authors attribute this effect to both genetic and epigenetic factors, including the habitual observation of other people moving their limbs, which may contribute to phantom limb experience in people with congenital limb difference. Finally, Aziz-Zadeh et al. (2012) found activation of brain areas associated with simulation, in addition to mentalising, in an individual with congenital limb difference when observing an action that she could not reproduce herself with her own body. The authors reason that the activation of the simulation network when observing actions out of someone's motor repertoire may be resulting from observation learning. Indeed, the subject had extensive visual and conceptual knowledge of the observed actions, which were also part of the normal human repertoire of actions. Therefore, according to this research, simulation through observation may be implicated in understanding others' minds and actions also in absence of self-other similarity. Furthermore, the research outlined above indicates that this mechanism may be able to activate sensorimotor cortices even in the absence of limbs and result in sensorimotor-driven embodiment. Given the above, we could thus speculate that the simulation mechanism may be flexible enough to have allowed participants with limb difference to engage in embodied simulation and understand mental states of individuals from the general population, i.e. others who differ from the self.

Furthermore, previous studies have introduced the concept of motor imagery (Gandola et al., 2019; Saruco et al., 2019), that is the "mental representation of an action without engaging its actual execution" (Saruco et al., 2019, p. 634). Mental simulation through motor imagery has been shown to activate sensorimotor networks even in the absence of (and which are comparable to) explicit motor outputs (Malouin & Richards, 2010; Saruco et al., 2019). Therefore, its training has been associated with counteracting the effects of missing or lost limbs and enhancing motor performances, both in people with upper and lower limb differences (Gandola et al., 2019; Malouin & Richards, 2010). Therefore, in accordance with the studies cited in the previous paragraph, these studies suggest the possibility of understanding others' minds by mental simulation also in the absence of body parts related to the observed person, which may support engagement in simulation in the limb difference population in my studies. I suggest that this extra process that individuals with limb difference require in order to interpret the world from the perspective of the majority of the population may lead to a relative strength, resulting in enhanced ToM. Finally, motor imagery ability was shown to be decreased following limb amputation or temporary disuse, e.g. limb immobilisation, and this was associated with a weaker mental representation of actions which was highly modulated by sensorimotor inputs (Malouin & Richards, 2010). This indicates that limb loss does not prevent motor imagery, but that it only makes it more difficult, which may explain the difference in performance seen between the congenital and acquired limb difference groups.

Nevertheless, my studies did not include a direct measure of visual experience and associated cortical activation in people with limb difference. Future research should take this into account to explore the hypothesis of this imaginary simulation as an extra step to understand the minds of others who have different embodiments and sensorimotor experiences. A further way to validate my results would be to conduct neuroimaging studies to determine whether brain areas associated with simulation, in addition to mentalising, are effectively recruited in participants undergoing the tasks used in this investigation to assess ToM.

Teleological for mentalising

The enhanced performance in individuals with limb difference may be attributed to a mechanism other than simulation, i.e. teleological for mentalising. This suggestion however does not exclude simulation as one of the mechanisms underlying ToM and present in both the general and limb difference populations. More in detail, the characteristic enhanced ToM effect seen in the limb difference population may be interpreted in terms of a supplementary role for motor experiences towards ToM, given the sensorimotor impairments seen in this population. Instead, my findings may be understood as highlighting a crucial role for the non-motor, more deliberative component of ToM, that I indicate here as *teleological for mentalising*.

As previously mentioned in Part 1 of this thesis, the teleological theory proposes that outcomes of actions can be recognised as their goals only if they are performed efficiently (Csibra & Gergely, 2007), and teleological reasoning has been previously indicated as a precursor of ToM (Csibra & Gergely, 2007). This stance is supported by studies in the literature showing increased activation of brain areas associated with mentalising (e.g. TPJ, mPFC) in typical adults in response to irrational vs rational actions (e.g. Brass et al., 2007; Marsh et al., 2014). These studies overall suggest that mentalising can be seen as a rationalisation of behaviour. More specifically, Marsh et al. (2014) suggested that the mPFC may be associated with rationality resolution, while the TPJ may be specifically involved in mentalising about the reasons and intentions underlying unusual behaviour. Indeed, the authors found both enhanced mPFC and TPJ activity in typical adults in response to the observation of scenarios involving irrational actions.

This mechanism is interesting and relevant for our discussion as irrationality is intended as "inefficient actions, given the environmental constraints" (e.g. Marsh et al., 2014, p. 82). If we applied this approach to describe the experiences of the limb difference population, such irrationality may be seen as "inefficient actions, given the *bodily* constraints" presented by this population. Therefore, it seems valid to speculate that mentalising based on a teleological mechanism may be applied by individuals with limb difference to understand mental states of others who differ from the self, rather than sole direct motor simulation. This interpretation seems to be supported by previous studies (Aziz-Zadeh et al., 2012; Cusack et al., 2012) that identified the activation of mentalising-related, in addition to simulation-related, brain areas in individuals with limb difference in response to the observation or imitation planning of unusual actions.

If we follow this rationale, it becomes clear why mentalising based on a teleological mechanism may result in enhanced ToM: individuals with limb difference are more likely to engage more often in rationalisation of behaviour when interacting with the general population; in other words, they are more trained in this non-motor, more deliberative component of ToM. In contrast, direct motor simulation may in a way "restrict" ToM ability, by constraining the possible interpretations of others' minds based on simulation on self-motor control policies and programming. Therefore, the motor component of ToM may interfere with or overtake the non-motor component of ToM, thus preventing or reducing the training of the latter in the general population.

The above interpretation may also explain the critical window during development observed for the enhancing effect of sensorimotor-driven embodiment on ToM. Specifically, from a developmental point of view, this assumption of motor simulation "restricting" ToM ability, as well as the differences seen in the congenital vs acquired vs control groups, may be explained through the process of "perceptual narrowing" in infancy. Briefly, perceptual narrowing is a developmental process which permits infants to narrow down the processing of sensory and perceptual inputs to those that are most relevant for their socioecological environment (Lewkowicz & Ghazanfar, 2011). This is considered a crucial part of the developmental process of humans and an adaptive response that permits infants to tune their sensory/perceptual abilities to best match their ecological setting (Lewkowicz & Ghazanfar, 2011). Therefore, when the concept of perceptual narrowing is applied to the mechanisms here discussed, we may speculate that infants, both with and without limb difference, may at first share the same representational framework of others' mental states. Successively, in the general population, the framework of reference may be narrowed down by motor simulation, given that their sensory and perceptual abilities are representative of most of the population. In contrast, individuals with limb difference may maintain such a general, shared representational framework of others' mental states, given their differing sensory/perceptual experiences compared to the general population. Ultimately, this may result in a more abstract and less constraint approach to understanding minds of others who differ from the self, through a shared representational framework, which ultimately may lead to such enhanced ability. This would be especially true for individuals with congenital limb difference, who never had any sensorimotor experience.

The above interpretation can be supported by the "like me" account for social cognition proposed by Meltzoff (2007b, 2007a), which points to the presence of a shared representational framework for understanding others and their mental states, rather than inferential processes driven by an initially solitary representation of the self, as suggested in other contrasting theories (e.g. see Piaget, 1952, 1954). Crucially, the "like me" framework can engulf the simulation mechanism, as it is more abstract and does not make any assumptions on the type of mechanism using the code (see also Part 3 of this thesis for more details on the "like me" assumption and simulation).

Nonetheless, I would like to highlight that this mechanism does not exclude the additional engagement in simulation for ToM; thus a potential role for simulation in the limb difference population. Indeed, in the previously described studies by Cusack et al. (2012) and Aziz-Zadeh et al. (2012), mentalising was recruited in individuals with limb difference *in addition to simulation* to, respectively, imitate and understand actions of individuals from the general population. Therefore, the two mechanisms may be interacting to achieve ToM.

Future studies involving neuroimaging in line with Cusack et al. (2012) and Aziz-Zadeh et al. (2012) are warranted to determine whether the findings of a mentalising mechanism, in addition to simulation, observed in action understanding and imitation of others who differ from the self could also be extended to the understanding of their minds. Furthermore, another interesting future study may be to compare ToM ability in infants with and without limb difference and monitor this throughout development, in an attempt to investigate this perceptual narrowing hypothesis for ToM. Finally, it should be explored whether these mechanisms are specific to the limb difference population or whether they extend to the general population when interacting with people with different bodies from their own, e.g. with people with limb difference.

Self-other control

The other two mechanisms which I propose may have resulted in enhanced ToM ability in people with limb difference both point to the same ability, that is self-other control. However, they are somewhat opposite and lead to different implications. In the paragraph below, self-other control will be first introduced. Two separate subsections will then discuss the two mechanisms possibly resulting in enhanced ToM.

To engage in successful ToM, a balance between self- and other-representations is required. When this balance is off, two scenarios arise. On the one hand, to engage both in cognitive and affective ToM, people are required to represent others’ mental states, thus to have an other-perspective. In order to accurately represent the world from the other-perspective, however, self-representations need to be inhibited, as they may at times be in contrast with the reality as seen from the other-perspective. Failure to inhibit self-perspective results in an Egocentricity Bias (EB), which leads to biased predictions about the others' mental states, especially in cases of incongruence between self- and other- states. This EB has been previously seen in children younger than 3 years of age, who are not successful at explicit ToM tasks as they fail at inhibiting their own representation of the world to correctly describe the false beliefs or wrong representations of the world from the other-perspective. Similarly, Riva et al. (2016) suggested an increasing difficulty to detach from self-perspective with aging, driven by a heightened emotional EB in the older population.

On the other hand, self-other balance can be impaired by the other-perspective overriding self-perspective, leading to self-other blurring (Lamm et al., 2016). Specifically, self-other blurring happens when the other-representations are not inhibited in favour of self-representations, which is generally considered as an adaptive and self-protective response. This leads instead to a lack of distinction, and thus confusion, between self- and other-perspective. An example is provided by individuals with mirror-touch synaesthesia, who experience tactile sensations on their body by only observing someone else being touched (Fritsch et al., 2021). Similarly, self-other blurring can happen in affective ToM, leading to greater emotional reactivity, accompanied with increased personal distress and emotional contagion (Lamm et al., 2016; Ward et al., 2018). In contrast, a clear distinction between self and other would prevent excessive personal distress from another's negative affective state. To conclude, self-other control, that is a balance between self- and other-representations, is needed to better understand others' minds. Supporting evidence towards this stance is also provided by Fritsch et al. (2021) who trained individuals with mirror-touch synaesthesia to increase self-other control, which resulted in higher empathy scores both in behavioural and implicit tasks.

Self-other blurring

The mechanism I propose here to possibly underlie enhanced ToM ability in people with limb difference seen in my study results stands if absent (or weak) self-representations (related to sensorimotor experiences) are accepted in this population. Specifically, I consider these lack (or weak) self-representations to lead to self-other blurring which, perhaps counterintuitively, may constitute an advantage in this population towards understanding others' minds.

A possible involvement of self-other blurring in the enhanced ToM ability seen in the limb difference population in my studies is supported by my results. Specifically, my Study 2 results suggest increased emotional reactivity in people with limb difference, which has been previously associated with self-other blurring (Ward et al., 2018). We may speculate that increased self-other blurring may be a result of two possible factors, i.e. people with limb difference (1) having no or (2) weak self-representations of sensorimotor experiences. According to the first factor, an absence of self-representations would in turn (a) prevent EB, considering that they would not have a self-perspective reference that can bias that of others; and (b) increase emotional reactivity, considering that they would not have a self-perspective to inhibit and contrast that of others; therefore, they may more readily take the other-perspective. According to the second factor, self-representations may be formed through e.g. imagery simulation (as described above) or prosthesis use (self-model creation through embodiment of prosthesis and sensorimotor representations (Fritsch et al., 2021)); however, these self-representations may be weaker. This would in turn result in (a) weak EB and (b) higher emotional reactivity compared to the general population, ultimately leading to improved ToM.

Overall, it is clear that decreased EB and increased emotional reactivity can result in enhanced ToM, and thus could support my findings in the limb difference population. However, as mentioned earlier, self-other blurring has also been associated with higher emotional contagion and personal distress, which may in turn hinder ToM ability. Indeed, individuals may reduce their engagement with others as an adaptive, protective response to prevent personal distress. Nonetheless, this adaptive measure may not need to be taken by people with limb difference, if they lack (or have weak) self-representations of states observed in others. Indeed, as a consequence, they would not be able to translate such states onto their own experiences, preventing this way emotional contagion and personal distress. This view may be supported by Aziz-Zadeh et al. (2012) who investigated pain observation in an individual with congenital limb difference in response to photos of a person without limb difference. The authors found that when the individual with congenital limb difference observed others' pain in a part of the body that she did not have herself, she was found to activate brain areas related to pain processing, e.g. insula, whereas somatosensory cortices were not recruited. This result suggests that individuals with limb difference can understand and are able to empathise with the mental state associated with pain of another person who differs from the self. However, they may not be able to map the person's pain onto their own somatosensory body representations. According to the authors, this may be due to the lack of sensory representations to process the pain in a localised bodily region. Furthermore, remarks from both Klimecki et al. (2013) and Ward et al. (2018) provide another explanation as to why self-other blurring may not always hinder ToM ability. Specifically, Ward et al. (2018) suggested that self-other imbalance and resulting dominant emotional reactivity could lead to both positive or negative outcomes for social behaviour, based on social context and/or variation in the coping strategies used by an individual. For example, Klimecki et al. (2013) demonstrated that socio-affective training, in the form of compassion training, represents an effective coping strategy that favours positive affect as opposed to personal distress following empathy. Therefore, some sort of coping strategy may be used by individuals with limb difference in response to increased emotional reactivity, which would lead to decreased emotional contagion and personal distress when engaging in such enhanced ToM.

Coping strategies were not directly addressed in these studies; it would be interesting to conduct future studies assessing coping strategies in both the limb difference and general population to determine whether the former have higher experience with this training and use it to modulate their responses to others' states during social interactions. In addition, further research on the extent of self-representations with regards to missing sensorimotor experiences is warranted to determine whether self-other blurring may be happening in people with limb-difference during social interactions or when trying to predict others' mental states.

Increased self-other distinction

The final mechanism which I propose to possibly underlie the enhanced ToM ability in people with limb difference points to self-other distinction, which is in contrast with my previous proposal. Specifically, I here accept the presence of self-representations in people with limb difference, which are very different compared to other-representations (related to the general population), and thus result in heightened self-other distinction.

Two possible interpretations result from this position. First, people with limb difference may not need to suppress self-perspective, as there is a clear distinction between self-representations and those of others, provided their substantially different sensorimotor-driven embodiment. According to this stance, (a) EB would not be an issue in the limb difference population, resulting in more accurate and faster understanding of others' minds; and (b) self-other blurring would not happen, provided that their substantially different representation of the self would not be confused with those of others. Second, people with limb difference may still require suppressing their self-perspective in order to achieve correct ToM, according to the self-other balance principles described above. However, the same interpretations in accordance with the first stance would be valid in this case, given their still distinctive representations of the self. Specifically, (a) EB would not present an issue in the limb difference population, which may be better at suppressing self-related representations in order to take the other person's perspective; and (b) self-other blurring would be avoided. A previous study by Aziz-Zadeh et al. (2012) could be interpreted as supporting the view of a heightened self-other distinction in people with limb difference. Specifically, the authors investigated action observation in an individual with congenital limb difference when observing an individual without limb difference. The authors observed activation in brain areas associated with simulation, including the premotor cortex and the IPL, in response to the observation of actions that the individual with congenital limb difference could execute herself, although using different effectors (i.e. different body parts) to complete the action. This result therefore highlights a clear distinction between self- and other- representations, which do not overlap and are maintained separate when understanding the behaviour of others who differ from the self.

Nonetheless, when accepting any of these two interpretations, it seems possible that limb difference, i.e. atypical embodiment driven by sensorimotor impairment, may lead to an increased self-other distinction and thus to enhanced ToM in this population. Previous studies have assessed the extent of self- vs other-representations during social cognition. For example, Jackson et al. (2006) investigated brain activation in healthy individuals in response to images of people in pain when asked to take self- or other-perspective and found differing, although overlapping, activation patterns associated with self- and other- pain processing. In line with this, it would be interesting for future research to investigate the extent of self-other distinction in people with limb difference when observing someone with a substantially different sensorimotor-driven embodiment to determine whether this is higher compared to the general population.

Limitations

The limitations of these studies will now be outlined; these are valid for all studies above presented, unless otherwise specified. First, details on the age of participants in the limb difference group were not collected; therefore, it was not possible to age-match control participants (although all participants were +18 years old). Considering that this may result in a possible impact of age gap on my results due to differing maturity and thus emotional, social and cognitive skills (including ToM), future studies comparing participants of the same age group are warranted to further validate my findings. Second, participants' verbal abilities and intelligence were not measured in this study, as these were not at the core of my research question and have not been previously included in all tasks here under investigation. However, these factors may have contributed to my participants' responses; therefore, readers should take this into consideration when interpreting my study results. Nonetheless, I believe this to be a question to be answered by studies interested in the interaction between language and/or intelligence and the development of ToM-related skills.

Therefore, I hope that my results can inspire future studies in this direction. Third, given the online nature of these studies, it was not possible to control participants' motivation and attention during all tasks, both within and between groups. However, given that this issue was present for the assessment of both groups, I consider a potential effect of these factors to be distributed amongst participants of both groups; thus, not to have greatly influenced my overall findings. Nonetheless, it would be interesting to conduct investigations in the future in the form of face-to-face interviews (as in the original SSFt study by Murray et al., 2017) to determine whether my findings will be maintained. Fourth, some of the measures utilised in these studies (e.g. empathy and perspective-taking measures) were obtained from self-report questionnaires and I cannot exclude that participants may have not accurately reported their abilities, or that participants' idea of their abilities may not reflect reality. However, this is a limitation inherent to all self-report questionnaires (Demetriou et al., 2015). Nonetheless, results from the self-report questionnaires support my other findings from behavioural measures, suggesting that they may be accurate enough to access participants' abilities. Fifth, previous studies suggested gender to impact measures of empathy utilised in this study (e.g. EQ: Baron-Cohen et al., 2003; Lawrence et al., 2004; EC-IRI: Krämer et al., 2010); therefore, my results may have been influenced by gender differences. Unfortunately, gender data was not collected in all participants; therefore, future studies investigating the effect that gender may have on these results are warranted. Finally, a power analysis confirms the suitability of my sample size for comparison between the limb difference and control groups (with a large effect size), which was bigger than that used in the original study (Murray et al., 2017). However, my sample size was too small to allow conclusive remarks with respect to my findings related to the limb difference subgroups (congenital vs acquired). Therefore, such findings can only be considered preliminary results which I aim to confirm in future studies.

Conclusions

To summarise, this series of studies identified enhanced ToM ability in people with limb difference when compared to the general population, with the congenital limb difference subgroup driving this advantageous effect. I suggested these findings indicate a role for sensorimotor-driven embodiment towards ToM ability and development, which may be critical at birth and flexible throughout adulthood, with sensorimotor impairment playing a part towards improved ToM. Furthermore, I showed this effect to be present in both the affective vs cognitive components of ToM, as well as in explicit vs implicit measures of ToM. Finally, I proposed four candidate mechanisms by which people with limb difference may present enhanced ToM ability, in virtue of their differing sensorimotor-driven embodiment, which are not mutually exclusive and could be interacting with each other.

To conclude, I would like to bring back our attention to one of the current questions to be tackled by research on cognitive embodiment raised by very influential researchers in the field:

"How, and how much, are sensorimotor and simulative processes reused for cognitive tasks" and how flexible are they (Pezzulo et al., 2013, p. 10)

I believe to have contributed to this topic with my findings, showing a relationship between sensorimotor-driven embodiment and ToM, which I propose may be mediated by the simulation mechanism, as well as other mechanisms such as teleological for mentalising; and I discussed the flexibility of such processes. I hope to inspire future investigations in this same direction.

3.2 False belief understanding in individuals with limb difference

Introduction

In the present research study, I once again focused on the limb difference population in an attempt to investigate ToM ability from a developmental point of view. In particular, this study focused on three objectives: (a) assess the role of mental rotation and perspective taking abilities for ToM in individuals with limb difference vs controls; (b) provide insights into the relationship between explicit and implicit ToM; and (c) investigate the feasibility of conducting the study online and inform laboratory studies. Given the Covid-19 restrictions preventing face-to-face experiments, I adapted the task developed by Xie et al. (2018) for online testing. Furthermore, I was unfortunately not able to address the first objective of this study, while the second could only partly be addressed. I will briefly provide some more details on the research questions and this task, prior to discussing my methodology and results.

First objective: mental rotation and perspective taking for ToM

This study aimed to identify whether mental rotation and perspective taking abilities are fundamental for engaging and developing ToM, through the assessment of individuals with limb difference. Specifically, it aimed to investigate if sensorimotor impairment seen in individuals with limb difference would impact their ToM ability as compared to controls. I refer the reader to Part 2, chapters 2.3-4 of this thesis for a detailed discussion on the rationale behind this study. Importantly, examining these abilities in individuals with limb difference can elucidate trajectories of change. This is because this population allows for a direct comparison between individuals with congenital vs acquired limb difference.

Unfortunately, due to the small size of the recruited limb difference population and the online nature of the study, I was unable to conduct any comparison between individuals with limb difference and controls. As a result, it is not possible – from the study's findings – to advance any conclusion about the role of mental rotation and perspective taking for ToM development and ability as assessed via sensorimotor impairment. Nevertheless, I report in this chapter the data obtained from 4 participants with limb difference and from the control sample (N = 14) in an attempt to inform and inspire future studies in this direction.

Second objective: explicit vs implicit ToM

This study also aimed to investigate the relationship between explicit and implicit ToM, by comparing participants' behavioural data (keyboard responses) with eye-tracking data (online webcam eye-tracking). While it is generally accepted that ToM is present by 4 years of age (Wellman et al., 2001), some studies indicated the presence of an implicit ToM from infancy (e.g. Onishi & Baillargeon, 2005; Southgate et al., 2007). Specifically, via assessing infants' looking behaviour in response to the changing beliefs of an observed agent, these studies suggested that infants may be able to represent and track others' mental states (e.g. beliefs) and use these to predict others' behaviours. Based on these findings, it has been indicated that implicit ToM competencies may exist from early infancy, and that looking behaviour may provide access to this ability. Only later during development, once children overcome verbal and inhibitory demands, this ability has been proposed to become explicit (e.g. Sodian et al., 2020).

Furthermore, implicit ToM has been previously indicated to be automatic and fast, although inflexible, while explicit ToM has been described as slower and effortful, but more flexible and dependent on language (e.g. Sodian et al., 2020). In line with this proposition, previous studies have indicated different performance in tasks involving implicit and explicit ToM understanding in adult individuals with ASD. For example, in Schuwerk et al. (2015), individuals with ASD showed typical behaviour in an explicit ToM task but scored significantly lower in an implicit ToM task. Similarly, Senju et al. (2009) identified a deficit specific to implicit ToM in individuals with Asperger syndrome using eye-tracking. Overall, it is debated whether implicit and explicit ToM may represent two separate cognitive processes underlying ToM or whether they fall along a continuum (Sodian et al., 2020; van Overwalle & Vandekerckhove, 2013). Please see Part 1 of this thesis for more details on this topic.

Although I was not able to investigate this relationship by contrasting performances between the limb difference population and controls as originally planned, in the present study I gathered data on the relationship between explicit and implicit ToM in 14 control individuals, allowing for a pilot online study and showing scope for obtaining some preliminary data.

Third objective: feasibility of online study and scope for laboratory study

To my knowledge, the present study is the first to implement a false-belief task online. Adapting an experimental task to online setting comes with challenges (see also infant study in Part 1, chapter 3 of this thesis), mainly resulting from a less controlled experimental setting due to the experimenter not being present during the testing session. For this reason, I believed it would be informative to determine the feasibility of implementing this task online and the replicability of the results obtained in the original study by Xie et al. (2018). Furthermore, this study used webcam eye-tracking, which has only recently become a tool of choice for online psychology studies

(Yang & Krajbich, 2020). Considering the novelty of webcam eye-tracking as a tool for psychology studies, some limitations remain for accurate recording and coding of participants' eye movements. For example, webcam eye-tracking has been previously reported to result in less accurate data and higher variance than that implemented in a laboratory setting (Semmelmann & Weigelt, 2018). Furthermore, it requires extensive calibration and validation procedures, and presents inconsistent temporal resolution (Semmelmann & Weigelt, 2018). In addition, webcam eye-tracking usually relies on an estimation of face and eye location averaged across timepoints to ensure accuracy, e.g. Papoutsaki et al. (2016); thus this methodology may hinder the accuracy of eye-tracking data if participants move, even slightly, throughout the experiment. Indeed, it is challenging for participants to keep still for the duration of a study, e.g. without a chinrest (as one would do in a laboratory setting), given the individual home setting. Notwithstanding these limitations, webcam eye-tracking has been previously shown to have great potential for online behavioural studies (e.g. Semmelmann & Weigelt, 2018; Yang & Krajbich, 2020), suggesting that its use may be informative for follow up investigations in laboratory experimental setting.

Research questions

To summarise, in this pilot study, I adapted Xie et al. (2018)'s task to an online setting and assessed its feasibility and the replicability of their findings with the limb difference population and control participants. First, I asked whether I could replicate, in my online study, findings from the original study conducted in a laboratory setting. Second, I asked whether their results could be extended by providing an analysis of participants' responses in a follow-up questionnaire on the mental rotation / perspective taking strategy they used to complete the study. Third, I asked whether their results could be extended by providing an analysis of participants' implicit false-belief understanding, as assessed through eye-tracking data, and by comparing it with their explicit false-belief understanding, as assessed through keyboard responses. Finally, I reported data from the limb difference population.

Methods

Participants

A total of 4 adults with limb difference and 14 adult controls, recruited from the Research for Limb Difference database (University of Essex), as well as internally at University of Essex or through social media, completed the study. Unfortunately, performing online recruitment of such a niche population (limb difference population) proved challenging and, given the online nature of this study, I was unable to control participants' engagement and attention throughout the task. Indeed, considering that this study was quite long (about 1 hour) and repetitive (20 trials per condition), participants might have found too demanding to complete in a home setting. In fact, additional 33 individuals with limb difference made an attempt at this online study but did not complete the study; thus were not included in the analysis. Furthermore, additional 7 controls made an attempt but did not complete the study and were thus excluded. Upon completion of the study, participants were compensated with a £5 Amazon Gift Card. Ethical approval was granted by the University of Essex Ethics Sub-Committee (ETH2021-0065).

False Belief Task

The task used in this study was adapted from the false-belief task created by Xie et al. (2018) to allow for online testing using the Gorilla platform for online behavioural studies. In the online version of the task, participants were asked to assess an agent's belief (true or false) regarding an object location, as observed in a video showed from either a 0° or 180° visual angle (mental rotation assessment). In addition, participants were asked to determine the object location based on their own perspective (perspective-taking assessment). Each trial began with a fixation cross at the centre of the screen (500ms). Either a true- or false-belief animation video (presented at a 0° or 180° visual angle) was then played for 8.5s. A blank screen with a fixation cross followed, in which participants were asked to take either their or the agent's perspective to indicate belief on object location. Participants provided their response using their keyboard, by pressing the space and arrow keys, respectively, to indicate the yellow and green boxes. Following participants' responses, a new trial started. Please see Figure 13 below for a visualisation of the events in a trial representative of this task. Participants had a time-limit of 3 second to provide their response; after this time, a new trial started regardless of whether participants provided their response. The experiment consisted of 160 trials (20 trials per condition – counterbalanced and their order randomised), which were separated by short breaks and differed regarding 3 factors: *(1)* object location; *(2)* visual angle; and *(3)* perspective. All participants included in this study completed at least 10 trials per condition. Throughout the experiment, participants' looking behaviour was detected using the webcam eye-tracker functionality built in the Gorilla platform. Specifically, participants' eye-gaze locations were detected on the screen in real time using Webgazer.js (Papoutsaki et al., 2016). Eye-tracking started and ended automatically, upon the start and the end of each trial, respectively.

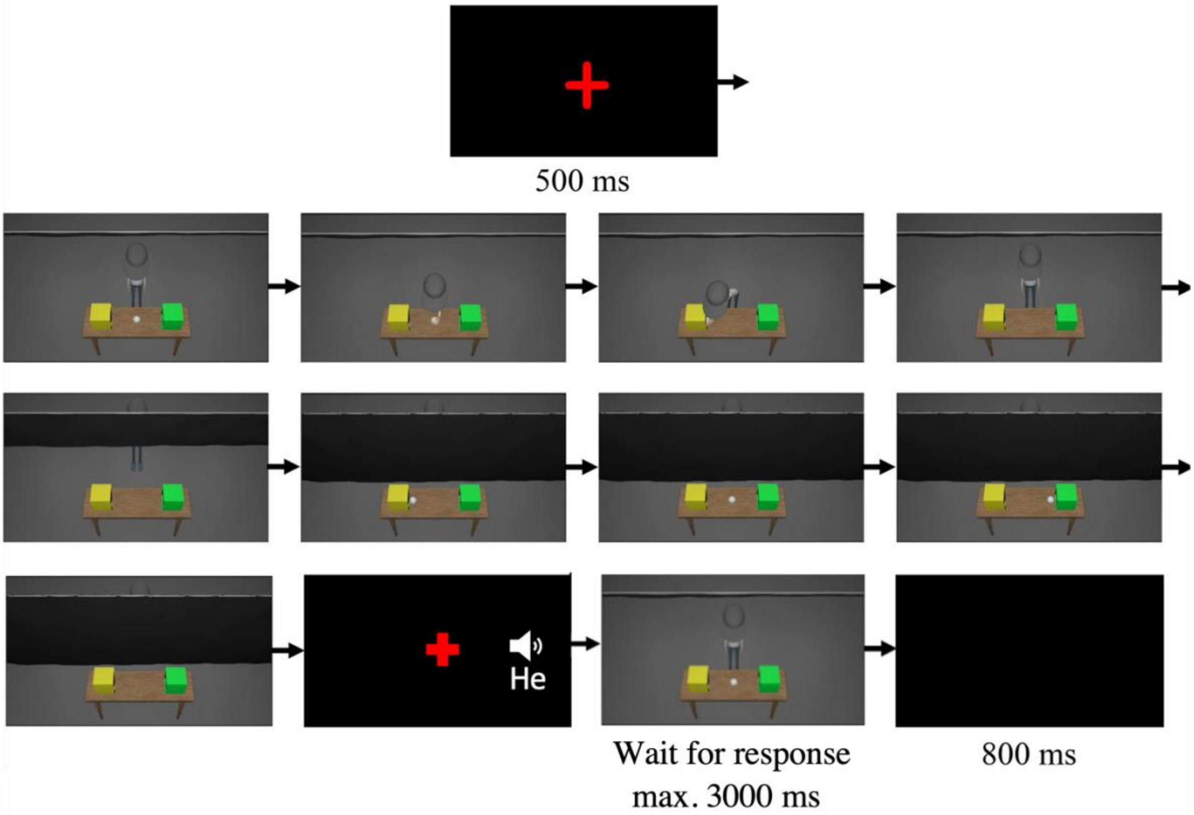


Figure 13. Sequence of events in a trial from my false-belief task (180° visual angle and agent’s perspective condition). Image adapted from Xie et al. (2018).

Self-report questionnaire

Upon completion of the false-belief task, participants were asked to complete a brief questionnaire which assessed the mental rotation / perspective taking strategy used to complete the study. Specifically, I asked participants “Which of the following better explains your thinking process while answering the questions related to "the agent" (HE) ?”, to which participants had the following four response options: (1) I imagined rotating the *AGENT*'s body position to my own body position to have the same perspective; (2) I imagined rotating *MY* own body position to the agent's body position to have the same perspective; (3) I did not imagine taking the agent' perspective; I used instead environmental landmarks to calculate his viewpoint (e.g. I relied on the position of the boxes); and (4) Other. The options introduced in my questionnaire were inspired by Surtees et al. (2013).

Procedure

Given the online nature of the study, individuals participated from their homes and there was no experimenter present. To start the study, participants accessed the Gorilla platform where the study was hosted. Next, participants read the study instructions, completed three practice trials (without feedback) and conducted eye-tracking calibration. Considering that this was not a laboratory study, the length of the study was reduced in an attempt to maintain participants' engagement with the task and increase data accuracy. Specifically, only the 0° and 180° visual angles were chosen in this task (while the following visual angles were discarded: 45°, 90°, 135°, 225°, 270°, and 315°); thus reducing the number of total trials from 640 to 160. This decision was taken following previous results and guidelines provided by Xie et al. (2018) who indicated these visual angles to be the most representative of their findings. Furthermore, participants completed the study in one session, as opposed to two sessions as in the original study, to reduce participants' drop-out. Upon completion of the false-belief task, participants filled the follow-up self-reported questionnaire on mental rotation / perspective taking strategies. Written consent was obtained from all included participants prior to the start of the study via email.

Data Processing

Both explicit (i.e. keyboard responses) and implicit (i.e. webcam eye-tracking) data from each trial, condition and participant were retrieved from the Gorilla platform.

With regards to the explicit data (i.e. keyboard responses), incorrect responses and response times shorter than 100 ms or longer than 2000 ms were discarded (Xie et al., 2018). Participants who did not complete at least 10 trials per condition were excluded. Thereafter, average response accuracies per condition per participant were calculated. Control participants who reported a total accuracy < 80% were excluded (Xie et al., 2018); the same was not valid for the limb difference population, for which there may exist potential deviations from the accuracy scores seen in the general population. Finally, average response accuracies per condition across participants were calculated. Prior to submitting the data for statistical analysis, based on the original study's methodology (Xie et al., 2018), I arcsine-transformed the average response accuracy data to make them more suitable for the ANOVA. Specifically, I used the function $Y' = 2 \times \text{arcsine}(Y^{1/2})$ in which Y' and Y were the transformed and original values, respectively. Before transformation, the extreme value 1 was replaced by $(1 - [1/4n])$ while n was the number of trials based on which accuracy was estimated for each combined condition, i.e. $n = 160/(2 \times 2 \times 2) = 20$.

With regards to the implicit data (i.e. eye-tracking), data were processed in two different ways. First, the percentage occupancies of the four quadrants of the screen were retrieved from the Gorilla platform. Each quadrant was associated with an area of interest in the screen, which related to the ball location and visual angle (see Figure 14). This data was then processed and the average percentage occupancy of each quadrant of the screen per condition per participant was calculated. Next, average percentage occupancies of each quadrant per condition across participants were also calculated. I was this way able to determine where in the screen participants looked (i.e. which box) and for how long. This information is indicative of their false belief understanding, as well as their perspective taking and mental rotation abilities (Xie et al., 2018). Second, a code was implemented on the RStudio software Version 1.4.1103 to analyse the eye-tracking data using the Saccades package from GitHub (https://github.com/tmalsburg/saccades). Eye-tracking data are provided by the Gorilla platform in prediction rows, each corresponding to a single eye-tracking sample for a given trial for a given participant. Therefore, prior to running the code, such predictions were filtered using a value >0.5 for the Support Vector Machine (SVM) classifier score for the face model fit. All data which did not meet this criterion were excluded, as this would suggest that the model's confidence in finding a face was low, and thus that the predicted eye movements would likely be inaccurate. After running the code, eye movement data were retrieved and plotted per participant, for each given condition, and on a trial-by-trial basis. Finally, a code was implemented on the MATLAB and Statistics Toolbox Release 2020b to visualise the eye movement data for a given condition with respect to the scene observed on the screen by the participant during the experiment (see Figure 15).

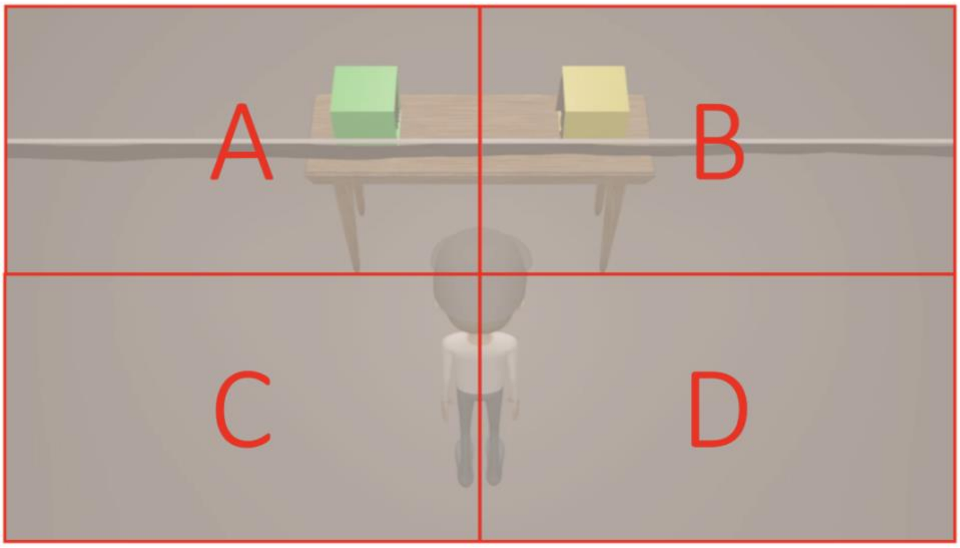


Figure 14. Areas of interest (AOIs) dividing the screen in four quadrants associated with ball location and visual angle in a trial from my false-belief task. In this case, AOI "A" is associated with green box, 0° visual angle; AOI "B" with yellow box, 0° visual angle; AOI "C" with quadrant opposite to green box, 0° visual angle; AOI "D" with quadrant opposite to yellow box, 0° visual angle. Average percentage occupancies of each quadrant of the screen (AOIs) per condition per participant and per condition across participants were calculated.

Data Analysis

Given the small population size, both explicit and implicit data from the limb difference population were only reported, and not analysed. In contrast, data from controls were submitted for statistical analysis. To determine whether an effect of any of the independent variables (object location, visual angle, and perspective) on participant's response accuracy (explicit measure) existed, a 2 x 2 x 2 repeated measures ANOVA was conducted. Another 2 x 2 x 2 repeated measures ANOVA was performed to determine the presence of an effect of any of the independent variables (object location, visual angle, and perspective) on participant’s looking behaviour (implicit measure).

Results

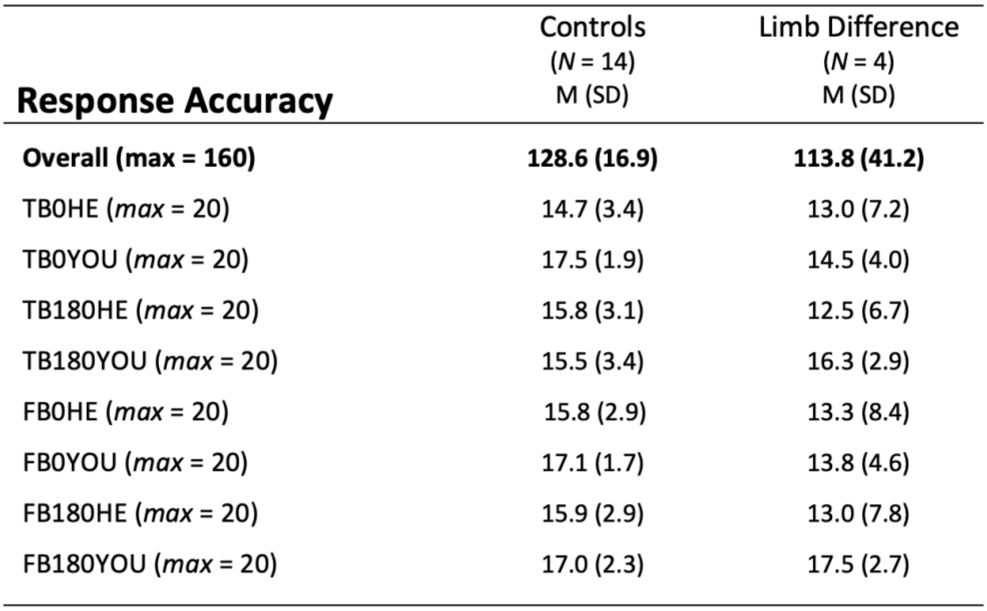
Feasibility of online study

Out of 79 participants (individuals with limb difference: N = 37, controls: N = 42) making an attempt at this online study, only 18 (~23%) (individual with limb difference: N = 4, ~11%; controls: N = 14, ~33%) successfully completed the study. Additional 9 participants (individuals with limb difference: N = 2; controls: N = 7) were tested but excluded as they did not meet the inclusion criteria of completing at least 10 trials per condition (individuals with limb difference: N = 2, ~5%; controls: N = 4, ~10%) or of achieving an overall response accuracy > 80% (controls: N = 3, ~7%). percentage of incorrect responses in the limb difference and control populations was 13.1 and 5.2%, respectively. Furthermore, ~15.8 and ~19.5% of trials were excluded in the limb difference and control populations, respectively, due to response times shorter than 100 ms and longer than 2000 ms (Xie et al., 2018).

Response Accuracy

Table 12 present a summary of the overall and average (per condition) response accuracies in the false-belief task across individuals with limb difference and controls.

Table 12. Total response accuracy and average response accuracy per condition across participants with limb difference and controls.



Response Accuracy	Controls (*N* = 14) M (SD)	Limb Difference (*N* = 4) M (SD)
Overall (max = 160)	**128.6 (16.9)**	**113.8 (41.2)**
TB0HE (*max* = 20)	14.7 (3.4)	13.0 (7.2)
TB0YOU (*max* = 20)	17.5 (1.9)	14.5 (4.0)
TB180HE (*max* = 20)	15.8 (3.1)	12.5 (6.7)
TB180YOU (*max* = 20)	15.5 (3.4)	16.3 (2.9)
FB0HE (*max* = 20)	15.8 (2.9)	13.3 (8.4)
FB0YOU (*max* = 20)	17.1 (1.7)	13.8 (4.6)
FB180HE (*max* = 20)	15.9 (2.9)	13.0 (7.8)
FB180YOU (*max* = 20)	17.0 (2.3)	17.5 (2.7)

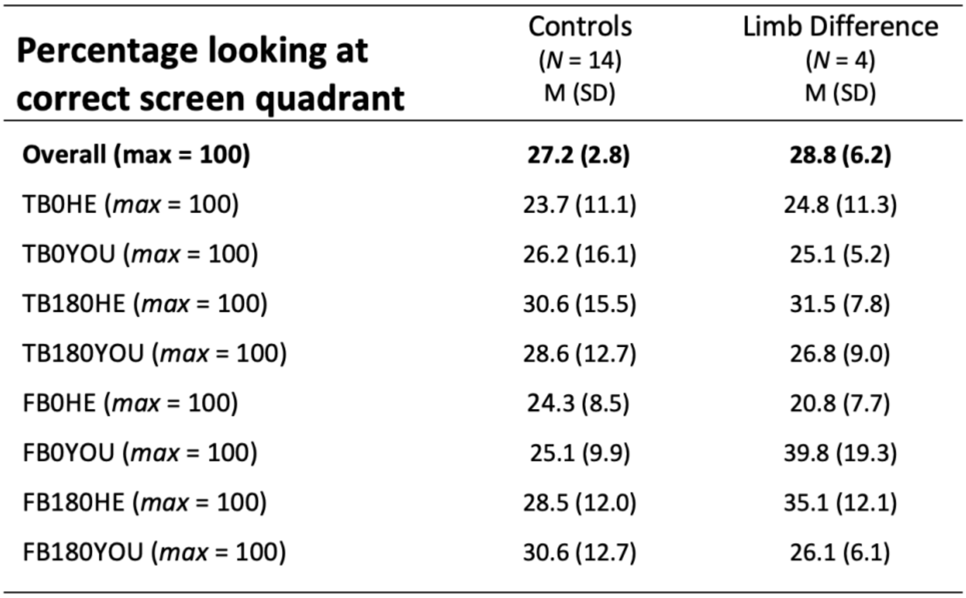
TB = True belief; FB = False belief; 0 = 0 degrees visual angle, 180 = 180 degrees visual angle, HE = agent's perspective, YOU = participant's perspective. E.g. TB0HE = true belief, 0 degrees visual angle, agent's perspective.

A 2 x 2 x 2 repeated measures ANOVA conducted on data from the control population reported a main effect of ball location, F(1, 13) = 5.948, *p* = .030, ηp^2 = .314, on participants' response accuracy, which was driven by the true belief condition (M = 2.723, SD = .050) vs false belief condition (M = 2.658, SD = .053). In contrast, visual angle, F(1, 13) = .282, *p* = .604, ηp^2 = .021, and perspective, F(1, 13) = .857, *p* = .372, ηp^2 = .062, did not impact participants' response accuracies. Furthermore, only the interaction between ball location and perspective reached significance, F(1, 13) = 4.686, *p* = .050, ηp^2 = .265. Significance was not observed for the interaction between ball location and visual angle, F(1, 13) = .521, *p* = .483, ηp^2 = .039, or visual angle and perspective, F(1, 13) = 1.118, *p* = .310, ηp^2 = .079. Similarly, a three-way interaction was not found, F(1, 13) = .115, *p* = .739, ηp^2 = .009.

Looking behaviour

Table 13 includes a summary of the overall and average percentage looking (per condition) at the correct screen quadrant in the false-belief task across individuals with limb difference and controls.

Table 13. Overall percentage looking occupancy and average percentage looking occupancy per condition at the correct screen quadrant across participants with limb difference and controls.



Percentage looking at correct screen quadrant	Controls (*N* = 14) M (SD)	Limb Difference (*N* = 4) M (SD)
Overall (max = 100)	**27.2 (2.8)**	**28.8 (6.2)**
TB0HE (*max* = 100)	23.7 (11.1)	24.8 (11.3)
TB0YOU (*max* = 100)	26.2 (16.1)	25.1 (5.2)
TB180HE (*max* = 100)	30.6 (15.5)	31.5 (7.8)
TB180YOU (*max* = 100)	28.6 (12.7)	26.8 (9.0)
FB0HE (*max* = 100)	24.3 (8.5)	20.8 (7.7)
FB0YOU (*max* = 100)	25.1 (9.9)	39.8 (19.3)
FB180HE (*max* = 100)	28.5 (12.0)	35.1 (12.1)
FB180YOU (*max* = 100)	30.6 (12.7)	26.1 (6.1)

TB = True belief; FB = False belief; 0 = 0 degrees visual angle, 180 = 180 degrees visual angle, HE = agent’s perspective, YOU = participant’s perspective. E.g. TB0HE = true belief, 0 degrees visual angle, agent’s perspective.

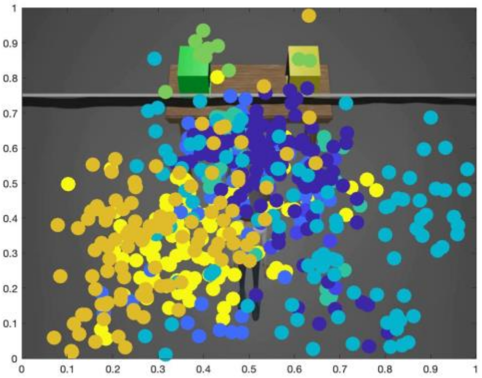
A 2 x 2 x 2 repeated measures ANOVA conducted on data from the control population reported an absence of main effects on participants' percentage looking at the correct quadrant in the screen and a lack of interactions between conditions. Following are reported the main effects of ball location, F(1, 13) = .020, *p* = .890, ηp^2 = .002; visual angle, F(1, 13) = .354, *p* = .562, ηp^2 = .026, and perspective, F(1, 13) = .104, *p* = .752, ηp^2 = .008.

Example visualisation and analysis of eye gaze location

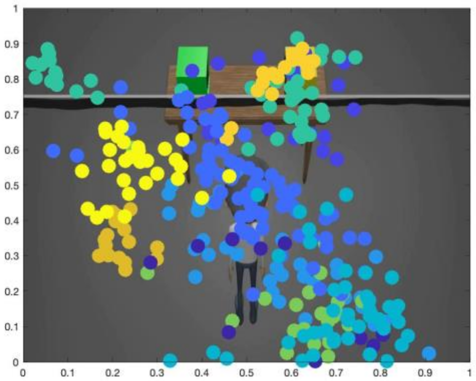
Figure 15 below shows an example visualisation of the looking behaviour across trials of a representative participant from the control group for each given condition. Specifically, it shows the predicted eye positions for a control participant in a scenario in which the ball was originally placed in the green box. While the ball remained in the green box in some trials (images a, c, e, g), it rolled in the yellow box in other trials without the agent seeing such change in location (images b, d, f, h), thus leading to a false belief in the agent. Furthermore, participants were asked to take their own perspective (images c, d, g, h) or that of the agent (images a, b, e, f). In addition, scenes were at times presented at a 0° visual angle (images a-d) or a 180° (images e-g) visual angle with respect to the participant.

While the data presented in the visualisation below are not representative of the whole control population included in this study, I provide an analysis of the eye position seen in this participant to show the potential of this technique and pave the way to future studies. Briefly, the estimated eye positions of this participant seem to suggest that they correctly looked more times at the green box, as opposed to the yellow box, in the true belief condition (where the ball was located), regardless of perspective and visual angle disparity (images a, c, e, g). Similarly, the participant correctly looked more times at the yellow box (as opposed to the green box) in the false belief condition (where the ball was located following the change in location of which the agent was unaware) when asked to take self-perspective, regardless of the visual angle disparity (images d, h). In contrast, while the participant correctly looked more times at the green box when asked to take the agent's perspective in the false-belief condition with a 180° visual angle disparity (image f), it seems that they incorrectly looked more times at the yellow box in the same condition but with a 0° visual angle disparity (image b). Therefore, we may conclude from these results that this participant may have additional difficulties in engaging in ToM, thus understanding another's false belief in this case (i.e. world from the agent's perspective), only when there is an absence of visual angle disparity between agent and participant. In turn, this may indicate that this participant struggled the most with taking the agent's mental perspective and understanding the agent's false beliefs when having the same visual perspective but having differing beliefs, possibly suggesting a lack of mental self-perspective suppression.

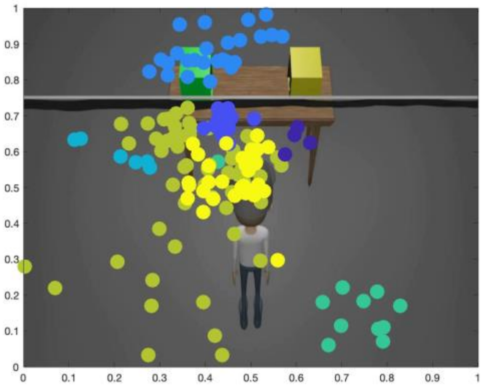
(a) TB0HE



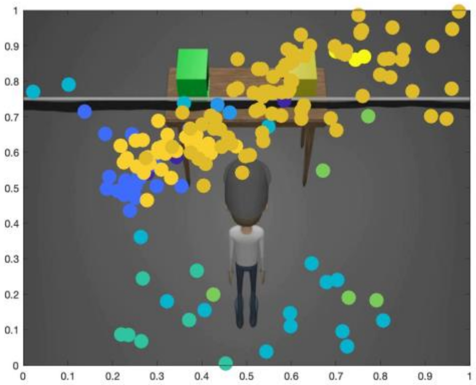
(b) FB0HE



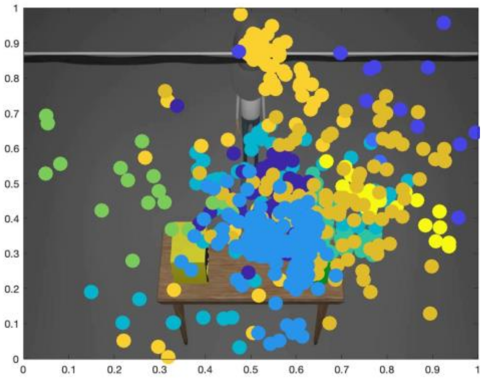
(c) TB0YOU



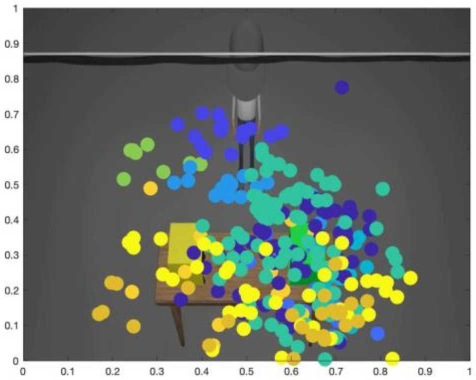
(d) FB0YOU



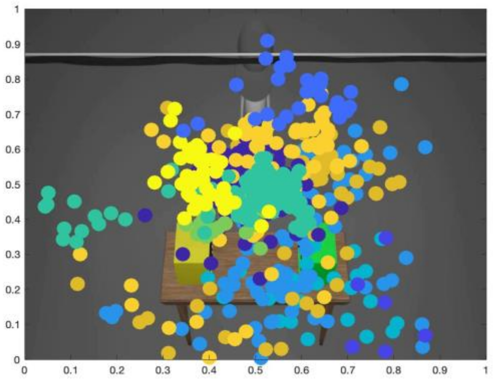
(e) TB180HE



(f) FB180HE



(g) TB180YOU



(h) FB180YOU

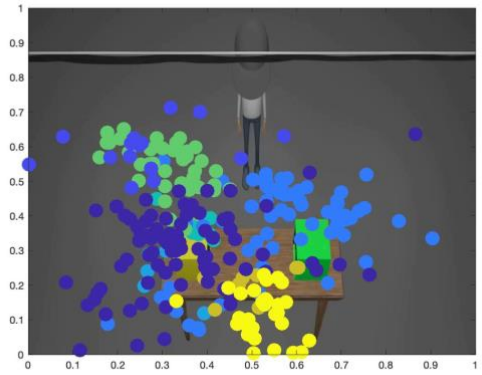


Figure 15. Visualisation of control participant's looking behaviour across trials for each given condition (green box order). TB = True belief; FB = False belief; 0 = 0 degrees visual angle, 180 = 180 degrees visual angle, HE = agent's perspective, YOU = participant's perspective. E.g. TB0HE = true belief, 0 degrees visual angle, agent's perspective.

Follow-up questionnaire

To determine which was the correct response during the false-belief task, 1 individual with limb difference reportedly relied on environmental cues, while 1 imagined rotating the agent's body position to their own body position to have the same perspective. The remaining 2 participants mentioned other strategies not relevant to this discussion, i.e. "got confused" and "I concentrated on the colours". With respect to the control group, out of 14 participants who completed the follow-up questionnaire, 8 relied on environmental cues, 2 imagined rotating the agent's body position to their own body position to have the same perspective, and 2 imagined rotating their own body position to the agent's body position to have the same perspective. The remaining 2 participants mentioned other strategies not relevant to this discussion, i.e. "whatever box he put it in is the one he would choose?" and "I memorized the colours of the boxes".

Discussion

In this pilot study, I investigated the replicability of findings from Xie et al., (2018) with regards to a role of perspective taking and mental rotation for false belief understanding in an online setting. Furthermore, I introduced a self-reported measure of mental rotation / perspective taking strategy used during the here implemented false-belief task to extend the results from the original study, and assessed implicit vs explicit measures of ToM.

The results from this preliminary study only partially supported Xie et al. (2018)'s original findings in the general population, only identifying an effect of belief type on participants' response accuracy (explicit measure). In contrast, I did not find an effect neither for mental rotation nor perspective taking from the results obtained both in the false-belief task and the self-report questionnaire. Finally, the newly introduced analysis of participants' looking behaviour (implicit measure) did not support an effect for belief type on percentage looking at the correct quadrant of the screen. Therefore, these results suggest that belief type may differently impact explicit and implicit ToM. However, given the small samples size and other limitations outlined below, future studies are warranted to validate my results. Finally, the feasibility of conducting this study online is briefly discussed.

Response accuracy

Xie et al. (2018) indicated that false belief performance (response accuracy) was affected by increased orientation disparity between the participants and the agent, suggesting involvement of embodied transformation (i.e. mental rotation). In contrast, my results only reported an effect of ball location, thus belief type, on participants' response accuracy. Specifically, participants scored significantly higher in the true vs false belief conditions. My results therefore suggest that belief type (true vs false belief) may have an impact on belief understanding, as assessed through explicit keyboard responses. However, my results do not support a relationship between mental rotation and/or perspective taking and ToM, which was indicated in the original study. Indeed, although I did find an interaction between ball location (thus belief type) and perspective, this only reached significance, thus highlighting this interaction, if existing, to be weak. Nonetheless, I cannot exclude that the small sample size included in this study and the reduced number of trials (M = 16.0, SD = 1.8 trials per condition) compared to the original study (20 per condition), as well as the online implementation of this study may have hindered significant findings. Both sample size and number of trials (and their interaction) have been previously indicated to highly influence statistical power in a study (Baker et al., 2021). See below for a more detailed discussion on the impact of conducting this study online on data. Future studies should replicate the study by closely matching sample size and number of trials, as well as experimental setting to answer this question.

Self-reported questionnaire

In this study, I aimed to extend the results from the original study (Xie et al., 2018) by explicitly asking participants about the mental rotation / perspective taking strategies they used to complete the task through a follow-up questionnaire upon finishing the study. The findings from this newly introduced follow-up questionnaire also do not support participants' reliance on mental rotation / perspective taking strategies to succeed in this task. Indeed, most of the participants reported that they relied only on environmental cues (or other techniques) to make their choices instead of mentally rotating the self to take the other's perspective. Considering that these measures were obtained from a self-report questionnaire, I cannot exclude that participants may have not accurately reported their strategy, or that participants' idea of their strategy may not reflect their actual strategy to complete the task. However, this is a limitation inherent to all self-report questionnaires (Demetriou et al., 2015). Nonetheless, these results seem to support the same view previously outlined by behavioural results on response accuracy, which also indicated an absence of a mental rotation / perspective taking strategy.

Eye-tracking

Finally, the here introduced analysis of participants' looking behaviour failed to reflect the above indicated effect of ball location (thus belief type) on participants' performance, as participants looked equally at the correct quadrant of the screen regardless of belief type. This therefore may indicate a different impact of belief type on explicit and implicit ToM. Indeed, while belief type was shown to affect participants' explicit keyboard responses, it did not seem to influence their implicit looking behaviour. Nonetheless, some differences in gaze locations between conditions were observed in some participants (e.g. the looking behaviour of the participant above described). Unfortunately, given the small sample size included in this pilot study, I am unable at this stage to draw conclusions as to whether these results may indicate that such explicit and implicit measures may support different cognitive processes. Therefore, a bigger sample size is warranted to validate this data and determine whether statistical significance can be reached with a more representative sample. Furthermore, it would be interesting to conduct this experiment in a laboratory setting, where participants' eye-tracking data can be collected more accurately and in a more controlled environment. Webcam eye-tracking was an interesting tool for us to use in this study as a complementary measure of response accuracy for participants' false-belief understanding. However, I am unable at this stage to determine whether the results seen in my study may be driven by the lack of accuracy of this tool online. Indeed, visualisation of the webcam eye-tracking data from participants suggested looking behaviours not to be extremely accurate, rendering the recognition of looking patterns challenging (see Figure 15 for an example visualisation). Therefore, future studies addressing these issues are needed, both to validate my findings and to provide further insights into the feasibility of using webcam eye-tracking.

Feasibility of online study

Finally, this pilot study suggests that implementing this false-belief task online is possible. However, my results also indicate that this experimental setting may result in less accurate data (both explicit and implicit) and higher participant drop-out or data loss. Specifically, with regards to explicit data, a higher percentage of inaccurate responses was observed in this online study compared to the original laboratory study (~5.2 vs ~3.6%, respectively). Similarly, a higher number of trials was excluded in the online study compared to the original study due to responses being provided either too fast or too slow (~19.5 vs ~11.1%, respectively). These results may suggest participants' lower engagement and attention throughout the task, or a lack of motivation when conducting a long and repetitive study in the comfort of one's own home. Furthermore, the still limited accuracy of webcam eye-tracking may have influenced the results of this online study and driven the absence of an effect of belief type on implicit ToM. Future studies should be conducted to determine whether this finding maintains in a laboratory setting and the feasibility of webcam eye-tracking for this task. Finally, a drop-out rate of 77% was observed in my study, which therefore calls for a limitation in the online implementation of this study.

Conclusions

To conclude, it is worth to briefly mention that, although a comparison between controls and individuals with limb difference was unfortunately not possible in this study, the approach here outlined represents a viable way to investigate false belief understanding, as well as the role of mental rotation and perspective taking for ToM, from an innovative perspective. I hope to complete this study in the future, possibly also implementing it in a laboratory setting.

3.3 Prevalent characteristics among individuals with limb difference: a population-based report

Introduction

In this subchapter, I report the prevalence of type, level, and side of limb difference among individuals with limb difference included in my database, which I also analysed by gender. In addition, I describe the prevalence of phantom limb experiences and prosthesis use among my population sample, both by type of limb difference and gender. Similarly, I report the prevalence of physical conditions and mental disorders among individuals included in my database, also by type of limb difference and gender. Finally, I introduce a new measure (i.e. language used to describe limb difference) for investigating the psychology of living with limb difference among the congenital and acquired subgroups. The aim of this report was to provide a unified summary of the prevalence of all such characteristics among my limb difference sample population, which are instead generally provided separately in the literature (see below). In addition, I aimed in this report to introduce a new measure for accessing varying experiences related to having a limb difference between the congenital and acquired subgroups, in an attempt to better understand the psychology of living with limb difference. Overall, I hope this report can inspire future studies in this direction, as well as inform individuals who require information on limb difference, both for academic and clinical purposes, or other. I will now briefly delineate four of the main motivations to conduct this study.

First, to my knowledge, research assessing the prevalence of all the above in the same population does not exist in the literature. To provide a few examples, Ziegler-Graham et al. (2008) provided data on the prevalence of acquired limb difference in the United States by aetiology of the limb difference, age, sex, and race. Furthermore, they estimated future prevalence based on specific incidence rates for amputation combined with mortality data. Mai et al. (2019) indicated the prevalence of congenital limb difference between 2010 and 2014 on another cohort in the United States by race and level of limb difference, although within a discussion on general major birth defects. Fraser (1998) instead reported laterality, gender and age differences in the frequency of congenital limb difference in the United Kingdom, however only focusing on upper (and not lower) congenital limb differences. Kyberd et al. (1997), in an attempt to inform the design of upper limb prostheses, investigated prevalence of congenital and acquired limb difference among attendees of the Oxford Limb Fitting Centre in the UK. They did so by age, gender, prosthesis use, cause and level of limb difference. However, the authors focused on individuals with upper limb differences only. Ephraim et al. (2003) reviewed articles in the literature to estimate the incidence of congenital and acquired limb difference worldwide. Data on the acquired limb difference population were also presented by age, race, as well as level and cause of amputation. However, those from the congenital limb difference population were only analysed by cause of limb difference and time of detection. Saadah and Melzack (1994) reported phantom limb experiences in only the congenital limb difference population. In contrast, Melzack et al. (1997) described phantom limbs in individuals with congenital and acquired limb differences, although without providing prevalence of the above-described characteristics. Finally, some studies including Datta et al. (2004) and Mckechnie and John (2014), investigated physical conditions and/or mental disorders among the limb difference population. However, on the one hand, Mckechnie and John (2014) only focused on individuals with acquired limb difference. On the other hand, Datta et al. (2004) described anxiety and depression among individuals with congenital and acquired limb difference, also in relation to prosthesis use and phantom limb experiences. Overall, it seems clear that a complete overview of the prevalence of characteristics associated with limb difference in the same population is currently lacking in the literature.

Second, most of the data present in the literature is provided by cohorts from different sides of the world, rendering their comparison and interpretation challenging. Indeed, a previous study identified different incidence rates of acquired limb difference between nations (Ephraim et al., 2003). Vuillermin et al. (2021) reported notable differences between their registry of individuals with congenital limb difference and those from the United States Midwest and Sweden. Furthermore, a lower number of reports on cases of congenital limb difference was found from developing vs developed countries (Ephraim et al., 2003). Overall, providing a unified summary of such data from the same population may allow a more accurate comparison of data and increase clarity on their relative prevalence in a specific country.

Third, it is currently challenging to find studies in the literature which directly describe characteristics of individuals with congenital vs acquired limb difference. Indeed, most of the studies provide data for either the congenital (e.g. Fraser, 1998; Mai et al., 2019; Vuillermin et al., 2021) or acquired (e.g. Heikkinen et al., 2007; Ziegler-Graham et al., 2008) limb difference groups, separately. Only a few studies in the literature (e.g. Ephraim et al., 2003; Kyberd et al., 1997; Melzack et al., 1997; Wilkins et al., 1998) included both the congenital and acquired limb difference populations in their investigations. However, Kyberd et al. (1997) only focused on the upper limb difference subgroup, not providing data on individuals with lower limb difference. Ephraim et al. (2003) instead investigated differences in prevalence of congenital and acquired limb difference worldwide. However, the authors mainly focused their analysis on the acquired limb difference subgroup (see above). While Melzack et al. (1997) only focused on phantom limb experiences, Wilkins et al. (1998) only investigated children and adolescents with limb difference instead. Overall, given the lack of a comprehensive description of the prevalence of the above characteristics in congenital vs acquired limb difference in the literature, I deemed it necessary to provide such an analysis in this report. Therefore, this report distinguishes itself from previous research in the literature as it aims at directly describing such prevalent characteristics among adults with congenital vs acquired limb difference.

Fourth and last, previous studies have conducted surveys in both individuals with limb difference and their families to investigate psychosocial dynamics among individuals with limb difference (see Part 2, chapter 1 for a detailed description of these studies). Briefly, Murray et al. (2007) identified the strengths, challenges and relational processes in families with children with limb difference. Bae et al. (2018) compared peer relationships and emotional states between children with congenital upper limb differences and the general population. Furthermore, Rybarczyk et al. (1995) reported body image, perceived social stigma and psychosocial adjustments in individuals who experience lower-limb amputation. However, literature suggests that we still know relatively little about the psychology of living with limb difference, especially on the potential different experiences resulting from having congenital vs acquired limb difference. In an attempt to shed some further light onto the matter, I included an innovative measure in this report, i.e. language used to describe limb difference. Specifically, while participants were simply asked to describe their limb difference, changes in language used may reflect varying representation and acceptance of limb difference by individuals, as well as self-awareness and openness to discuss about their limb difference. By being an implicit question (as it did not directly ask participants how they experience their limb difference), this measure may provide some new insights into potential varying experiences related to having and living with congenital vs acquired limb difference. Overall, I introduced this measure in an attempt to show scope for further research and inspire future studies in this same direction.

Methods

Participants

A total of 259 adults with limb difference, whose characteristics were investigated and described below, populated the “Research for Limb Difference” database (University of Essex). This database was newly created specifically for this research project given that a database including people with limb difference was not available at the time this report and the above studies were conceived. Participants were recruited through social media and through the support of partner associations supporting individuals with limb differences, including REACH Charity, IAMPOSSIBLE Foundation, Steelbones and LimbPower. Ethical approval was granted by the University of Essex Ethics Sub-Committee (ETH2021-0065).

Qualtrics Survey

Upon registration to the database, participants were asked to complete an online questionnaire, hosted on the Qualtrics platform. This questionnaire collected (a) personal information, such as participants’ gender and location; (b) details on their limb differences, including type and level of limb difference, age of occurrence of limb difference, phantom limb experiences, and use of prostheses; and (c) information on participants’ physical conditions and/or mental disorders.

Procedure

Upon registering their interest in signing up to the Research for Limb Difference database, participants were sent an email with a link to the online survey hosted on the Qualtrics platform. Participants completed the online questionnaire from their homes. Written consent was obtained from all included participants prior to the start of the questionnaire through the Qualtrics platform.

Data analysis

Data was retrieved from the Qualtrics platform, and a descriptive analysis was conducted.

Results

Limb difference by type, level, side, and gender

First, the prevalence of type, level, and side of limb difference was assessed. This analysis was conducted by gender as well.

A. Prevalence of type of limb difference

Out of the 259 people with limb difference who signed up to the database, 44.4% had congenital limb difference. Overall, these results indicate a higher prevalence of acquired vs congenital limb difference among individuals included in the database.

When conducting the same analysis by gender, I observed a slightly higher prevalence of males (56.5%) vs females with congenital limb difference in the database. In contrast, there were no differences in prevalence of acquired limb difference between males and females, as both genders were equally represented in this group (50%). See Table 14 below for a summary of prevalence of limb difference in the database by type of limb difference and gender.

Table 14. Prevalence of limb difference among individuals in the database by type of limb difference and gender.

	Participants (N)	Prevalence (%)
Acquired Limb Difference	**144**	**55.6**
Female	72	50
Male	72	50
Congenital Limb Difference	**115**	**44.4**
Female	50	43.5
Male	65	56.5
Total	**259**	**100**

B. Overall prevalence of level and side of limb difference

Table 15 presents a summary of the overall prevalence of level of limb difference, as well as the prevalence by type and gender. Results indicate a similar prevalence of lower (48.3%) and upper (45.6%) limb difference among the limb difference population included in the database, whereas a lower prevalence of mixed upper and lower (6.2%) limb difference was found. Furthermore, a similar prevalence of side of limb difference (left: 46%; right: 47%) was observed overall, whereas a lower prevalence of mixed left and right (7%) limb difference was seen.

When conducting the same analysis by gender, I observed that 59% of females within the limb difference population had lower limb difference, while 34.4% had upper limb difference and 6.6% had mixed upper and lower limb difference. In contrast, about 38.7% of males in this population had lower limb difference, while 55.5% presented upper limb difference and 5.8% had mixed upper and lower limb difference. A similar prevalence of side of limb difference was instead observed among male and female individuals included in the database.

Table 15. Prevalence of level and side of limb difference among participants included in the database by type and gender.

	Participants (N)	Prevalence (%)	F : M	Prevalence (%)
Acquired Limb Difference	**144**	**55.6**	**72 : 72**	**50 : 50**
L Lower	43	29.9	25 : 18	58.1 : 41.9
L Upper	21	14.6	9 : 12	42.9 : 57.1
R Lower	37	25.7	23 : 14	62.2 : 37.8
R Upper	32	22.2	10 : 22	31.3 : 68.7
R Lower, L Lower	4	2.8	2 : 2	50 : 50
R Upper, R Lower	2	1.3	0 : 2	0 : 100
R Upper, L Upper, R Lower, L Lower	5	3.5	3 : 2	60 : 40
Congenital Limb Difference	**115**	**44.4**	**50 : 65**	**43.5 : 56.5**
L Lower	19	16.5	3 : 16	15.8 : 84.2
L Upper	36	31.3	13 : 23	36.1 : 63.9
R Lower	26	22.6	21 : 5	80.8 : 19.2
R Upper	26	22.6	8 : 18	30.8 : 69.2
R Upper, L Upper	3	2.5	2 : 1	66.7 : 33.3
R Upper, L Upper, R Lower	1	0.9	1 : 0	0 : 100
R Upper, L Upper, L Lower	1	0.9	1 : 0	60 : 40
R Upper, L Upper, R Lower, L Lower	1	0.9	1 : 0	50 : 50
L Upper, R Lower	1	0.9	0 : 1	0 : 100
L Upper, R Lower, L Lower	1	0.9	0 : 1	0 : 100
Total	**259**	**100**	**122 : 137**	**100**

F : M = female : male; L = left, R = right

C. Prevalence of level and side of limb difference in the congenital subgroup

With respect to the congenital subgroup, about 39.1% of individuals presented lower limb difference, while 56.5% had upper limb difference and 4.4% had mixed upper and lower limb difference. These results indicate a higher prevalence of upper limb difference in the congenital subgroup. Finally, a similar prevalence of limb difference occurring on either side of the body was seen among individuals with congenital limb difference (left: 47.8%, right: 45.2%, left and right: 7%).

When conducting the same analysis by gender, 48% of females presented lower limb difference, while 46% had upper limb difference and 6% had mixed upper and lower limb difference. In contrast, 32.3% of males had lower limb difference, while 64.6% presented upper limb difference, and 3.1% had mixed upper and lower limb difference. Finally, within the female participants of the congenital limb difference group, a higher prevalence of right vs left limb difference (58 vs 32%) was observed; the opposite was valid for the male participants of the congenital limb difference group (60% of left limb difference cases).

D. Prevalence of level and side of limb difference in the acquired subgroup

With respect to the acquired subgroup, 55.6% of individuals presented lower limb difference, while 36.8% had upper limb difference and 7.6% had mixed upper and lower limb difference. These results indicate a higher prevalence of lower limb difference in the acquired limb difference subgroup. Finally, a similar prevalence of limb difference occurring on either side of the body was seen among individuals with acquired limb difference (left: 44.5%, right: 47.9%, left and right: 7.6%). See Figure 16 below for a visualisation of the prevalence of level and side of limb difference in individuals with congenital vs acquired limb difference included in the database.

When conducting the same analysis by gender, 66.7% of females presented lower limb difference, while 26.4% had upper limb difference and 6.9% had mixed upper and lower limb difference. In contrast, 44.4% of males had lower limb difference, while 47.2% presented upper limb difference, and 8.5% had mixed upper and lower limb difference. Finally, a similar prevalence of side of limb difference was found within the female participants of the acquired subgroup (left: 51.4%, right: 45.8%). In contrast, a slightly higher prevalence of right-side limb difference was observed within the male participants of the acquired subgroup (left: 41.7%, right: 50%).

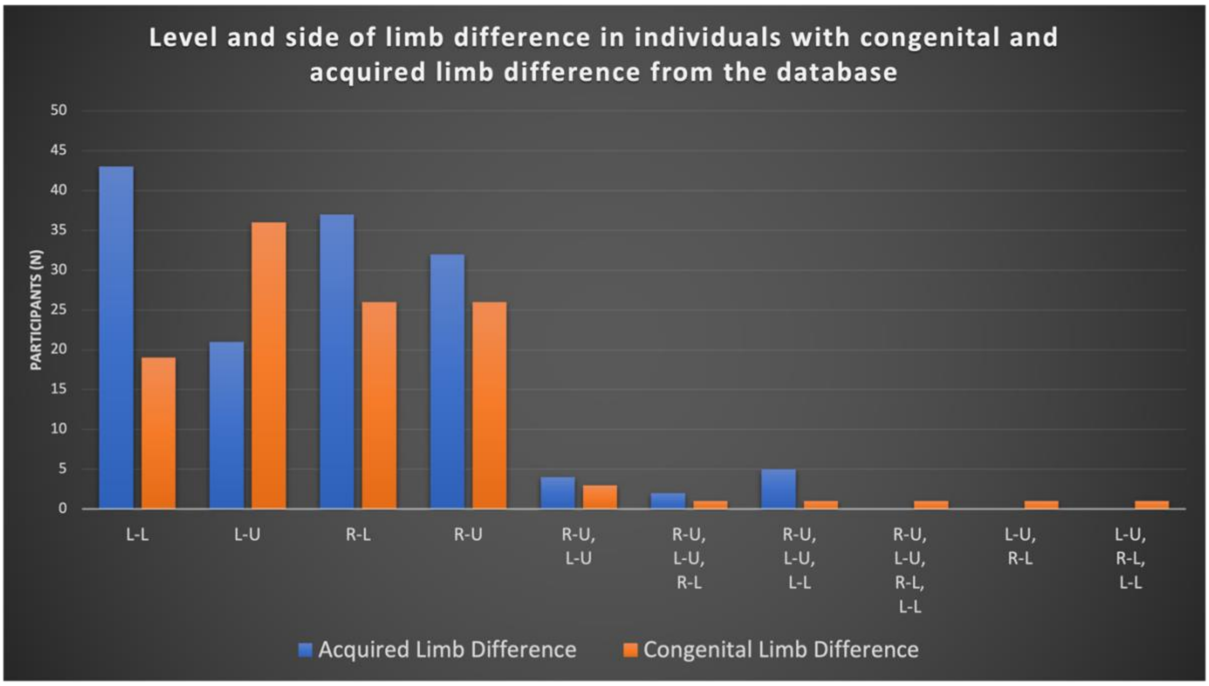


Figure 16. Visualisation of data on level and side of limb difference from participants with congenital and acquired limb difference in the database. L-L: left lower body; L-U: left upper body; R-L: right lower body; R-U: right upper body.

Phantom limbs

Next, the prevalence of phantom limb sensation and pain was assessed, as well as their prevalence by type of limb difference and gender. Please see Table 16 below for a summary of this data.

Table 16. Prevalence of phantom limb experiences in individuals from the database by type of limb difference and gender.

	Participants **N (F : M)**	**Prevalence** **%**
Acquired Limb Difference	**144 (72 : 72)**	**55.6 (50 : 50)**
No Phantom Limbs	6 (3 : 3)	4.2 (50 : 50)
Phantom Limbs Experience	138 (69 : 69)	95.8 (50 : 50)
Sensation	53 (26 : 27)	36.8 (36.1 : 37.5)
Pain	32 (10 : 22)	22.2 (13.9 : 30.6)
Sensation and Pain	53 (33 : 20)	36.8 (45.8 : 27.8)
Congenital Limb Difference	**115 (50 : 65)**	**44.4 (43.5 : 56.5)**
No Phantom Limbs	54 (40 : 14)	47 (74.1 : 25.9)
Phantom Limbs Experience	61 (10 : 51)	53 (16.4 : 83.6)
Sensation	38 (7 : 31)	33.0 (14.0 : 47.7)
Pain	22 (3 : 19)	19.1 (2.6 : 29.2)
Sensation and Pain	1 (0 : 1)	0.8 (N/A : 1.5)
Total	**259 (122 : 137)**	**100**

A. Overall prevalence of phantom limbs

Out of 259 total participants with limb difference included in the database, only 23% had never experienced phantom limbs. In addition, out of the 77% of individuals with limb difference who did report to have experienced phantom limbs in the past, 73% were continuing to experience them at the time of the survey was conducted. More in detail, 46% only presented phantom limb sensation, 27% reported phantom limb pain only, while the rest (27%) presented mixed phantom limb sensation and pain at some point in their life. See Figure 17 below for a visualisation of the differences in phantom limb experiences among the limb difference population assessed in this report.

When conducting the same analysis by gender, results seem to indicate that females may experience phantom limbs less frequently than males (absence of phantom limbs: 35% vs 12%, respectively). Finally, phantom limb sensation had a higher prevalence among males (Male: 48% vs Female: 42%), as well as phantom limb pain (Male: 34% vs Female: 17%), whereas the opposite was valid for experiences of mixed phantom limb sensation and pain (Female: 39% vs Male: 18%).

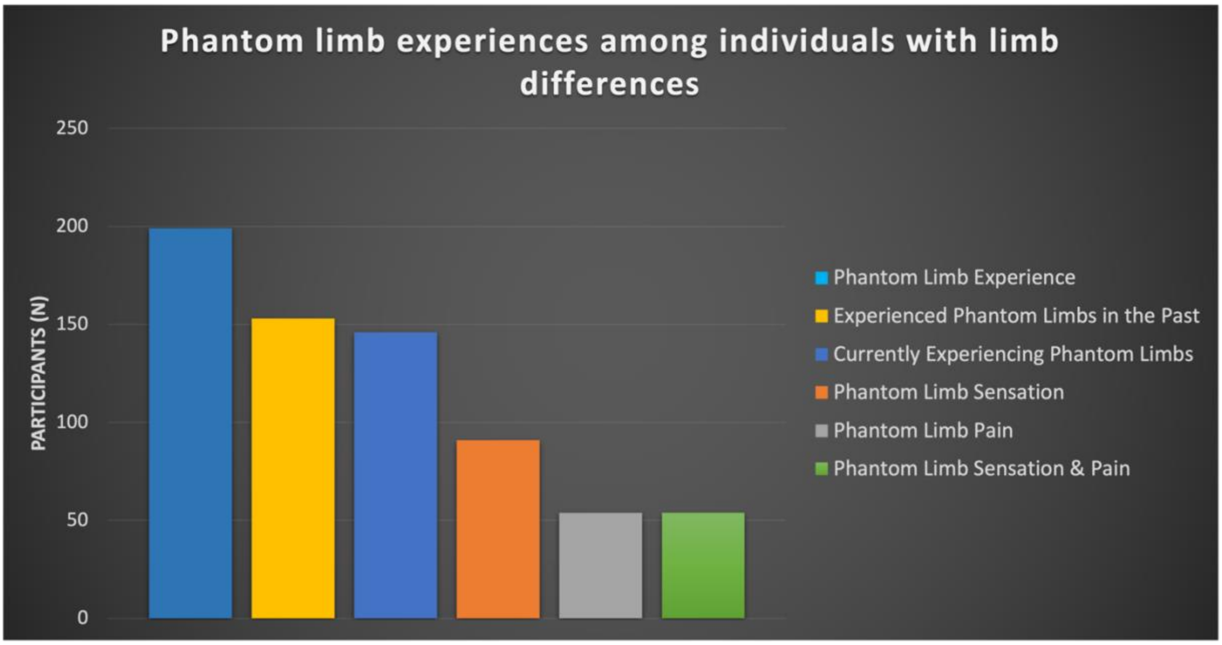


Figure 17. Difference in phantom limb experiences among individuals with limb differences in the database.

B. Prevalence of phantom limbs in the congenital subgroup

With respect to the congenital limb difference subgroup, out of 115 individuals, 47% had never experienced phantom limbs at the time of the survey. In contrast, of the individuals who did experience phantom limbs at some point during their life in this subgroup, 62% had only experienced phantom limb sensation, 36% only presented phantom limb pain, while the rest (2%) had mixed phantom limb sensation and pain.

When conducting the same analysis by gender, the female population of this subgroup was found to experience phantom limbs less frequently than males (absence of phantom limbs: Female: 80%, Male: 22%). In more detail, a higher prevalence of phantom limb sensation was seen in males (Female: 14% vs Male: 48%), as well as of phantom limb pain (Female: 3% vs Male: 29%) and mixed phantom limb sensation and pain (Male: 2%).

C. Prevalence of phantom limbs in the acquired subgroup

With respect to the acquired limb difference subgroup, out of 144 individuals, only 4% had never experienced phantom limbs at the time of the survey. In contrast, of the individuals in this subgroup who did experience phantom limbs at some point during their life, 37% had only experienced phantom limb sensation, 22% only reported phantom limb pain, while the rest (37%) had mixed phantom limb sensation and pain. See Figure 18 for a visualisation of the prevalence of phantom limbs experiences in individuals with congenital vs acquired limb difference included in the database.

When conducting the same analysis by gender, no difference in prevalence of phantom limbs experience by gender was found (absence of phantom limbs: Female: 4.2%, Male: 4.2%). Similarly, no difference in prevalence of phantom limb sensation by gender was seen. In contrast, a higher prevalence of phantom limb pain was observed in males (Female: 14% vs Male: 31%), while the opposite was valid for mixed experiences of phantom limb sensation and pain (Female: 46% vs Male: 28%).

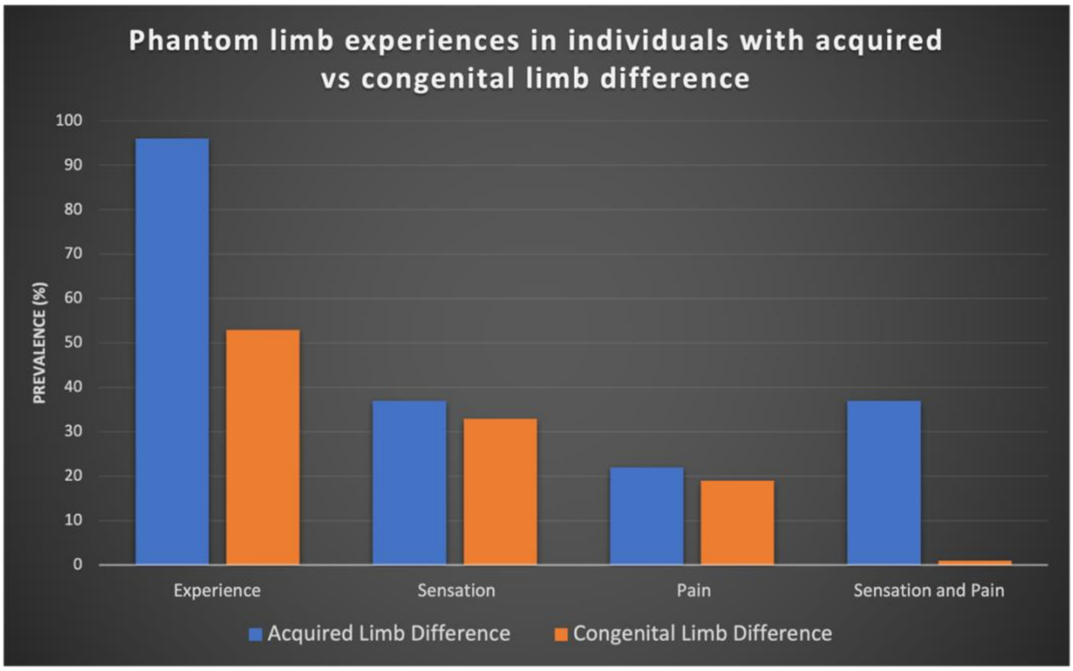


Figure 18. Difference in phantom limb experiences in individuals with acquired vs congenital limb difference from the database.

Use of prostheses

Use and type of prostheses among participants here investigated, as well as their prevalence by type of limb difference and gender. Please see Table 17 below for a summary of this data.

Table 17. Prevalence of use and type of prosthesis among participants of the database, by type of limb difference and gender.

	Participants **N (F : M)**	**Prevalence** **%**
Acquired Limb Difference	**144 (72 : 72)**	**55.6 (50 : 50)**
No use prostheses	**20 (10 : 10)**	**13.9 (13.9 : 13.9)**
Use prostheses	**124 (62 : 62)**	**86.1 (86.1 : 96.1)**
Cosmetic	5 (2 : 3)	3.5 (2.8 : 4.2)
Functional	110 (56 : 54)	76.4 (77.8 : 75.0)
Cosmetic, Functional	9 (4 : 5)	6.3 (5.6 : 6.9)
Congenital Limb Difference	**115 (50 : 65)**	**44.4 (43.5 : 56.5)**
No use prostheses	**45 (24 : 21)**	**39.1 (48.0 : 32.3)**
Use prostheses	**70 (26 : 44)**	**60.9 (52.0 : 67.7)**
Cosmetic	5 (2 : 3)	4.3 (4.0 : 4.6)
Functional	62 (21 : 41)	53.9 (42.0 : 63.1)
Cosmetic, Functional	2 (2 : 0)	1.7 (40.0 : 0)
Functional, Other	1 (1 : 0)	0.9 (2.0 : 0)
Total	**259 (122 : 137)**	**100**

A. Overall prevalence of use and type of prostheses

Out of 259 total participants with limb difference included in the database, only 25% did not make use of prostheses. Out of the 75% of individuals who did use prostheses, 5% used cosmetic prostheses, 89% functional prostheses, and 6% a mix of cosmetic and functional prostheses. Furthermore, no difference in prosthesis use by gender was observed. Please see Figure 19 below for a visualisation of the differences in prostheses use among the limb difference population assessed in this report.

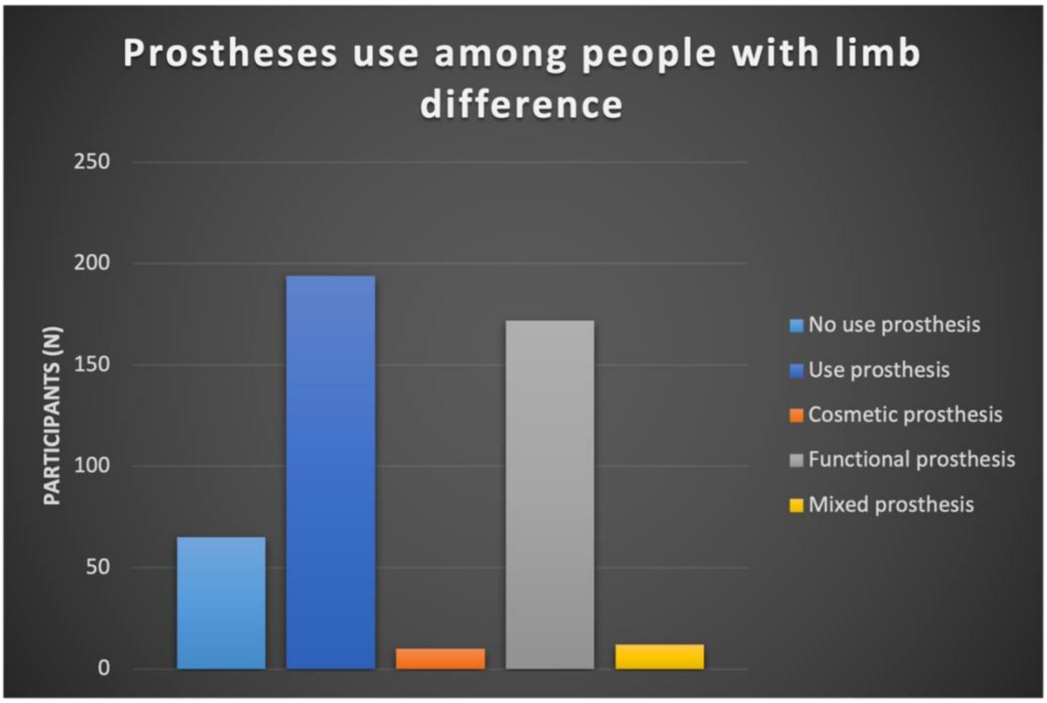


Figure 19. Difference in prostheses use among people with limb difference from the database, by type of prosthesis.

B. Prevalence of use and type of prostheses in the congenital subgroup

With respect to the congenital limb difference subgroup, out of 115 individuals included in the database, 39% did not report the use of prostheses. Nonetheless, among the 70 individuals in this subgroup who did utilise prostheses, the functional type of prosthesis was the most common (54%), regardless of gender.

C. Prevalence of use and type of prostheses in the acquired subgroup

With respect to the acquired limb difference subgroup, out of 144 individuals included in the database, only 14% did not make use of any type of prostheses. Among the individuals in this subgroup who did make use of prostheses, the functional prosthesis remained the most commonly used (76%). Please see Figure 20 below for a visualisation of the prevalence of use and type of prostheses among the acquired vs congenital limb difference population samples.

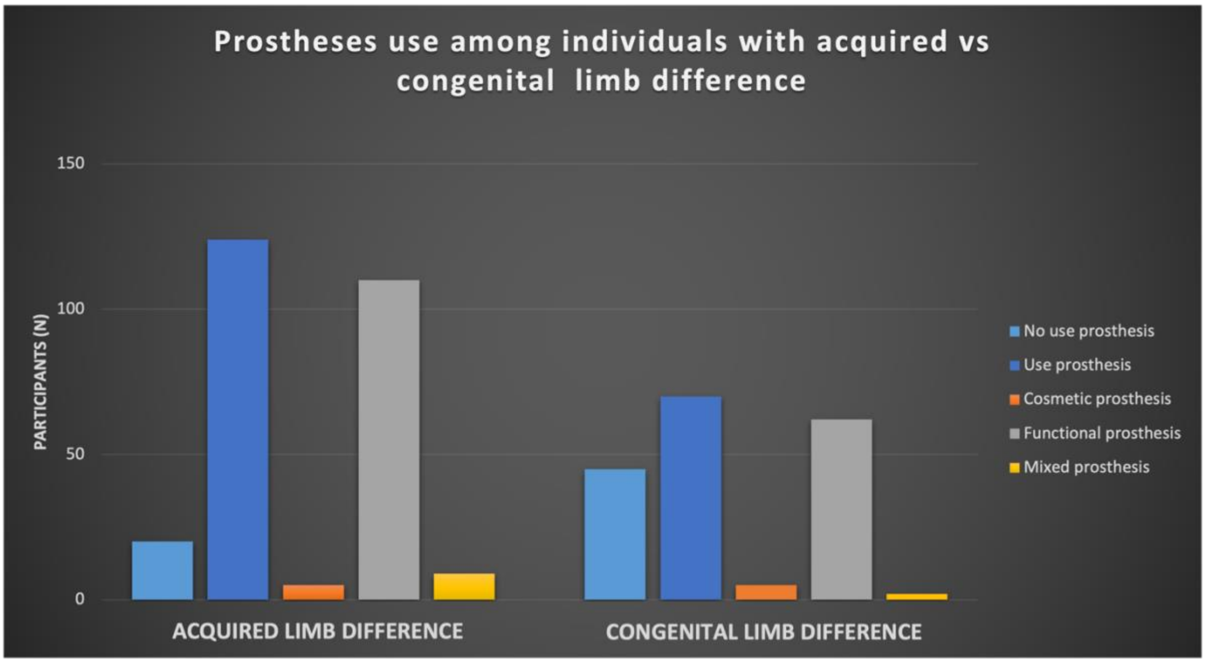


Figure 20. Difference in prostheses use among individuals with acquired vs congenital limb difference from the database.

Physical conditions and mental disorders

Next, an analysis was conducted to determine the prevalence of physical conditions and mental disorders in the limb difference population assessed in this report, also by type of limb difference and gender. Please see Table 18 below for a summary of the most common psychical conditions and mental disorders among participants.

Table 18. Prevalence of physical conditions and mental disorders among participants of the database, by type of limb difference and gender.

	Participants N (F : M)	Prevalence %
Acquired Limb Difference	**99 (49 : 50)**	**68.8 (68.1 : 69.4)**
Depression	28 (11 : 17)	19.4 (15.3 : 23.6)
PTSD, Depression	20 (14 : 6)	13.9 (19.4 : 8.3)
Migraine	7 (4 : 3)	4.9 (5.6 : 4.2)
PTSD	5 (3 : 2)	3.5 (4.2 : 2.8)
PTSD, Depression, Other	4 (2 : 2)	2.7 (2.8 : 2.8)
Depression, Other	3 (0 : 3)	2.1 (0 : 4.2)
Congenital Limb Difference	**68 (20 : 48)**	**59.1 (40.0 : 73.9)**
Depression	31 (8 : 23)	27.0 (16.0 : 35.4)
PTSD	16 (1 : 15)	13.9 (2.0 : 23.1)
Migraine	6 (2 : 4)	5.2 (4.0 : 6.2)
Anxiety, Depression	5 (5 : 0)	4.3 (10 : 0)
Head trauma	4 (1 : 3)	34.8 (2.0 : 4.6)
Migraine, Depression	4 (2 : 2)	34.8 (4.0 : 7.7)
Total	**167 (69 : 98)**	**64.5 (56.6 : 71.5)**

Abbreviations: PTSD: post-traumatic stress disorder.

A. Overall prevalence of physical conditions and mental disorders

Out of 259 total participants with limb difference included in the database, 65% reported to have some physical condition or mental disorder. In more detail, the most prevalent conditions were (a) depression: 35%, (b) post-traumatic stress disorder (PTSD): 13%, (c) comorbid depression and PTSD: 14%, (d) migraine: 8%, and (e) anxiety: 3%.

When conducting the same analysis by gender, I found males to be more likely to report mental disorders as opposed to females (72 vs 57%, respectively). In more detail, the following were observed to be more frequent among males with limb difference vs females (a) depression (16 vs 29%, respectively) and (b) PTSD (3 vs 29%, respectively). Compared to males, females were more likely to report the following mental disorders: (c) comorbid depression and PTSD (13 vs 6%, respectively) and (e) anxiety (4 vs 0%, respectively). Prevalence of (d) migraine was instead similar between genders (5%).

B. Prevalence of physical conditions and mental disorders in the congenital subgroup

With regards to the congenital limb difference group, out of 115 participants included in the database, 59% reported some physical condition or mental disorder. In more detail, the most prevalent mental disorders in this subgroup were (a) depression: 45% and (b) PTSD: 24%.

When conducting the same analysis by gender, the prevalence of physical conditions and mental disorders was higher in males vs females within this subgroup (74 vs 40%, respectively). In concordance, depression was more frequent in males vs females with congenital limb difference (35 vs 16%, respectively); the same was valid for PTSD (23 vs 2%).

C. Prevalence of physical conditions and mental disorders in the acquired subgroup

With respect to the acquired limb difference subgroup, out of 144 participants included in the database, 68.8% reported some physical condition or mental disorder. In more detail, the most prevalent disorders in this subgroup were (a) depression: 28% and (b) comorbid PTSD and depression: 20%. See Figure 21 below for a visualisation of the prevalence of physical conditions and mental disorders among the acquired vs congenital limb difference groups assessed in this report.

When conducting the same analysis by gender, I observed a similar prevalence of physical conditions and mental disorders between genders within this subgroup (Female: 68.1% vs Male: 69.4%). However, depression was most common in males vs females with acquired limb difference (24 vs 15%, respectively), while comorbid PTSD and depression were more frequent in females than males with acquired limb difference (19 vs 8%).

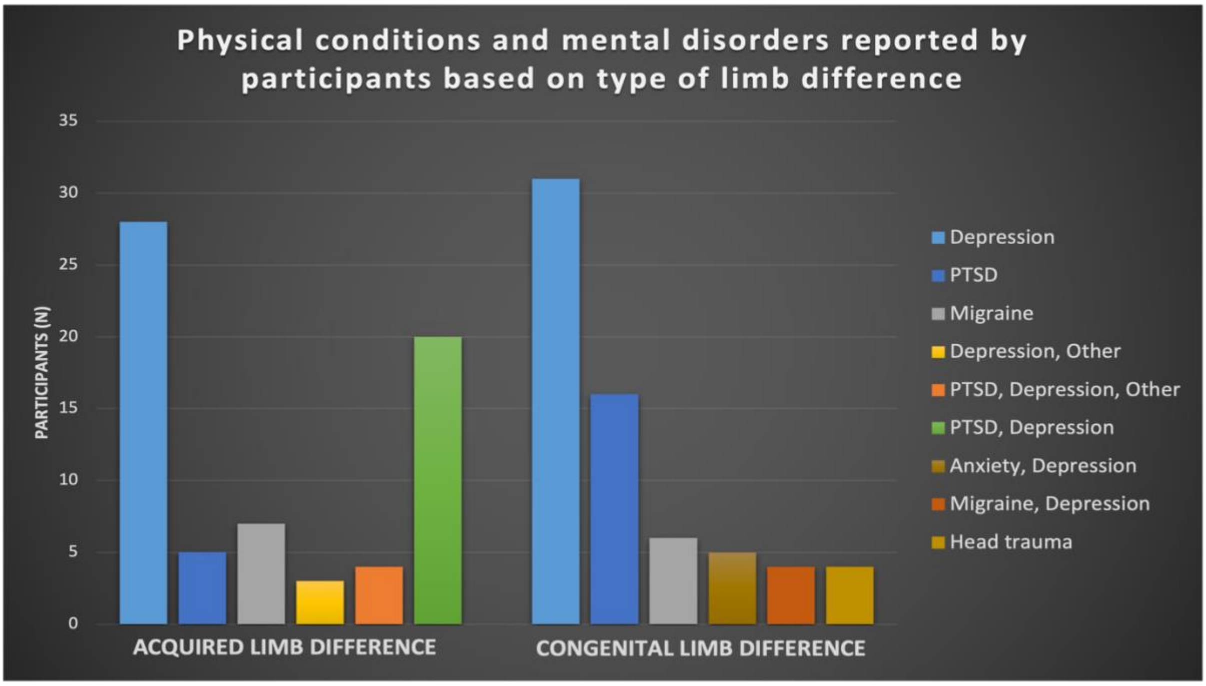


Figure 21. Difference in prevalence of physical conditions and mental disorders among the acquired vs congenital limb difference population from the database. PTSD: Post-traumatic stress disorder.

Language used to describe limb difference

Finally, the language used to describe limb difference was compared between individuals with congenital vs acquired limb difference. Specifically, I asked the following question to participants: “Please specify further where is your limb difference and what parts of your body are affected”. Responses were analysed according to the following four categories: (1) self-referential language (e.g. “my”, “mine”, “I have”, “I am”, “I lost”, “I miss”, etc.); (2) emotional language (e.g. “inconvenience”, “unfortunately”, “loss”, “missing”, “worse”, “affects”, etc.); (3) use of words associated with ownership (e.g. “my” or “mine”); and (4) use of words associated with non-possession (e.g. “not having”, “lost”, “miss”). Please see Figure 22 below for a visualisation of the prevalence of language used to describe limb difference in individuals with congenital vs acquired limb difference included in the database.

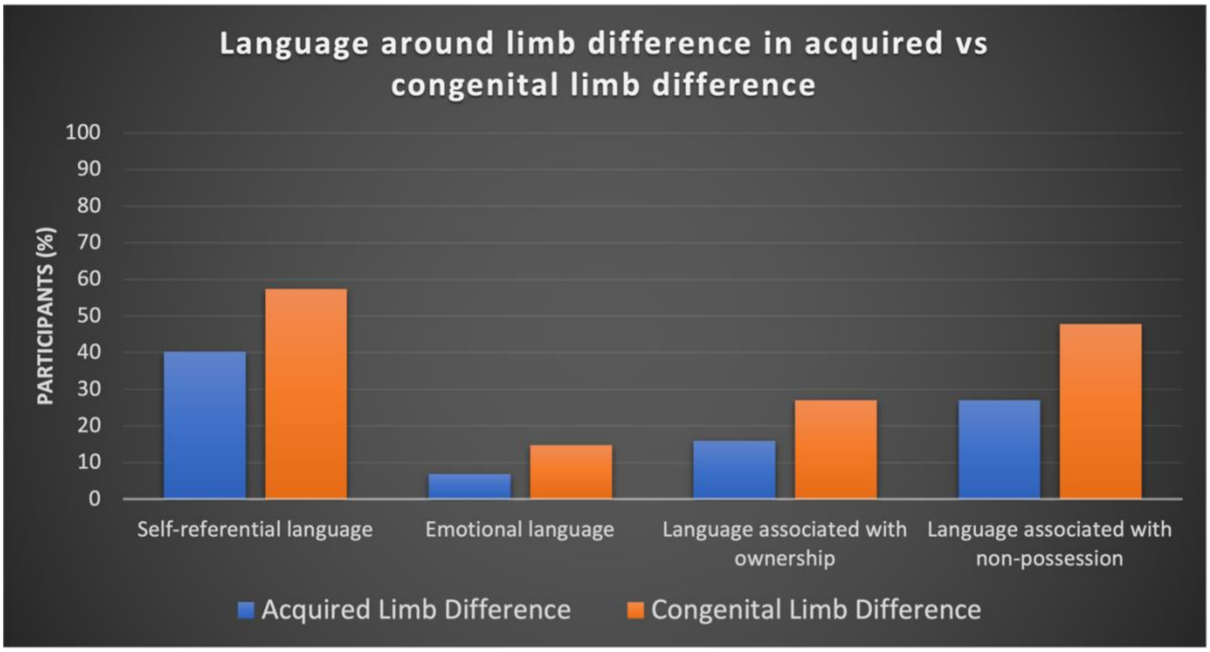


Figure 22. Differences in language used by individuals with acquired vs congenital limb difference to describe their limb difference from the database.

Overall, results suggest a heightened use of all categories of language above identified among individuals with congenital vs acquired limb difference. Specifically, I observed in the congenital limb difference group a higher prevalence of self-referential language (57.4%), emotional language (14.8%), use of words associated with ownership (27%), and use of words associated with non-possession (47.8%), compared to the acquired limb difference group.

Discussion

In this report, I investigated the prevalence of type, level and side of limb difference among participants included in the Research for Limb Difference database. I analysed these also by gender. Similarly, I assessed the prevalence of phantom limb experiences, prosthesis use, and physical conditions and mental disorders in the limb difference population. I analysed these also by type of limb difference and gender. Finally, I compared the language used to describe limb difference between individuals with congenital vs acquired limb difference included in the database. I will now discuss these findings in light of previous literature.

Prevalence of type of limb difference

Overall, a higher prevalence of acquired vs congenital limb difference was identified in this report. To my knowledge, there are no studies in the literature directly reporting the prevalence of congenital vs acquired limb difference within a sample of individuals with limb difference to which I can compare my findings. This is mainly driven by the fact that only a few studies in the literature include both individuals with congenital and acquired limb differences in their cohort (Ephraim et al., 2003; Kyberd et al., 1997; Melzack et al., 1997; Wilkins et al., 1998). For comparison purposes, I calculated the prevalence of type of limb difference in study by Melzack et al. (1997) which reported the number of participants within their limb difference sample who had congenital vs acquired limb difference. As a result, my findings are in accordance with data provided by Melzack et al. (1997), where a higher prevalence of individuals with acquired vs congenital limb difference was seen (76 vs 45%). Future studies assessing the relative prevalence of congenital vs acquired cases within a limb difference sample are warranted to provide more accurate estimates and inform future studies.

Furthermore, while gender seemed to be equally represented in the acquired limb difference subgroup, a higher prevalence of male individuals was observed in the congenital limb difference subgroup. With regards to such prevalence, mixed results are present in the literature. More specifically, Ephraim et al. (2003) and Fraser (1998) indicated a higher prevalence of male individuals in the acquired and congenital subgroups, respectively. In contrast, Kyberd et al. (1997) found a higher prevalence of female individuals with congenital limb difference. My results seem to support only findings by Fraser (1998). Kyberd et al. (1997) proposed that the prevalence of gender in acquired vs congenital limb difference may be driven by individuals' working and activities choices. Specifically, the authors suggested that a greater proportion of male individuals engage in more dangerous work and leisure activities compared to females, leading to traumatic amputations (thus the prevalence of male individuals with acquired limb difference seen in their study). My results do not support this interpretation and show instead a higher prevalence of male individuals in the congenital subgroup. At this stage, it remains unclear why such bias may occur.

Prevalence of level of limb difference

Overall, I found a similar prevalence of lower or upper limb difference among the limb difference population included in the database. However, when conducting this analysis by subgroups, a higher prevalence of lower limb difference was seen within the acquired limb difference subgroup. In contrast, a higher prevalence of upper limb difference was found within the congenital limb difference subgroup. These results partly support previous studies suggesting a higher prevalence of lower and upper limb difference in the acquired (Ziegler-Graham et al., 2008) and congenital (Mai et al., 2019) limb difference populations, respectively. My results are only partly support such studies as I found a lower prevalence both compared to Ziegler-Graham, et al. (2008) (acquired: 55.6 vs 90%, respectively) and Mai, et al. (2019) (congenital: 56.5 vs 66.6%, respectively).

Furthermore, both male and female individuals in the congenital limb difference group mostly presented upper limb difference, whereas those in the acquired limb difference group had lower limb difference. Overall, these results seem to indicate that gender may not influence the level of limb difference.

Prevalence of side of limb difference

A similar prevalence of side of limb difference (i.e. left or right side of the body) was observed overall among the limb difference population included in the database, as well as in both subgroups separately.

However, I observed an influence of gender on prevalence of side of limb difference. Specifically, right limb difference was more frequent in females with congenital limb difference, whereas left limb difference was mostly seen in males with congenital limb difference. In contrast, a similar prevalence of side of limb difference was observed in male and female individuals with acquired limb difference.

These results are partially supported by previous studies. Specifically, similarly to the findings here reported, Kyberd, et al. (1997) identified a left-side bias in male individuals with congenital limb difference. However, in contrast to my findings, female individuals with congenital limb difference were also found to have a left-side bias in both Kyberd, et al. (1997) and Fraser (1998).

Fraser (1998) hypothesised that such left-side bias may be resulting from exposures to negative factors during early embryo development when left limbs develop (rather than at later stages when right limbs develop). My results do not seem to support this hypothesis and possibly suggest that gender instead may play a role in such prevalence. It would be interesting to conduct future studies directly addressing a potential influence of gender during embryo development on side of congenital limb difference.

Finally, Kyberd, et al. (1997) identified a right-side bias in male individuals with acquired limb difference, which was however not identified in this report. As mentioned earlier, the authors proposed a tendency in the male population to conduct more dangerous work. Therefore, this result of theirs may be seen as supporting their proposal, considering that right-hand dominance is generally most common (hence the prevalence of male individuals with right-limb amputation). However, my results once again do not support this interpretation.

Prevalence of phantom limbs

Overall, my results indicate that most individuals with limb difference in the database have experienced phantom limbs at some point in their life, with a higher prevalence of phantom limb sensation vs pain vs mixed experiences of phantom limb sensation and pain. These results support findings by Krane & Heller (1995), although different sample sizes and proportion of individuals with congenital vs acquired limb difference were used. Nonetheless, Wilkins, et al. (1998) reported phantom limb pain and/or sensation in less than half of their population, thus contrasting my results. However, they assessed a much younger population (children and adolescents) compared to the adult sample here investigated, and age may have impacted results.

More in detail, results by type of limb difference suggest a higher prevalence of general phantom limb experiences in individuals with acquired vs congenital limb difference (difference of 43%). This is in accordance with previous studies (e.g. Melzack et al., 1997; Saadah & Melzack, 1994; Wilkins et al., 1998) reporting phantom limb experiences in individuals with congenital limb difference, although with a lower prevalence compared to that seen among individuals with acquired limb difference. The presence of phantom limbs in the congenital limb difference population has not been much evidenced in the literature. This is because it was not until recently that phantom limbs were only considered possible after the loss of a body part, as resulting from neural memory of the previously healthy limb (Price, 2006), and thus to be a prerogative of individuals with acquired limb difference. However, Saadah and Melzack (1994) suggested that neural representation of the body may be in part genetically determined, hence the presence of phantom limb experiences in individuals with congenital limb difference. An alternative explanation is provided by previous studies (e.g. Brugger et al., 2000; Melzack et al., 1997; Price, 2006), which proposed that phantom limbs in the congenital limb difference population can develop throughout development via low-level resonance and contagion and that body image can be shaped by visual and tactile modalities. Findings from Part 3, chapter 3.1 of this thesis may support this stance, by indicating increased emotional reactivity and possible simulation in individuals with congenital limb difference. Ultimately, the results here reported support the presence of phantom limb experiences in individuals with congenital limb difference.

Specifically, my results indicate a higher prevalence of phantom limb sensation and pain, as well as mixed experiences of sensation and pain, in individuals with acquired vs congenital limb difference. These findings are once again in concordance with previous studies (e.g. Saadah & Melzack, 1994; Wilkins et al., 1998) confirming the presence of both phantom limb pain and phantom limb sensation in individuals with congenital limb difference, although to a lesser extent compared to the acquired limb difference group. I speculate that the higher prevalence of phantom limb sensation and pain in the acquired limb difference population may result, respectively, from possibly stronger neural representations of the previously healthy limb (compared to the representations developed from resonance and visual and tactile modalities) and cortical reorganisation of motor and somatosensory cortices (MacIver et al., 2008). Additional studies addressing these phantom limb experiences further in the congenital limb difference population are warranted to elucidate and evidence through empirical data the underlying mechanisms.

Finally, my results indicate that the lower prevalence of phantom limb experiences in the congenital group is driven by the female population, possibly suggesting a relationship between gender and phantom limb experiences in the congenital limb difference population. Given the scarcity of literature on gender differences among individuals with congenital limb differences and phantom limb experiences, future studies addressing this are warranted to elucidate whether candidate mechanisms for the presence of phantom limbs in the congenital limb difference population may be directly influenced by gender, or factors interacting with gender. With respect to the acquired limb difference subgroup, a higher prevalence of phantom limb pain was seen among male vs female individuals, supporting previous studies in the literature (e.g. Hirsh et al., 2009).

Prevalence of prosthesis use and type

Overall, my results suggest that most of the people (75%) with limb difference included in the database made use of prostheses at the time the survey was conducted. Data on prosthesis use among the general limb difference population are difficult to find in the literature; therefore, directly compare my results with previous research is challenging. Nevertheless, Kooijman et al. (2000) indicated prosthesis use in 72% of their population with *upper* limb difference. Similarly, Datta, et al. (2004) identified a rejection rate of 33.75% among patients with acquired and congenital *upper* limb difference.

Furthermore, my results indicate that *functional* prostheses are most commonly used among individuals with limb difference included in the database (89%) vs *cosmetic* prostheses (5%) vs a mix of the two (6%). Nonetheless, a previous study conducted on individuals with *upper* limb difference identified a similar number of users of cosmetic and functional prostheses (Kyberd, et al. 1997).

More in detail, my results suggest a lower prevalence of prosthesis use among the congenital vs acquired limb difference population (75% decrease in use). This result has been previously proposed e.g. by James et al. (2006) to be associated with the extent of the impact of prosthesis use on individuals' lives. Specifically, the authors suggested that prostheses do not improve function and quality of life of individuals with congenital limb difference to the same extent that they do for individuals with acquired limb difference. Indeed, as also supported from this report's findings, congenital limb difference most frequently involves partial or complete absence or malformation of *upper* limbs, while acquired limb difference involves loss of *lower* limbs. Therefore, function and quality of life are more often nearly normal in individuals with congenital limb difference compared to those with acquired limb difference. For this reason, individuals with congenital limb difference may present a higher rejection rate of prostheses. In contrast, prostheses are generally more functional for individuals with acquired limb difference, thus possibly explaining such heightened prostheses use.

Finally, it seems therefore not unexpected that I observed a higher prevalence of functional prosthesis use among individuals, regardless of type of limb difference or gender. Nevertheless, there exists studies in the literature suggesting a higher prevalence of cosmetic prosthesis use in a cohort of individuals with acquired upper limb difference (e.g. Kooijman et al., 2000) or among females with acquired upper limb difference (Resnik et al., 2020).

Prevalence of physical conditions and mental disorders

Overall, my results suggest that 65% of individuals with limb difference who participated in this survey had some additional physical condition and/or mental disorder, with the male population suffering to a greater extent than female individuals from these. Furthermore, my results identified depression and PTSD to be common among this population, with a higher prevalence in the male group, while anxiety was higher in the female group. These results seem to be consistent with previous studies identifying significantly higher levels of depression and anxiety compared to the general population in individuals following a traumatic amputation (Mckechnie & John, 2014) or with congenital and acquired limb difference (Datta et al., 2004). Similarly, PTSD has also been previously indicated in the limb difference population, e.g. among soldiers with acquired limb difference in Sri Lanka (Abeyasinghe et al., 2012). Migraine was also found in about 5% of both the male and female populations. This was the only factor which was not affected by gender, thus possibly suggesting that gender influences mental disorders rather than physical conditions in the limb difference population included in this database.

In more detail, my results suggest a slightly higher prevalence of physical conditions and mental disorders among the acquired vs congenital limb difference population (68 vs 59%). While the prevalence of conditions and disorders was similar between genders in the acquired limb difference population, males with congenital limb difference were found to more frequently report a mental disorder compared to females with congenital limb difference. Therefore, my results possibly indicate a relationship between having a limb difference and the development of mental disorders guided by gender. For example, in the general population, depression is more prevalent among females (Abate, 2013), while my results suggest a higher prevalence of this disorder among males with limb difference in both groups. Nonetheless, age of both males and females has been associated with the prevalence of different disorders among the general population of Europe (King et al., 2008). Unfortunately, I did not collect this information from participants; therefore, I cannot provide conclusive remarks at this stage. However, I hope to inspire future studies in this direction.

Finally, these results suggest a different prevalence of common mental disorders among the acquired and congenital limb difference populations, with PTSD being more frequent in the acquired group and depression in the congenital group. The reason behind the higher prevalence of PTSD in acquired limb difference may be that generally limb amputation is a traumatic experience (Mckechnie & John, 2014). Furthermore, limb amputations are often reported in soldiers and PTSD has been recorded as one of the most common mental disorders in this population given their traumatic experiences during service (Abeyasinghe et al., 2012). It is more challenging to speculate, in contrast, on the reasons behind the high prevalence of depression in both groups, and especially in the congenital limb difference population. Previous studies have indicated many factors to come into play towards the psychosocial functioning of individuals with limb difference, e.g. social environment of the school setting (Varni et al., 1991), social deprivation (Wall et al., 2021), and perceived social stigma (Rybarczyk et al., 1995). Future studies addressing the reasons behind the different prevalence of mental disorders in the congenital vs acquired limb difference are warranted to identify and develop appropriate prevention schemes and coping strategies.

Prevalence of language used to describe limb difference

Overall, individuals with congenital limb difference assessed in this report more often referred to their limb difference in a self-referential way. This possibly suggests an increased ability of these individuals, compared to individuals with acquired limb difference, to speak of and refer to themselves or a heightened self-awareness. A reason behind this difference may be the changing body representations in individuals with acquired limb difference. Indeed, (Cowan, 1998) described self-reference to occur as the self represents a schema on which it is easier to understand others' schemas. Therefore, we may speculate that body schema of individuals with acquired limb difference may be blurred or changing, thus resulting in less frequent self-referential language. Nonetheless, time from amputation may have an influence on these results, provided that increased time from amputation may result in different representations of the body after limb loss or phantom limb experiences (MacIver et al., 2008), as well as the development of coping mechanisms following amputation (Oaksford et al., 2005). Unfortunately, I did not collect this information from participants; therefore, future studies addressing the existence of such correlations are warranted. Nonetheless, better scores at self-reported measures in studies presented in Part 2, chapter 3.1 of this thesis may support the stance of increased self-awareness among individuals with congenital vs acquired limb difference.

Furthermore, my results indicate a more frequent use of emotional language in the congenital vs acquired limb difference group when describing their limb difference, possibly suggesting a relation between limb difference vs limb loss and reference to emotions. Specifically, we could speculate that this increase in emotional language among individuals with congenital limb difference may be related to a more frequent exposure to emotional language associated with their limb difference throughout development compared to individuals with acquired limb difference, who face amputation at different stages in their life. This explanation may indeed be motivated by previous findings in psychology on an association between increased mental state language (including emotions) during development and enhanced mental state language later in life in the general population (e.g. Taumoepeau & Ruffman, 2006). Taumoepeau and Ruffman (2006) suggested an association between mother use of desire language with 15-month-old children and children's mental state language and emotion task performance at 24 months of age. While studies in Part 2, chapter 3.1 of this thesis may support this view by indicating higher scores at measures of mental state language in people with congenital vs acquired limb difference vs controls, it would be interesting for future studies to address the use of emotional language specific to individuals' description of their limb difference.

Finally, I aimed to better explore whether the limb difference was seen by participants as part of their body or as a missing part. Specifically, I did so by analysing the use of words associated with ownership and non-possession. By performing higher in both these categories, we might speculate that my results indicate that individuals with congenital limb difference may indeed have a mental representation of their limb difference as part of their body, which they however see as missing. In contrast, individuals with acquired limb difference may no longer, or to a lesser extent, consider their lost limb as part of their body and they less frequently consider it as missing from their body. This last point may seem counterintuitive, given that their limb was indeed once part of their body and that it is missing following amputation. While this effect may possibly be a result of coping mechanisms following amputation, at present, this remains a speculation for future studies to experimentally investigate.

Future studies addressing the validity of these findings in the limb difference population are warranted to determine whether varying limb difference may mediate different mechanisms associated with self-referential and emotional processing, as well as representations of the missing limb(s) or coping strategies. For example, (De Pisapia et al., 2019) identified the involvement of the mPFC in affective self-referential reasoning, with a key function of this brain region towards the processing of negative attributes. Therefore, studies addressing varying activity of this brain area in people with congenital vs acquired limb difference vs the general population may shed some light into potential differences in self-processing in virtue of having limb difference between groups and validate my data.

Conclusion

To conclude, in this report I provided a unified summary of prevalent characteristics among my limb difference population sample and directly described the congenital vs acquired groups. Furthermore, I introduced a new measure to be investigated in the limb difference population to better understand the psychology of having a limb difference. Considering that the results here presented not always replicated previous findings in the literature, care should be taken when interpreting them. Nevertheless, this report included data from a large sample of individuals with both types of limb difference, thus conferring relevance to my results. Overall, I believe that my results show scope for further research and can be informative for both academic and clinical purposes.

Part 3

On Theory of Mind: robots

1. Background

In the last decades, the fields of artificial intelligence and robotics have greatly advanced, resulting in the development of increasingly sophisticated virtual and physical intelligent agents. Furthermore, their increasingly complex abilities and behaviours have allowed their application to several everyday scenarios (e.g. Bhat et al., 2016; Görür et al., 2017; Hoffmann et al., 2017; Milliez et al., 2014). Indeed, interactions between AI and humans have nowadays become ubiquitous and heterogeneous, extending from autonomous driving cars to voice assistants or recommender systems supporting the online experience of millions of users. However, their application to sectors involving collaborative (or competitive) and communicative scenarios (e.g. assistive robotics, surveillance, entertainment, human-robot interaction (HRI), decision support, etc.) remains limited, and mutual understanding between robots and humans is a relevant issue to be addressed by research (Taniguchi, 2016).

Nonetheless, the integration of AI and robots among humans is still far from optimal, that is an ongoing issue for which two main explanations can be provided. On the one hand, AI and robotic agents' increasingly advanced motor and perception skills, as well as their improved humanoid features have enhanced humans' positive attitude towards them. However, the still limited social capabilities have a negative impact on humans' trust and acceptance of robots as social companions (uncanny valley effect) in their daily lives (Abubshait & Wiese, 2017; but also see Cavallo et al., 2018; Di Dio et al., 2020; Zanatto et al., 2019, 2020 for studies investigating trust and acceptance in HRI). On the other hand, humans have been often seen as a source of complexity, disturbance, and unpredictability that could affect autonomous systems' performance (e.g. Alami et al., 2005; Kulić & Croft, 2005; Mainprice et al., 2011; Sisbot et al., 2007), thus limiting their application.

Therefore, several clever robotic architectures have been created to equip robots with social skills and improve HRI (e.g. Demiris, 2007; Devin & Alami, 2016; Görür et al., 2017; Vanderelst & Winfield, 2018). While some of these architectures only aimed at equipping robots with social skills for an effortless interaction with humans (Lemaignan et al., 2017), others were inspired by our knowledge of human social understanding providing plausible models of human cognition (e.g. Patacchiola & Cangelosi, 2016; Scassellati, 2002; Vinanzi et al., 2019; Winfield, 2018). Indeed, getting inspiration from human behaviour, we know that efficient and natural interactions among humans are generally associated with their successful interpretation and prediction of others' mental states, which guide their behaviour (Taniguchi, 2016). Therefore, to achieve optimal integration of AI and robots in the society, systems with similar capabilities (thus having a ToM) need to be developed. ToM is at the base of most human higher-level social skills, such as collaboration, communication, imitation (Frith & Frith, 1999; Rakoczy, 2017; Tomasello et al., 2005) and is widely recognised to impact humans' success during social interactions (Devaine et al., 2014; Kovács et al., 2010). Nevertheless, building robots with a ToM, thus with the ability to understand others' intentions, beliefs and desires, remains amongst the "Grand Challenges of Science Robotics" (Yang et al., 2018, p. 9).

The implementation of some human-inspired ToM functions within a robotic architecture has been previously shown to provide robots with increasing social abilities (e.g. Demiris, 2007; Devin & Alami, 2016; Görür et al., 2017; Hiatt et al., 2011; Patacchiola & Cangelosi, 2016; Vanderelst & Winfield, 2018). Nonetheless, developing adaptive agents with the ability to *autonomously learn* how to model others' mental states and associated behaviours from a limited amount of data remains to be addressed. This ability has been elsewhere referred to as "Machine Theory of Mind"

(Rabinowitz et al., 2018) and has been proposed to enable more flexible interactions between robots and their human partners. Furthermore, integrating this adaptive ToM in robots has been envisaged to increase their application to situations in which specific data are not presently available due to high variability or volatility of environments and agents (e.g. searching and rescuing during disasters or everyday social settings) (Bianco & Ognibene, 2019; Lake et al., 2017).

Recently, an experiment part of a larger computational study about belief-based behaviours prediction performed by Rabinowitz et al. (2018) explored the learning of explicit belief representations (thus mental states). Specifically, they did so through a meta-cognitive observer (i.e. able to represent others' mental states) who predicted others' behaviours after the observation of a few of their previous behaviours. However, the authors deemed this approach to be limited due to the high computational demands connected with the complexity of beliefs representation. While recent algorithms have made the control and deliberation (not adaptation) of beliefs-based social behaviours more computationally affordable (e.g. Kominis & Geffner, 2017), Rabinowitz et al. (2018) also suggested that the supervisory signal needed to explicitly learn beliefs, as accessed through one's own mental states made available by meta-cognition, may be too biased. Furthermore, Rabinowitz et al. (2018) did not study sample efficiency of this approach nor compared the learning trajectory of their models compared to that of belief-based behaviours prediction models that did not explicitly learn to predict beliefs using their own mental states as teaching signal. This is particularly important considering the use of deep convolutional networks that are known to be data hungry, the variability of social interactions, and the high ecological cost that may be associated with misinterpreting others' actions in "social animals" like humans. Therefore, it remains to be determined which architectural principles could be suitable to endow intelligent systems with adaptive, autonomous and interpretable ToM.

Currently, the best example of autonomous development in real environments of an adaptive ToM is provided by humans, representing once again the “role model” for the advancement in robotics (Asada et al., 2009; Lungarella et al., 2003; Shimoda et al., 2022). Indeed, such cognitive skill has been suggested to be present from an early age during human development (e.g. Luo, 2011; Onishi & Baillargeon, 2005; Senju et al., 2011; see also Part 1 of this thesis). However, the developmental path of ToM leading to its emergence in humans remains unclear. Therefore, this makes it challenging to develop computational models reflecting the development of a human-inspired adaptive ToM. In fact, while it is now widely accepted that children older than 4 years of age show evidence of ToM ability, whether this cognitive ability can be observed at younger ages is yet to be confirmed (Apperly & Butterfill, 2009; Ruffman & Perner, 2005; Sodian & Kristen, 2016; Wellman et al., 2001). Findings in support of early ToM ability in infants comes from behavioural, as well as a few neuroimaging and computational studies (e.g. Hyde et al., 2018; Kampis et al., 2015; Luo, 2011; Onishi & Baillargeon, 2005), which evidenced *belief* understanding and tracking abilities in infants as young as 6 months of age. However, these interpretations have been contrasted by other non-ToM-related theories, including the low-level processing and behavioural rules theories (e.g. Heyes, 2014a, 2014b; Perner & Ruffman, 2005). Specifically, these theories propose that infants’ successful performance in such ToM-related tasks is not guided by their representation of others’ mental states but rather by their memory, attention and perception, statistical regularities or simple expectation of behaviour.

One psychology account that has been previously proposed to allow ToM engagement is the “like me” assumption (Meltzoff, 2007b, 2007a). Meltzoff (2007a) proposed that infants, as well as adults, can use themselves and their self-experience as a framework for understanding others and their mental states. Specifically, this account suggests the presence of a shared “supramodal” code which allows the understanding of others’ mental states and behaviours by representing them with a similar structure or “ontology” to that used to encode own mental states and behaviours. Therefore, it facilitates reasoning on and allows easier comparison between others’ representations and representations of own behaviours and underlying mental states. This approach therefore highlights that shared representations are at the base of social cognition, rather than the result of complex developmental and inferential processes driven by an initially solitary representation of the self, as suggested in contrasting theories (e.g. see Piaget, 1952, 1954). While this “like me” assumption has been adopted in infant studies assessing action, perception, emotion and imitation (Meltzoff, 2007b, 2007a), it has also been previously implicated in computational mechanisms proposed to underlie ToM, including association (e.g. Prinz, 1997), simulation (e.g. Meltzoff, 2007b), and teleological (Csibra & Gergely, 2007) mechanisms. Crucially, the “supramodal” code, part of the “like me” framework, can be considered more abstract than the above-mentioned mechanisms as it does not make specific assumptions on the type of mechanism using the code. Therefore, it is able to engulf all above mechanisms (see Figure 23 below).

To conclude, the “like me” assumption for others’ *beliefs* representation supports the early ToM account. Specifically, it does so by providing a means to explicitly and interpretably represent and infer others’ mental states through autonomous learning from an early age, without resorting to communication or other forms of external, hand-crafted learning signals. Therefore, the implementation of this assumption in robotic and AI architectures may represent one of the ways forward to further socio-cognitive capabilities of adaptive robots and intelligent agents, contributing to the resolution of one of the main current challenges in robotics.

1.1 Related work

Previous attempts (e.g. Devin & Alami, 2016; Giese & Rizzolatti, 2015; Görür et al., 2017; He et al., 2016; Kennedy et al., 2009; Rabinowitz et al., 2018; Raileanu et al., 2018; Vanderelst & Winfield, 2018; Winfield, 2018) exist in the literature creating social agents able to predict their partner's intentions or computational models which allow the study of the development of ToM functions. However, they mainly focused on the final performance of intention prediction, disregarding the developmental trajectory and the learning of explicit representations of others' beliefs for intention and behaviour prediction. See Table 19 below for a summary of relevant robotic or computational implementations in the literature addressing inference vs learning of others' mental states vs learning of others' beliefs. These are described following.

Table 19. Summary of relevant robotic or computational implementations in the literature addressing inference vs learning of others' mental states vs learning of others' beliefs.

Study	**Inference of others' mental states**	**Learning of others' mental states BUT beliefs**	**Learning of others' beliefs**
Devin & Alami (2016)	✓	X	X
Görür et al. (2017)	✓	X	X
Hiatt et al. (2011)	✓	X	X
Demiris & Khadhouri (2006)	✓	X	X
Winfield (2018)	✓	X	X
Baker et al. (2017)	✓	X	X
Hamlin et al. (2013)	✓	X	X
Patacchiola & Cangelosi (2016)	✓	X	X
Asakura & Inui (2016)	✓	X	X
Ramirez & Geffner (2011)	✓	X	X
Kominis & Geffner (2015, 2017)	✓	X	X
Zeng et al. (2020)	✓	X	X
Raileanu et al. (2018)	✓	✓	X
He et al. (2016)	✓	✓	X
Rabinowitz et al. (2018)	✓	✓	✓
Breazeal et al. (2009)	✓	✓	✓
Kennedy et al. (2009)	✓	✓	✓

Inference and representation of mental states (including beliefs), but not autonomous learning

Previous studies exist in the literature presenting compelling artificial architectures and computational models which allow the representation or inference of others' mental states, including beliefs, for predicting or evaluating others' behaviours. Nevertheless, they did not focus on the learning of explicit representations of others' beliefs, but rather on the final performance of intention prediction, while some assumed a learning free simulation-based approach or availability of explicit communication channels.

For example, Devin and Alami (2016) integrated a ToM model in their robot control architecture which permitted the estimation of both the state of the environment and human partners' internal states, in particular goals and plans, which were considered by the robot for successful human-robot shared plans performance. A hand-crafted list of possible actions and goals, as well as associated status was readily provided to the robots. Görür et al. (2017) used a Hidden Markov Model (HMM) to estimate actions performed by interacting agents in a collaborative task and incorporate human emotional states. The authors provided robots with a hand-crafted set of action states, while the emotional states were acquired by the robot in the form of human reactions to the robot's judgment on each action.

Hiatt et al. (2011) implemented ToM for equipping robots with the ability to interpret human behaviour variability and unexpected actions during team operations. Specifically, using the simulation approach, the authors allowed their robot to simulate hypothetical models of others which differ in their knowledge (beliefs) about the world. Furthermore, the authors relied on communication between robot and interacting actors to disambiguate actors' unexpected actions and behaviours. Another example is provided by Demiris and Khadhouri (2006), who advanced a similar approach in their work also using the simulation approach. Briefly, the authors described a simulation-inspired computational architecture that allows robots to select and execute an action, as well as understand it when shown by a demonstrator (thus predicting the agent's goals and future behaviour). Winfield (2018) also implemented a simulative computational model for ToM. Specifically, the authors provided robots with a simulation-based internal model of itself and its environment, including other agents, which could test (i.e. simulate) next possible actions and anticipate likely consequences, both of the robot itself and others.

Baker et al. (2017) presented a compelling Bayesian computational model of ToM (BToM), which was found to accurately infer mental state judgements of human participants in the "food track" experiment. Briefly, this computational model formalises ToM as a Bayesian inference about unobserved mental states (beliefs, desires, percepts) of a POMDP agent, based on observed actions. Hamlin et al. (2013) also used a Bayesian ToM computational model to identify whether 10-month-old infants' social evaluation was driven by an analysis of the mental states that motivated others' behaviours. Crucially, by showing that the ToM model better explained participants' behaviour, as opposed to other non-ToM models, the above studies provided evidence of reliance on the others' beliefs representations for predicting or evaluating others' behaviours.

Patacchiola and Cangelosi (2016) also relied on a Bayesian framework to infer others' mental states. Briefly, the authors made use of Bayesian Networks, including representations of self and others' beliefs and actions, to investigate the relationship between ToM and trust. Interestingly, adopting a developmental approach and Bayesian inference in a simulated environment, the authors provided evidence of differing associations between others’ actions and beliefs based on the presence of a mature or immature ToM ability. Ultimately, such differing association resulted in differing ability to distinguish helpers and hinderers, thus to determine who to trust.

Asakura and Inui (2016) provided a Bayesian framework for false-belief reasoning in children, which integrated both the simulation and teleological theories. Specifically, the authors indicated that children rely on teleological-based reasoning to infer others’ beliefs states (other model), but they highlighted that this reasoning also entails simulation-based reasoning based on an internal model of one’s own mind (self model). The authors found their model to provide a good fit to a variety of children ToM data.

Ramirez and Geffner (2011) focused on the inference of a probability distribution over possible goals of an agent whose behaviour results from a POMDP model shared between the agent and the observer. Kominis and Geffner (2015, 2017) applied this approach to multiagent settings, in which agents share a common goal and plan with beliefs about the world and consider nested beliefs of others in an online fashion.

Zeng et al. (2020) proposed a brain-inspired model of ToM, based on the implementation of the functions of different brain areas implicated in false-belief understanding as seen through human empirical evidence (e.g. STS: sensitive to biological motion; pSTS: implicated in understanding others’ actions and perspective taking, precuneus: implicated in mental imagery, etc). The authors focused on the belief reasoning pathway between such brain areas. Nevertheless, their model was rather false-belief task-specific as it was based on object permanence and visual access pathways, which are important in false-belief tasks but not necessarily crucial for understanding others' intentions in more generic beliefs-driven behaviours (e.g. searching).

Learning of mental states other than beliefs

Other studies made use of neural network-based models to investigate the learning of others' intentions for predicting others' behaviours; however, they did not investigate the learning of beliefs.

For example, I already reported about ToMnet, an artificial neural network developed by Rabinowitz et al. (2018) that, using meta-learning, was able to learn to predict differences in others' behaviours after the observation of a few of their previous behaviours (without explicit beliefs representation). In their studies, the authors investigated the network's ability to characterise different features of the observed agents, such as field of view, and its generalisation ability across different classes of agents.

Raileanu et al. (2018) and He et al. (2016) used reinforcement learning instead for equipping artificial agents with the ability to learn to collaborate with others based on their intentions. Briefly, Raileanu et al. (2018) introduced the Self Other-Modeling (SOM) approach to solve multi-agent adversarial and cooperative tasks, driven by a reward function based on intentions of both agents in the task. Similarly to other simulation models (e.g. Demiris, 2007; Ognibene & Demiris, 2013), the SOM approach allowed the agent to use its own policy to predict the other agent's actions and underlying intentions in an online manner. During the game, the agent inferred the other agent's hidden goal by directly optimising over the goal using its own action function to maximise the likelihood of others' actions. He et al. (2016) instead focused on jointly learning a policy and the behaviour of opponents, as well as their strategies using a feed-forward network. Specifically, the authors of this paper resorted to multitasking to explicitly model other agents' actions and strategies, using the following two supervisory signals, i.e. others' actions in the current state and mode (defensive vs offensive).

Learning of beliefs

Only a few papers in the literature directly addressed the learning of explicit beliefs representation for the prediction of others' behaviour. Considering beliefs for an adaptive ToM architecture is important, given that others' behaviours aimed at achieving goals are determined by actors' current beliefs.

As mentioned earlier, in a side experiment, Rabinowitz et al. (2018) explored the learning of explicit belief representations using a meta-learning approach. However, the authors deemed their approach likely not to scale to real-world situations due to the challenges of acquiring a supervisory signal for learning beliefs. Furthermore, the authors did not investigate the learning trajectory of their model. Therefore, it remains to be determined whether this approach is usable in online adaptive and social robots.

Breazeal et al. (2009) and Kennedy et al. (2009) attempted to address the acquisition of a supervisory signal for learning beliefs using the "like me" assumption. Specifically, Breazeal et al. (2009) provided their robots with a representation of beliefs, in the form of representations integrating perceptual features from raw sensory information, such as spatial relationships between observations and other metrics of similarity (knowledge of the world). The robot developed beliefs about the world from its own reference frame and, through a simulation mechanism, it reused the same mechanisms used for its own belief modelling but transformed and filtered the incoming data stream from others' visual perspective to predict others' beliefs. Kennedy et al. (2009) also used a "like me" simulation to represent and predict others' beliefs. The authors equipped their robot with the ability to create an imagined representation of the world by taking the agent's visual perspective. The robot then used its own generated goal for the simulation and associated it to the decision-making of the interacting agent. Following this, a new declarative fact was created for the specific situation and the response saved for future use. This therefore led to learning instances, remembering them and then using them to generate new goals.

While these studies provided compelling architectures for social robots, they integrated strong and less flexible forms of the "like me" assumption (i.e. simulation), which use the same mechanisms to both control and recognize actions (Demiris & Khadhouri, 2006; Giese & Rizzolatti, 2015; Ognibene & Demiris, 2013). Due to the correspondence problem previously outlined (Brass & Heyes, 2005; Nehaniv & Dautenhahn, 2002), as well as the limits of generative methods when applied to inputs that do not satisfy their assumptions (Ng & Jordan, 2002; Prasad et al., 2017), these models provide a limited solution for adaptive ToM. This is true if not considering compensatory mechanisms that may aid the understanding and prediction of behaviours of others who differ from the self.

Furthermore, similarly to Rabinowitz et al. (2018), these studies did not investigate the impact of learning to explicitly represent others' beliefs on inferring others' intentions and behaviours. Nevertheless, these simulative models do not offer a fair evaluation of such an effect given that their recognition performance steeply decays for actions that the observer himself would not be able to represent and produce (Nehaniv & Dautenhahn, 2002). As a result, beliefs representation is a prerogative of these models to recognize beliefs-driven behaviours and the removal of such representations would result in extensive architectural and performance changes.

Overall, the developmental trajectory (or learning) of explicit representations of others' beliefs and the nature of their impact on others' intention prediction performance is yet to be identified. In other words, whether the “like me” approach for learning to explicitly represent others' beliefs provides an advantage towards predicting others' behaviours, thus resulting in a usable approach for adaptive and social robots with increasing ToM skills, remains to be addressed.

2. Theoretical and Methodological Contributions

2.1 Proposal

In this thesis, I propose that learning to explicitly represent others' beliefs provides an advantage towards predicting others' behaviours, in line with the ToM account previously outlined. Learning to explicitly represent beliefs may impose demanding and complex constraints on performance and the computational mechanisms associated with others' behaviour prediction (as suggested e.g. by Rabinowitz et al., 2018). However, I hypothesise this ability may help render the prediction of others' behaviour more accurate and improve the learning trajectory of this skill. This will resolve in part the limitations associated with the high variability of behaviours in different contexts without increasing the amount of training samples (which are difficult to obtain and ecologically expensive) and bridge the gap between human and AI learning speeds.

This hypothesis is driven by our knowledge from the psychology "like me" assumption of social cognition, which I adapt in this thesis to a "like them" assumption to show that developing the capability of learning to explicitly represent beliefs through a shared representational framework may (a) pose architectural demands that are simpler than previously believed, (b) contribute in speeding-up the acquisition of socio-cognitive prediction skills, (c) strongly improve the interpretation of beliefs-driven behaviours, and (d) increase the generalisation ability to predict behaviours of others acting in different environments. Overall, my proposal aims to provide architectural suggestions for social artificial intelligent systems, such as virtual companions and social robots, and it is also relevant for shedding some light on developmental neuropsychology debates surrounding ToM.

2.2 The “like them” assumption

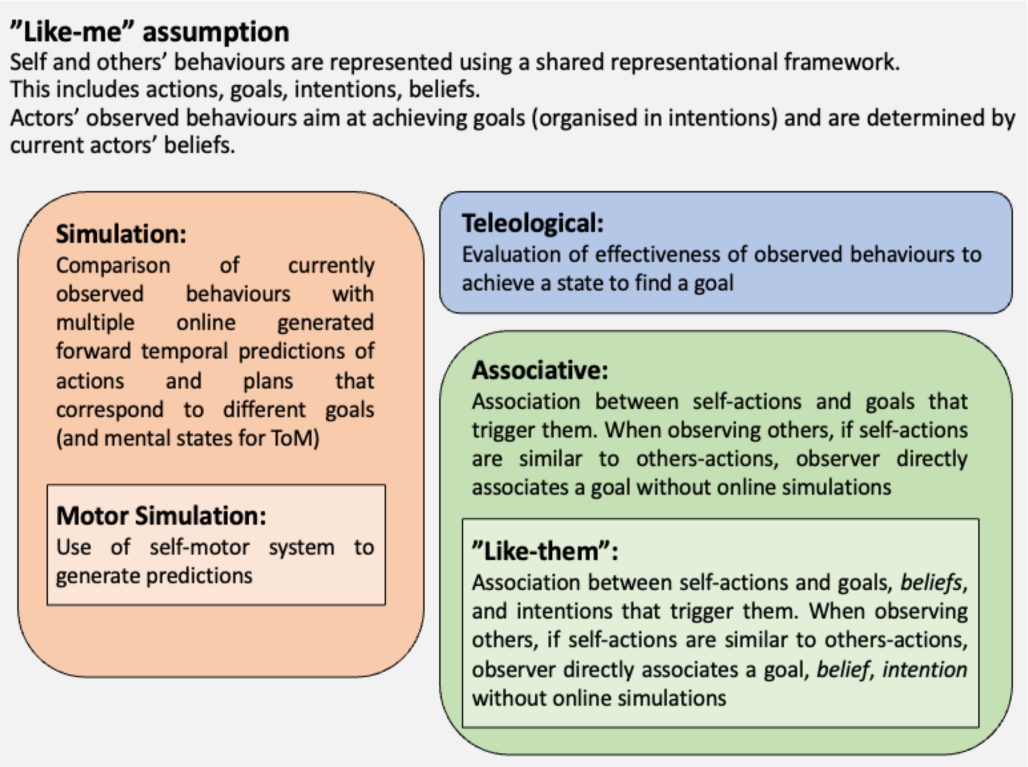
To tackle the demanding and complex constraints for artificial systems of not having a readily available supervisory signal for learning to explicitly represent beliefs (Rabinowitz et al., 2018; Schlinger, 2009; Spradlin & Brady, 2008), I resort to the “like me” assumption of social cognition. However, I propose an adaptation of such assumption that I name the “like them” assumption, which provides the rationale for the experimental research conducted in the next chapter (Part 3, chapter 3).

In the “like them” approach, I maintain the “like me” assumption of a shared representational framework between self and others to reason about mental states. However, to understand observed others’ behaviours (and underlying mental states), this variation does not rely on shared representations as simulated by the observer’s control and planning systems, with the assumption of self-other similarity (as in the direct-matching hypothesis). In contrast, this relation is inverted. Rather than generating trajectories on the fly, an individual’s social perception system is prepared or better trained before another’s action is observed, thus more readily able to understand others and predict their behaviour.

Under the “like them” assumption, through a mechanism of self-observation, one’s own mental states while performing a task are assumed to be similar to future putative mental states and actions of other agents (“like them”) performing the same or similar tasks. Therefore, own (mental) states and actions provide a pair of supervisory signals and observed samples to train a predictor of others' mental states. In other words, this means that self-representation of beliefs, which is formed through self-experience and inference processes, acts as a supervisory signal which, through the shared representational framework, more readily allows the understanding and prediction of others’ beliefs and, ultimately, behaviour.

My “like them” proposal can also be seen as in line with the associative hypothesis, which has been described by James (1890) as a mechanism linking action-effect representations through bidirectional associations. My approach follows previous research (e.g. Hommel et al., 2001, as described in Csibra & Gergely, 2007) which extended this associative hypothesis to intention interpretation. This approach proposes that associations between behaviours (actions) and underlying intentions (effect) develop from an early age and can be used later in life as a means to infer and predict others’ intentions and behaviours. The associative account has also been previously related to others’ understanding by proposals of its involvement in mirror neurons development (e.g. Heyes, 2010), as well as in sensorimotor matching for imitation (Decety & Chaminade, 2003). My work differs from these accounts as it focuses on beliefs (Demiris, 2007). It relies on associations between explicit belief representations learnt through self-experience and consequent behaviours to improve prediction performance of others’ beliefs-driven behaviours. See Figure 23 for a diagram of the “like me” assumption and the mechanisms using such shared representational framework, including my “like them” assumption.

Overall, this “like them” assumption may thus be able to overcome the lack of an explicit supervisory signal for the prediction of others’ mental states. Nonetheless, using oneself as a model for the prediction of others’ behaviour may still lead to bias. However, Gilbert and Malone (1995) have previously suggested that, while such bias may have negative consequences, it can still be beneficial towards the understanding of others’ behaviours, even when the resulting predictions from this bias are not completely accurate. Furthermore, the main demand of the “like them” architectural approach (i.e. the propagation of teaching signals from the control and executive processes to the social perception ones) is less constraining than the ones required by a full direct-matching approach. Indeed, the latter re-uses most of the same neural circuitry for both processes (Giese & Rizzolatti, 2015) and would result in increasingly biased and noisy predictions due to the correspondence problem (Brass & Heyes, 2005; Nehaniv & Dautenhahn, 2002). Ultimately, while my approach does not imply the direct matching of self-other representations, it can still support high-level cognitive skills by ensuring representational matching through a shared “supramodal” code.



”Like-me” assumption

Self and others’ behaviours are represented using a shared representational framework.
This includes actions, goals, intentions, beliefs.
Actors’ observed behaviours aim at achieving goals (organised in intentions) and are determined by current actors’ beliefs.

Simulation:
Comparison of currently observed behaviours with multiple online generated forward temporal predictions of actions and plans that correspond to different goals (and mental states for ToM)

Motor Simulation:
Use of self-motor system to generate predictions

Teleological:
Evaluation of effectiveness of observed behaviours to achieve a state to find a goal

Associative:
Association between self-actions and goals that trigger them. When observing others, if self-actions are similar to others-actions, observer directly associates a goal without online simulations

”Like-them”:
Association between self-actions and goals, *beliefs*, and intentions that trigger them. When observing others, if self-actions are similar to others-actions, observer directly associates a goal, *belief*, *intention* without online simulations

Figure 23. “Like me” assumption and comparison of the characteristics of the mechanisms using this shared representational framework.

2.3 Multi-task learning

While my approach is inspired by the “like me” developmental psychology hypothesis, its theoretical basis comes from machine learning, where learning by self-observation can be seen as a form of transfer learning (Pan & Yang, 2010). Specifically, learning by self-observation allows individuals to use the observation of one’s own mental states while performing tasks to support prediction of others’ intentions in similar tasks. In turn, this results in the assumption of one's own and others’ actions and beliefs as sampled from related distributions. Transfer learning can be seen as a subclass of meta-learning (Langdon et al., 2022), i.e. the ability to use existing models and knowledge, e.g. derived from first-person experience, to efficiently solve new tasks and interact with novel agents (Langdon et al., 2022).

I consider for the current implementation another form of meta-learning, multi-task learning (Caruana, 1998; Crawshaw, 2020; Ruder, 2017). Multi-task learning has been indeed indicated by Omidshafiei et al. (2017) as a means to increase the applicability of deep learning models to real-world scenarios and to deal with tasks involving non-observable components. Multi-task learning involves the simultaneous learning of multiple tasks by a shared model in order to improve learning performances (Crawshaw, 2020; Ruder, 2017). Specifically, this approach has been associated with several advantages, including improved data efficiency and implicit data augmentation through reduction of noise patterns, reduced overfitting through shared representations, reduced representation bias, attention focusing on relevant features, and faster learning in virtue of complementary information between related tasks (Crawshaw, 2020; Ruder, 2017). Furthermore, multi-task learning has been previously suggested to be the most relatable type of learning to human learning, as humans are rarely presented with single tasks in isolation and they instead rely on information from different modalities to build their knowledge (Crawshaw, 2020).

In the simultaneous multi-task learning approach commonly adopted in deep neural networks, multiple output sub-networks (*heads*) receive information from a shared input subnet (trunk). During learning, the shared trunk simultaneously faces the informational requirements, in the form of error backpropagation, of the different heads. While this may increase the complexity of the task that the trunk must face, if the predictions from different heads share common aspects, and thus pose compatible constraints, the presence of multiple tasks increases learning efficiency by helping to discard solutions that would lead to overfit on one head, thus providing a form of regularisation. This approach easily allows the removal of one prediction output, through the elimination of one *head*, without directly affecting the functioning of the other *heads*. Therefore, this allows an easy estimation of its impact on learning prediction of the other variables by directly comparing performance between models that present or not that *head*.

Given the above, I hypothesise that using multi-task learning would improve the generalization of actor's target prediction exploiting the domain-specific information coming from the prediction of actor's beliefs. In other words, I hypothesise that additionally learning to explicitly represent beliefs would result in more effective predictions of others' intentions.

3. Experimental Contributions

This chapter is structured as a sequence of experiments of increasing complexity comparing predictive performance of others' behaviour by a "ToM observer", who is able to explicitly represent others' beliefs, and a "simple observer", who does not have this ability. Following the rationale presented in the previous chapter of this thesis (Part 3, Chapter 2), I hypothesise that the "*ToM observer*" will exhibit an advantage towards others' behaviour prediction, also when faced with varying observers, actors, and environmental complexities. This would be a direct result of its ability to learn to explicitly represent beliefs following a "like them" approach, which supposedly has a beneficial effect on predicting others' intentions and behaviours. I expect this to be valid both in terms of accuracy and sample complexity or learning speed, allowing for important implications in different research fields.

Material and methods

The architectures

In this thesis, I employed two neural architectures which differ in their ability to learn to predict (or not) others' beliefs while they are performing a task, resulting in the *Beliefs* and *NoBeliefs architectures*, respectively. The implementation of these architectures is adapted from Rabinowitz et al. (2018). These neural architectures present a shared torso network, that processes an input tensor representing recent states of the observed actor as well as the environment and objects configuration, and three *prediction heads*, including the (1) target position, (2) action, and (3) state, to predict respectively the position of the observed actor's target, the next action and the next state of the observed actor. However, the *Beliefs* architecture has an additional (4) belief *head*, resulting in a total of four *prediction heads*, which is used to predict the beliefs of the observed actor with regards to the target position. All these *heads* are fed by the torso output. See Figure 24 below.

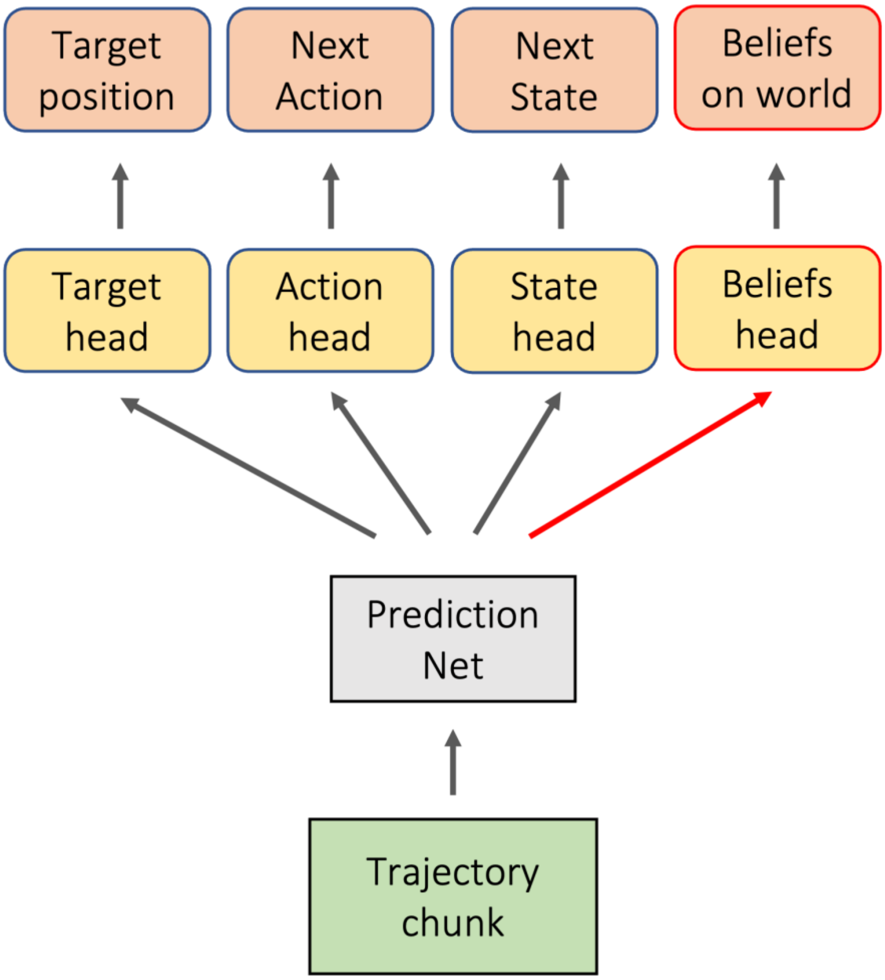


Figure 24. Visualisation of the architecture utilised in the here reported studies, formed of a shared prediction net torso and subsequently of separate prediction heads. For the *NoBeliefs* architecture the following prediction heads are considered: 1. Target position, 2. Actor's next action, and 3. Actor's next state. For the *Beliefs* architecture, the 4. belief prediction head (in red) is also considered.

Environment and observed actors

For all experiments in this study, the environment consisted of 11x11 grid world maps, which varied in the location of walls, columns, and free cells to move around the map (see Figure 25 for a visualisation of example grid world maps). To assess a developmental trend in my experiments, I built training datasets comprising different numbers of maps; i.e. 5, 10, 15, 20, 25, 30, 60, 120, and 300 maps. The environment also enabled a common action space (north, east, south, west, northeast, northwest, southeast, southwest, stay) with deterministic results. A total of 30 trajectories were generated per grid world map by randomly selecting the initial locations of both the actor and target for each trajectory. For example, when considering the dataset with 60 grid world maps, a total of 1800 actors' behaviours were created, while 9000 total behaviours were created for that with 300 maps. At training and test time, 3 distractor objects, identical to the target, were randomly positioned (unless otherwise specified) at different empty locations in the map. This enabled a wider environment and behavioural diversity with no additional computational cost.

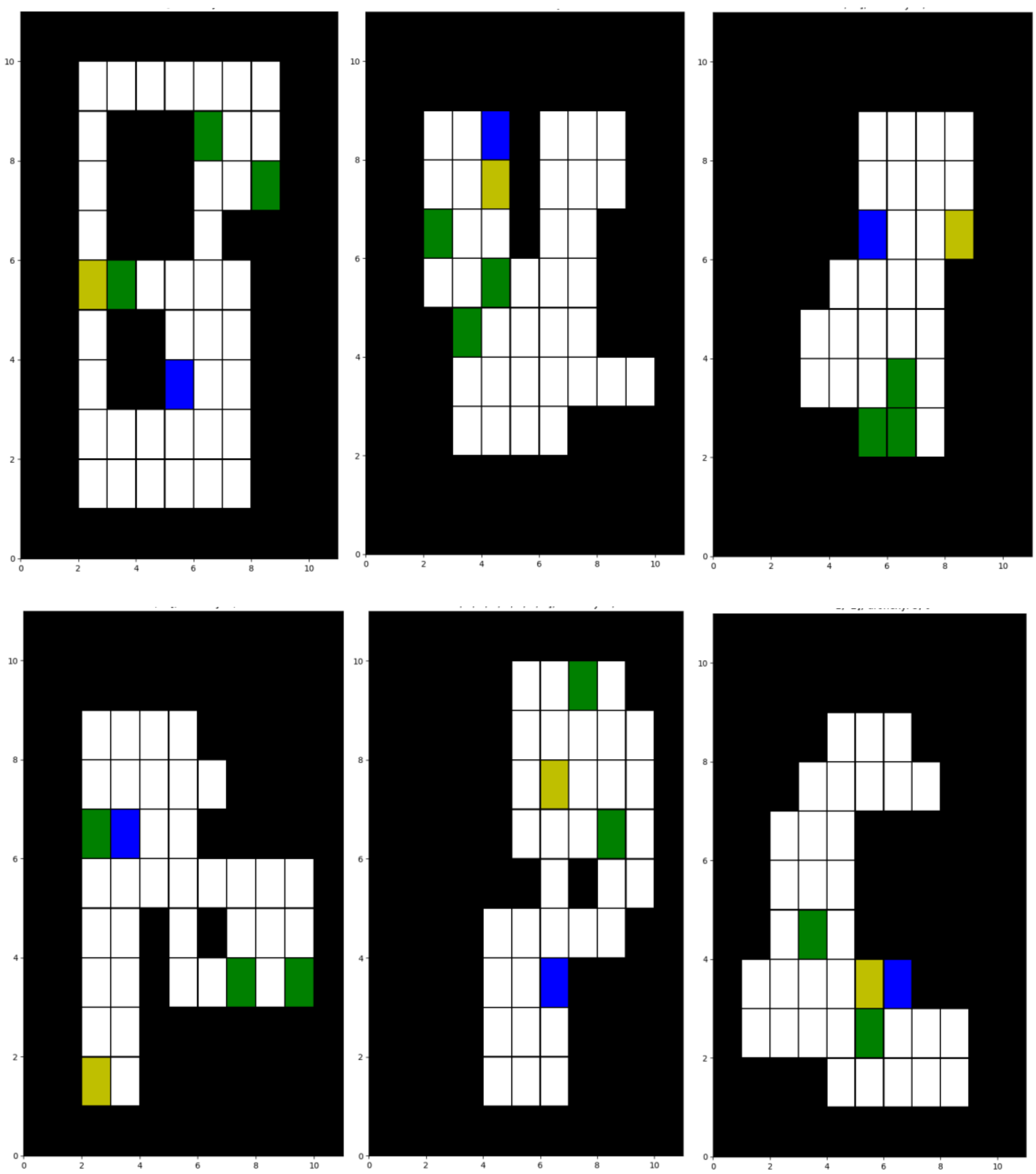


Figure 25. Visualisation of example 11x11 grid world maps, which varied in the location of walls, columns and free cells to move around the map. Colour code - Black: *walls*; White: *empty cells*; Yellow: *target*; Green: *distractor objects*; Blue: *current actor's position*.

Actors included in these experiments had partial observability over the target position in the environment, i.e. they could see it only when it was in their 5x5 field of view, but they were fully informed of their position and that of the walls. To account for the resulting uncertainty and related information-gathering behaviours (Friston et al., 2015), the actors' trajectories were generated using the POMDP planner based on Montecarlo tree search presented in Ognibene et al. (2019). It also integrated a Bayesian filter that explicitly represented the actor's beliefs about the state of the task. The distractor objects were not represented in the beliefs as they would not affect the actor's behaviour. As the actor knew both its own location and the map, beliefs ultimately represented the probability distribution of the target location in the map, which was instead unknown to the actor. Note that while the beliefs design of the observed actor are necessarily determined by its task, this does not affect the generality of the observer's performance because it is blind to the task and belief design assumptions.

Observers

For the experiments in this study, I built observers who, in each episode, have access to a set of behavioural trajectory chunks of an actor comprising five past steps. The observer's goal is to make predictions about the observed actors' future behaviour, with a specific interest on the target position. I trained two types of observers to infer the actors' target position based on the actors' overt behaviour, according to the two architectures described above. Specifically, the first type of observer, i.e. the *NoBeliefs* observer, predicts an actor's target position, next action, and next resulting state; the *NoBeliefs* architecture was therefore used. In contrast, the *Beliefs* observer, was asked to predict (in addition to the previous) also the actor's beliefs; thus utilising the *Beliefs* architecture. Observers are referred to in this thesis as “simple observer” and “ToM observer”, respectively. To note, I trained the observers to infer observed actors’ future behaviour from an allocentric perspective. Specifically, I assumed that both the observed behaviour and corresponding beliefs are produced by the actor itself and utilised as a teaching signal using an allocentric representation. As previously mentioned, to investigate the developmental trajectory of learning to predict other actors’ behaviours and beliefs, the observers were exposed to differing numbers of training samples.

Input sensing and routing

The observers can observe themselves or others; the input vector for the system can then be provided by a common reference frame to represent either the proprioception or self-localisation state of the observer or the physical state of the other actor. Several architectures have studied the problem of how to switch between the processing of own and others’ data and how to acquire others’ physical states. Note that in this case this function, while important, poses less constraints to behaviour performance as it feeds not the execution process but the learning one.

Input encoding

Pre-processing. The shared *prediction net* takes inputs formed by a number (a total of 5) of past steps of a full trajectory on a single grid map, including information on the actor trajectory, presence of walls and objects’ locations. Observed actions\states pairs are combined through a spatialisation-concatenation operation, whereby actions are tiled over space into a tensor and concatenated to form a single tensor of shape (11 x 11 x 20). While 11 x 11 represents the size of the grid world environments, 20 vectors are provided as inputs consisting of information regarding (a) actions (9 possible actions in experiments, thus 9 vectors); (b) objects coordinates, including the target position (4 objects, thus 4 vectors); (c) actor's position in past steps (5 past steps, thus 5 vectors); (d) 1 feature plane for the walls in environments; and (e) 1 vector for the actor's current position.

Prediction Net. Following spatialisation-concatenation operation, tensors are passed through a 2-layer ResNet with 32 channels, leaky ReLU nonlinearities, and batch-norm.

Target Location Prediction Head. The output from the torso is inputted into a 1-layer Convnet with 32 channels and leaky RELU, another 1-layer Convnet with 16 channels and leaky RELU, followed separately by (a) a fully connected layer to 121-dim logits (11 x 11 grid world) and (b) another 1-layer Convnet with 4 channels to 1. These are then summed.

Next Action Prediction Head. The output from the torso is inputted into a 1-layer Convnet with 32 channels and leaky RELU, followed by average pooling, and 2 fully connected layers to 9-dimensions (9 possible actions).

Next State Prediction Head. The output from the torso is inputted into a 1-layer Convnet with 32 channels and leaky RELU, another 1-layer Convnet with 16 channels and leaky RELU, followed separately by (a) a fully connected layer to 121-dim logits (11 x 11 grid world) and (b) another 1-layer Convnet with 4 channels to 1. These are then summed.

Beliefs Prediction Head. The output from the torso is inputted into a 1-layer Convnet with 32 channels and leaky RELU, another 1-layer Convnet with 16 channels and leaky RELU, followed separately by (a) a fully connected layer to 121-dim logits

(11 x 11 grid world) and (b) another 1-layer Convnet with 4 channels to 1. These are then summed.

Training

All architectures (both *Beliefs* and *NoBeliefs*) were trained with the Adam optimiser, with varying learning rates (we tested 6 levels from 0.00015 to 0.001), using batches of size 32. A learning rate scheduler with the following parameters was used in all experiments: milestones = [30, 60, 80, 160], gamma = 0.5. The Cross Entropy loss function was utilised for all *heads* training, except for the Belief *head*, for which the Kullback–Leibler divergence loss function was used instead. This choice was driven by the fact that the action, state and target prediction *heads* all required one-hot encoding to identify one single position in the map, whereas a distribution of probabilities over each map location was needed for the belief *prediction head*. Prior to conducting the main study described in this thesis, L1 and L2 regularisation tuning was performed using the L1 and L2 factor search; the final values of the L1 and L2 parameters used in all experiments were 0.005 and 0.001, respectively. In addition, early stopping was also integrated during training as a means to prevent overfitting. Early stopping was performed to obtain the best total model and best total loss and accuracy models separately, as well as best loss and accuracy models for each *prediction head* (i.e. target position, state, action, belief). A validation set comprising 10 maps (with 30 associated behaviours per map) was utilised for early stopping. Furthermore, balance between *prediction heads* was achieved through factor tuning. Specifically, normalisation factors for each *prediction head* were calculated from the losses obtained on the best total model from each *prediction head*, utilising the following formula: (1/best loss model for *prediction head*)*100[1]. The resulting normalisation factors which were utilised in all experiments were: 0.000718612, 0.000702, 0.114846, and 0.00219 for the target, action, state, and belief (when present) *prediction heads*, respectively. Finally, I trained the nets with a varying number of samples.

3.1 "Like Them": Developmental synergy between behaviour prediction and explicit representations of others' beliefs in a deep-learning model of Theory of Mind

In this study, I conducted a series of experiments to assess the role of learning explicit belief representations towards understanding others' intentions and behaviours throughout development. Specifically, the ability (or inability) to process beliefs was implemented by building two neural architectures differing in whether (or not) they included a Belief *head* (representation and prediction of beliefs), i.e. the *Beliefs* and *NoBeliefs* architectures. These correspond to the "ToM observer" and "simple observers", respectively. The developmental point of view (thus the study of the learning trajectory) was instead implemented by providing the neural networks with differing numbers of environmental settings (or maps) for observed behaviours during training.

In all experiments, varying learning rates were included (6 levels from 0.00015 to 0.001) and multiple subjects testing was conducted (achieved by initialising the

[1] Please note that for each *prediction head*, normalisation factors were calculated based on the best loss models measured with respect to *each specific head* separately.

network with 18 different weights, that is equivalent to 18 subjects) for both architectures and for any given condition.

Our results are summarised in Table 1, which shows the best target prediction accuracy (averaged between runs, or subjects, with the same belief processing ability) and the associated variance for both the *Beliefs* and *NoBeliefs* architectures. Furthermore, Table 20 indicates the learning rates which resulted in the best performances for both architectures based on the number of maps that were made available during training. See also Figure 26 for a visualisation of the target prediction accuracies by number of maps in the *Beliefs* and *NoBeliefs* architectures.

Table 20*.* Best target prediction accuracies by best learning rate obtained for the *Beliefs* vs *NoBeliefs* architectures based on the number of maps made available during training. Accuracies were calculated as averages across 18 initial network weights; the associated variances were reported.

	BEL			NoBEL			BEL-NoBEL (%)	*p-value*
Train Maps (N)	Best LR	Avg Acc (%)	Var	Best LR	Avg Acc (%)	Var		
5	0.001	59.26	18.84	0.001	61.51	20.30	-2.25	***.043***
10	0.001	65.75	8.99	0.001	65.46	11.82	0.29	.488
15	0.001	67.54	2.79	0.001	66.85	1.84	0.69	.077
20	0.001	68.74	2.08	0.001	67.47	1.51	1.27	***.001***
25	0.001	69.45	1.63	0.001	67.57	1.44	1.87	***<.001***
30	0.001	69.49	1.60	0.001	67.79	1.76	1.69	***<.001***
60	0.00075	70.49	1.23	0.001	69.32	1.33	1.17	***<.001***
120	0.00075	71.47	0.87	0.0005	70.46	0.89	1.00	***<.001***
300	0.00015	72.43	0.28	0.00015	71.59	0.43	0.84	***<.001***

BEL: *Beliefs* architecture; NoBEL = *NoBeliefs* architecture

Both within- and between-architectures, as well as developmental analyses were performed to interpret data. This is valid also in all subsequent experiments.

Within-architectures analysis

As shown in Table 20, I observed a steady trend in both the architectures for achieving better performance with an increasing number of maps made available during training (see also Figure 26). These results suggest that better target prediction accuracies can be achieved with increasing experience. Furthermore, the above results suggest a trend in both architectures with regards to the learning rate required to achieve best performance. Specifically, a lower learning rate was more performant in both architectures with an increasing number of maps. This is in conformance with the fact that training with an increasing number of maps results in longer learning and more learning updates, which in turn call for a low learning rate to achieve best performance (in this case, a learning rate of 0.00015). The opposite is valid for training with a small number of maps (in this case, a learning rate of 0.001). Finally, high variance was observed in both architectures trained with small numbers of maps (5 and 10 maps), showing overfit with a limited number of samples (a total of 150 and 300 behaviours, respectively).

Between-architectures analysis

With regards to the role of beliefs for intention prediction, results in Table 20 indicate better performance (highly statistically significant in most cases) by the *Beliefs* architecture (see Figure 26 for a visualisation of this data). This is valid for all except for the smallest number of maps included in this experiment (5 maps), which however was identified to have high variance and low statistical significance (likely due to insufficient training data and resulting overfit). Performance improved by up to 1.87% (p < .001) (25 maps) when including the beliefs *head*, thus when allowing belief processing. Overall, these results indicate an important role for beliefs in target prediction accuracy, suggesting that adding the Belief *head* results in regularisation and, as a consequence, improved performance.

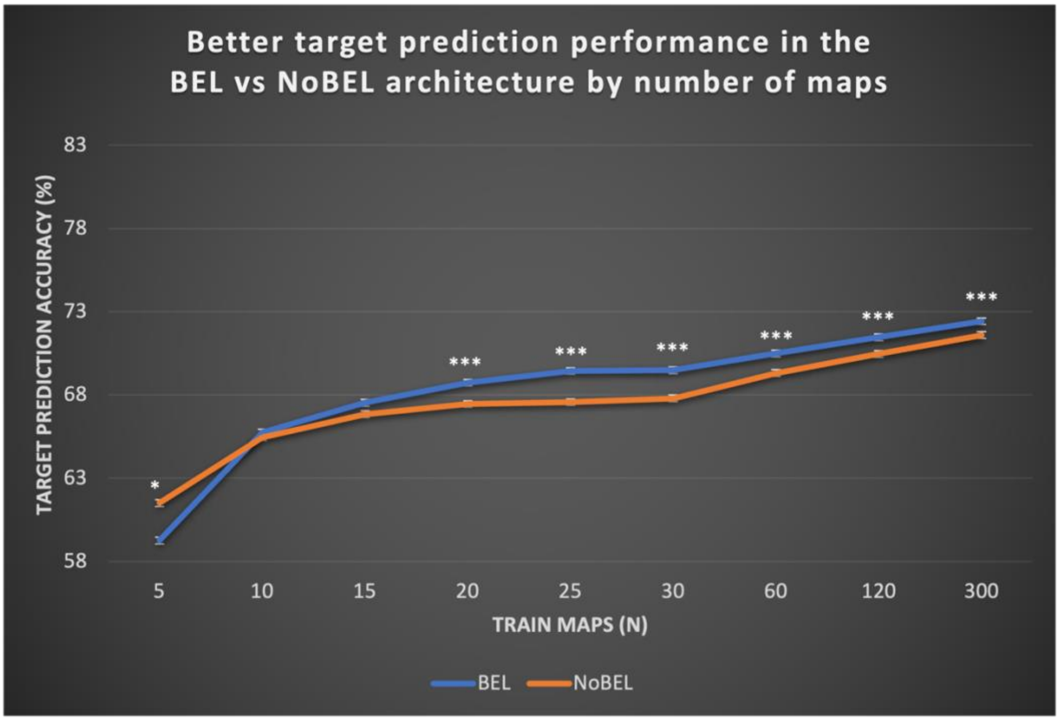


Figure 26. Visualisation of increase in performance by the *Beliefs* vs *NoBeliefs* architectures with an increasing number of training maps. Average target prediction accuracy at each number of map; ***: significant at the .001 level (2-tailed); *: significant at the .05 level (2-tailed).

Further supporting this hypothesis are the results shown in Table 21 below. Specifically, the above target prediction accuracies were obtained through early stopping on the best target accuracy and best loss on actor's target model in both architectures (as I was interested in determining best performance with regards to target prediction). However, the data presented in Table 21 report the target prediction performance obtained on the best loss on actor's beliefs model, meaning the model which achieved best prediction of beliefs. These results suggest a comparable performance in target prediction accuracy obtained in the best target and beliefs models. Indeed, results follow a similar trajectory, both in terms of overall and experience-based performances, confirming multi-task-induced regularisation. This can therefore be considered the reason behind improved performance by the *Beliefs* architecture. In other words, these results indicate a useful exchange of information between the Belief and Target *prediction heads*, which evidently represents an advantage towards predicting others' behaviour.

Table 21. Target prediction performance of the *Beliefs* architecture obtained on the best loss on agent's beliefs model, supporting multi-task-induced regularisation. Accuracies were calculated as averages across 18 initial network weights; the associated variances were reported.

	BEL		
Train Maps (N)	Best LR	Avg Acc (%)	Var
5	0.001	58.79	18.65
10	0.001	65.83	7.60
15	0.00075	67.06	5.86
20	0.001	68.01	1.97
25	0.001	68.81	1.50
30	0.001	69.54	1.23
60	0.00075	70.43	1.30
120	0.00075	71.33	1.26
300	0.00015	72.41	0.45

BEL: *Beliefs* architecture; NoBEL: *NoBeliefs* architecture; LR: learning rate

Developmental analysis

From a developmental perspective, my results suggest an interplay between amount of experience and target prediction, with increasing experience resulting in improved target prediction accuracy.

With regards to the impact of learning explicit beliefs representation on target prediction accuracy, I instead identified a developmental trend. Specifically, I observed the improvement in performance when including the Belief *head* in my architecture to follow an increasing trend from 5 to 25 maps, to then convert into a decreasing trend from 25 to 300 maps. This supports the regularisation hypothesis: there is a significant contribution of the Belief *head* towards learning target prediction, which increases until reaching a plateau (e.g. 25 maps), after which, with enough data (large number of maps, e.g. 300 maps), regularisation is decreasingly needed. In contrast, regularisation is not yet effective with very few training samples due to overfit, as also highlighted from the high variance in the results. Crucially, in this simplified environment (only a few objects, localised without noise, in a simple grid map of 11x11 cells), the performances of both models kept increasing, even if slightly, with quite an extensive training set (i.e. 300 maps, with 30 behaviours). Therefore, in realistic environments that are much more complex and diverse than grid worlds, it is likely that the performance gain due to learning to predict others' beliefs together with the target may extend for a significant amount of time. Future studies are warranted to assess whether this is the case. Nonetheless, I investigated the generalisation of my architectural choices in a more complex environment in Part 3, chapter 3.3 of this thesis.

From a human behaviour point of view, these results indicate that beliefs processing starts playing an increasingly important role for understanding others' behaviours with increasing experience, until reaching a possible plateau of maximum impact after which beliefs again gradually become less useful for predicting others' behaviour, although still representing a good source of information (see Figure 26). This result could be interpreted as follows. While the correct recognition of an actor's target requires increasing experience, beliefs processing may be an ability that develops early in humans and can be used to aid the interpretation of others' behaviours from an early age, prior to extensive experience. Overall, these findings may indicate a beneficial role for beliefs in predicting others' intentions throughout development from an early age, supporting early ToM emergence[2].

[2] The actual dynamics may depend on the complexity of the environment (see Part 3, chapter 3.3 of this thesis).

Conclusions

- Learning explicit representations of others' beliefs, taking a 'like them' approach, is beneficial for predicting others' behaviours and results in improved accuracy of others' intentions prediction.
- This effect is driven by multi-task-induced regularisation between beliefs and target processing, which results in faster learning with a lower number of training samples (i.e. more efficient learning).
- This beneficial effect follows a developmental trend, i.e. increasing impact seen with increasing experience. A plateau of maximum impact is reached, after which this decreases, while remaining important for improved performance.
- With respect to human behaviour, these results indicate that not only beliefs processing is computationally possible from an early age, but that it also beneficial towards predicting others' behaviours. Overall, these findings support early Tom emergence and indicate it to be advantageous for understanding others.

3.2 Conditions which maximise multi-task-induced regularisation between target and beliefs processing

In the next studies, I explored the specific conditions in which including in the architecture the learning of explicit beliefs representations has the most beneficial effects towards predicting others' intentions.

Exp. 1: Target Visibility

First, I conducted further testing on both architectures with simulations which varied according to whether the target was (or not) *visible* by the actor for the whole behavioural chunk processed by the "*neural network observers*". In all cases, all distractor objects present in the environment were generated and randomly positioned within the actor's field of view (FOV)[3].

This current manipulation resulted in the following conditions: (1) Target *visible* + 3 Objects *visible*, and (2) Target *not visible* + 3 Objects *visible*. All trained neural networks (i.e. from both configurations and trained with different numbers of maps and initial weights) were tested for these new conditions. Results are summarised in Tables 22A,B.

[3] The impact of also changing the number of steps done before or after seeing the target and the visibility of objects are investigated in the next studies (Exp. 2-6 of this chapter).

Tables 22. Best target prediction accuracies for the *Beliefs* vs *NoBeliefs* architectures by number of training maps in the conditions with 3 distractor objects visible and either target ***(A)*** *visible* and ***(B)*** *not visible* by the actor. Accuracies were calculated as averages across 18 initial network weights; the associated variances were reported.

A.

	Target *visible* – 3 Objects *visible*					
	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	79.33	48.59	81.17	2.97	-1.83	.286
10	85.50	0.97	84.89	0.58	0.61	***.045***
15	86.50	0.50	85.83	0.38	0.67	***.005***
20	87.11	0.22	86.56	0.26	0.56	***.002***
25	87.17	0.26	86.44	0.26	0.72	***<.001***
30	87.33	0.24	86.61	0.25	0.72	***<.001***
60	87.94	0.06	87.72	0.21	0.22	.080
120	88.28	0.21	88.00	0.00	0.28	***.020***
300	88.94	0.06	88.28	0.21	0.67	***<.001***
				avg.	0.29	
				max.	0.72	

B.

	Target *not visible* – 3 Objects *visible*					
	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	37.44	96.97	38.56	31.44	-1.11	.681
10	52.61	29.31	48.83	17.44	3.78	***.025***
15	58.17	27.09	51.11	16.69	7.06	***<.001***
20	56.94	8.29	54.17	12.85	2.78	***.015***
25	57.11	8.58	51.56	13.91	5.56	***<.001***
30	56.22	8.18	53.22	6.42	3.00	***.002***
60	59.39	19.08	54.44	12.73	4.94	***.001***
120	59.39	13.08	55.11	10.46	4.28	***.001***
300	61.61	5.19	63.11	4.34	-1.50	***.047***
				avg.	3.20	
				max.	7.06	

BEL: *Beliefs* architecture; NoBEL: *NoBeliefs* architecture

Within-architectures analysis

In this experiment, I observed a general increase in performance for both architectures when the target was *visible* by the actor compared to the original study 1 (Part 3, chapter 3.1) (~17% increase at 300 maps). In contrast, I found a decrease in performance when the target was *not visible* (~10% decrease at 300 maps) (see Tables 22A,B vs Table 20). These fluctuations in performance can be expected, considering that study 1 included a mix of cases with target *visible* and *not visible*, thus resulting in an average performance between the two conditions analysed in this experiment[4].

Overall, it is clear that behaviour prediction improves when the target is *visible* by the observed actor. These results may be explained by the fact that when the target is *visible*, the actor may be engaging in a "*goal-oriented or directed"* behaviour (i.e. covering the shortest path between positions), thus in an efficient way (Csibra & Gergely, 2007). In contrast, in the absence of a known target position, the actor's navigation behaviour would rather be "*information gathering*" (e.g. target-seeking or exploration of a wider area, often not adopting the shortest path between visited positions) (Morash, 2016). Indeed, even an optimal explorative actor may perform u-turns, zigzag and spiralling movements to localize a target with an unknown position. As a result, an observer may be able to use this information to predict the actor's behaviour. Specifically, it may be easier for an observer to predict the actor's target when his behaviour is recognised to be *goal-directed*, as opposed to an *exploratory* behaviour, leading to improved performance. Interestingly, the neural networks

[4] Indeed, in the testing set of study 1 (Part 3, chapter 3.1 of this thesis), a steps *visible* / steps *not visible* ratio of 1.35 : 1 was observed. Please note that the distribution of these conditions in the training set of study 1 is not known, thus the real effect of the gain in performance may be bigger than expected.

seemed to be able to recognise a difference between these two behaviours, resulting in different performances when the target was (or not) *visible* by the actor.

Between-architectures analysis

By manipulating whether the target was *visible* by the actor (Tables 22A,B), I identified a great advantage of including belief processing for intention prediction when the target was *not visible* and all distractor objects were *visible* (see also Figure 27). Indeed, a statistically significant big positive impact of including the Belief *head* in my architecture was observed in that condition, reaching up to 7.06% difference in target prediction accuracy (p < .001) compared to the *NoBeliefs* architecture (15 maps).

This result may be driven by the fact that beliefs collapse as redundant information on target position when the target is *visible* by the actor. Specifically, the actor's *goal-directed* behaviour may provide enough information to an observer with regards to the actor's target. Overall, a beneficial role for learning to explicitly represent beliefs, specifically towards *exploratory* behaviour, was here identified.

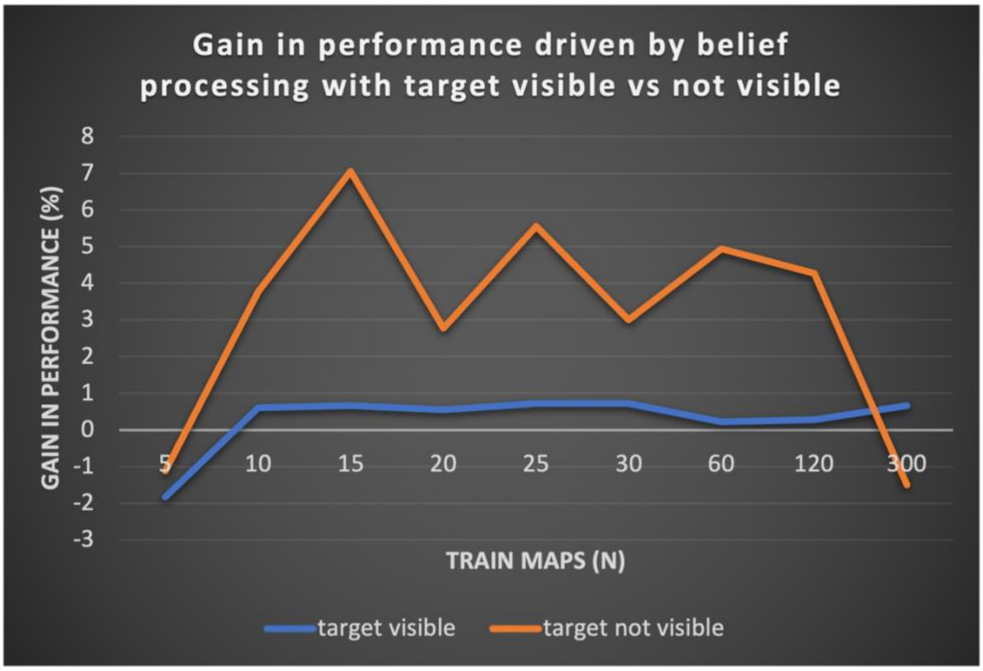


Figure 27. Visualisation of gain in performance driven by the Beliefs *head*, by number of training maps, in the experimental condition with 3 distractor objects *visible* and target *visible* vs *not visible* by the observed actor.

Developmental analysis

From a developmental perspective, results from this experiment indicate that the recognition and prediction of goal-directed, as opposed to exploratory, behaviours may be skills that develop earlier during development. Indeed, while target prediction accuracy improves with experience when observing an actor engaging in both behaviours, performance is always worse when interpreting *exploratory* behaviours.

Nonetheless, these results also suggest that being able to process beliefs may be mostly informative when observing actors engaging in *exploratory* behaviours, which significantly accelerate the learning process. In *exploratory* conditions, actors indeed need to rely more strongly on their beliefs; thus, if an observer is aware of them (i.e. is endowed with a ToM), a substantial gain in performance may be obtained.

Finally, with respect to the impact of explicitly learning beliefs, I also observed the same (although here less linear) developmental trend identified in study 1, with the highest impact of beliefs processing at a medium level of experience. These results thus support the previously identified dynamics of multi-task-induced regularisation. From a human developmental perspective, these findings support early development of beliefs representation and indicate their beneficial role towards interpreting behaviours of actors with partial knowledge about the environment and engaging in exploratory behaviours (especially when the observer has had some experience with similar situations).

Exp. 2: Steps-driven performance and Beliefs Processing

Next, I further assessing whether predictive performance was influenced by the number of steps performed by the observed actor and associated beliefs, both in the target *visible* and *not visible* conditions.

Specifically, I considered, for each given condition, a varying number of steps (from 1 to 5) in the agent's simulated trajectories with target *visible* or *not visible*, resulting in the following conditions: Target *visible* and 5, 4, 3, 2, or 1 steps *visible;* Target *not visible* and 5, 4, 3, 2, or 1 steps *not visible*. This experiment was conducted for two sets of maps only, i.e. 25 and 120 maps. This choice was motivated by the fact that these were identified as the required number of maps for *Beliefs* to be most influential (to peak) and to decrease consistently from the peak, respectively, in study 1. Therefore, these were considered to well represent the window of experience during which the main changes in the previously identified (developmental) trend occurred. Results are shown in Tables 23A,B.

Tables 23. Best target prediction accuracies for the *Beliefs* vs *NoBeliefs* architectures based on target visibility and number of past steps with target visibility, when considering ***(A)*** 25 train maps and ***(B)*** 120 train maps. Best target prediction accuracies from an adaptive architecture are also presented in the condition with 120 train maps ***(B)***. Accuracies were calculated as averages across 18 initial network weights; the associated variances were reported.

A.

	BEL		NoBEL		BEL-NoBEL	*p-value*
Target *visible*	Avg Acc (%)	Var	Avg Acc (%)	Var		
5 past steps with target *visible*	89.83	4.38	89.67	10.71	0.17	.857
4 past steps with target *visible*	90.50	1.72	90.94	5.00	-0.44	.509
3 past steps with target *visible*	91.89	2.10	92.22	3.36	-0.33	.549
2 past steps with target *visible*	97.78	0.18	97.78	0.18	0.00	1.000
1 past step with target *visible*	96.56	0.26	96.56	0.61	0.00	1.000
Target *not visible*						
5 past steps with target *not visible*	55.67	2.12	52.22	3.48	3.44	***<.001***
4 past steps with target *not visible*	55.83	1.68	52.72	2.45	3.11	***<.001***
3 past steps with target *not visible*	53.44	0.85	51.06	3.11	2.39	***<.001***
2 past steps with target *not visible*	50.18	1.28	48.33	3.29	1.84	***.001***
1 past step with target *not visible*	46.22	1.71	44.39	4.37	1.83	***.003***

B.

	BEL		NoBEL		BEL-NoBEL	*p-value*	Adaptive	
Target *visible*	Avg Acc (%)	Var	Avg Acc (%)	Var			Avg Acc (%)	Var
5 past steps with target *visible*	91.61	2.84	92.44	2.61	-0.83	.139	92.56	1.91
4 past steps with target *visible*	92.61	2.02	92.94	0.88	-0.33	.412	92.50	1.79
3 past steps with target *visible*	93.89	1.52	94.33	0.82	-0.44	.226	94.06	0.88
2 past steps with target *visible*	98.11	0.10	98.00	0.00	0.11	.163	98.00	0.00
1 past step with target *visible*	97.78	0.18	97.78	0.18	0.00	1.000	97.33	0.24
Target *not visible*								
5 past steps with target *not visible*	58.33	1.53	56.89	1.16	1.44	***.001***	58.47	1.14
4 past steps with target *not visible*	57.89	1.40	56.28	1.27	1.61	***<.001***	57.78	1.59
3 past steps with target *not visible*	55.44	2.50	53.78	2.07	1.67	***.002***	55.11	2.69
2 past steps with target *not visible*	52.00	2.71	50.56	2.61	1.44	***.012***	51.89	3.40
1 past step with target *not visible*	47.78	3.01	47.06	3.11	0.72	.224	47.67	4.59

BEL: *Beliefs* architecture; NoBEL: *NoBeliefs* architecture

Within-architectures analysis

Tables 23A,B show a striking difference in performance between the target *visible* vs *not visible* conditions (see also Figure 28 for an example visualisation of this data), in both architectures and regardless of the number of maps, supporting findings from Exp.1 ("*Target visibility*"). Specifically, better performance is achieved in both the

Beliefs and *NoBeliefs* architectures in the conditions with the target *visible*, as opposed to *not visible*.

More in detail, when the target is *not visible* by the actor (thus during exploratory behaviours), target prediction performance increases with an increasing number of steps with target *not visible* by the actor (see Figure 28). This is valid for both architectures and regardless of the number of maps. These results therefore suggest that the more steps the actor performs while the target is *not visible* (thus showing exploratory behaviour), the easier it is for an observer to predict his behaviour. This is probably due to additional information made available to the observer during such exploratory behaviour. For example, distractor objects neglected by the actor in additional steps can be discarded as non-targets. In contrast, with fewer steps, only a limited area can be covered by the actor's FOV and the same probability of being the target is assigned to the objects outside of his FOV. Note that the current architecture has a limit of 5 steps in memory. Actor's object neglect as a source of information during goal-directed and exploratory behaviours is further assessed in the next experiments.

When the target is *visible* (thus during goal-directed behaviours), target prediction performance decreases with an increasing number of steps with the target *visible* by the actor instead. However, this is not a linear relationship, as an increase from 1 to 2 steps with target *visible* is seen (see Figure 28). This is valid for both architectures and regardless of the number of maps. These results indicate that it is easier for an observer to predict the behaviour of an actor who has seen his target for a couple of steps (e.g. 1 vs 2 steps). However, if the actor completes several steps with a *visible* target, it may become challenging for an observer to predict his target, considering that the actor should take the most efficient path to the target once *visible* (thus showing a goal-directed behaviour)[5]. This is something that remains consistent with increasing experience.

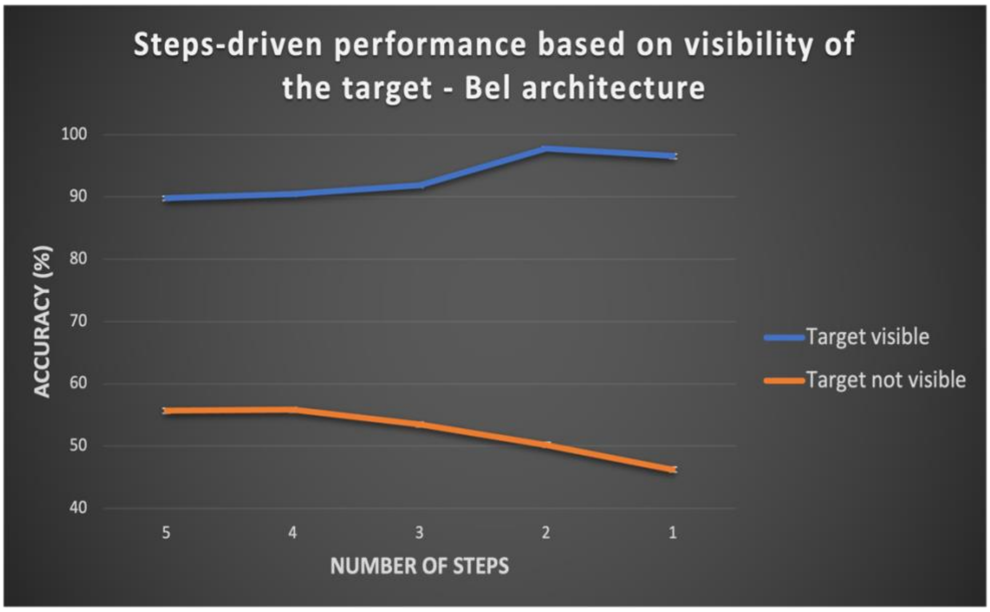


Figure 28. Example visualisation of the steps-driven performance in the *Beliefs* architecture based on actor's target visibility, by number of steps completed by the actor given the target visibility condition. Average target prediction accuracy at each number of steps.

Between-architectures analysis

Tables 23A,B indicate that the role of beliefs for behaviour prediction becomes once again more evident in conditions with target *not visible* (thus during exploratory behaviours) where the *Beliefs* architecture significantly outperforms the *NoBeliefs* architecture (maximum gain in performance: 3.4%, p < .001, 5 past steps with target

[5] Based on a follow-up analysis of steps statistics (see supplementary material 1), this condition happens very rarely in my simulated trajectories, as the actor generally takes a direct path to the target once *visible*. Therefore, this confirms that conditions with an additional number of steps with target *visible* may confuse the observer and render behaviour prediction more challenging.

not visible, 25 train maps) (see also Figure 29). Furthermore, this seems to follow a trend: the gain in performance driven by beliefs processing seems to increase with an increasing number of steps when the target is *not visible*, regardless of the number of maps.

In contrast, when the target is *visible* by the actor (thus during goal-directed behaviour), target prediction accuracies reached with the *Beliefs* architecture are the same or worse than those obtained with the *NoBeliefs* architecture (maximum decrease in performance due to beliefs processing: ~-0.8%, p = .139, 5 past steps with target *visible*, 120 maps). However, none of these results are statistically significant; therefore, further studies are warranted to elucidate the role of beliefs processing for step-driven target prediction performance when the target is *visible*. Nonetheless, a possible explanation for these results is the following. When the target is *visible* by the actor, beliefs are substantially redundant with the target position output (and related training signal) and their contribution to multi-task-induced regularisation may thus be limited. Therefore, an absence of substantial difference in performance between the two architectures when the target is *visible* is expected.

Overall, these findings suggest that there is an increasing role for beliefs processing (towards predicting others' behaviour) when the observed actor has not seen the target and has interacted with the environment for a longer period of time. This increasing role of beliefs processing with time may also be a result of the low informativity of beliefs at the beginning of the trial, when their information is redundant with the current actor's position but still informative with respect to the actor's FOV. In other words, beliefs are mostly beneficial in exploratory behaviours, especially following longer exploration.

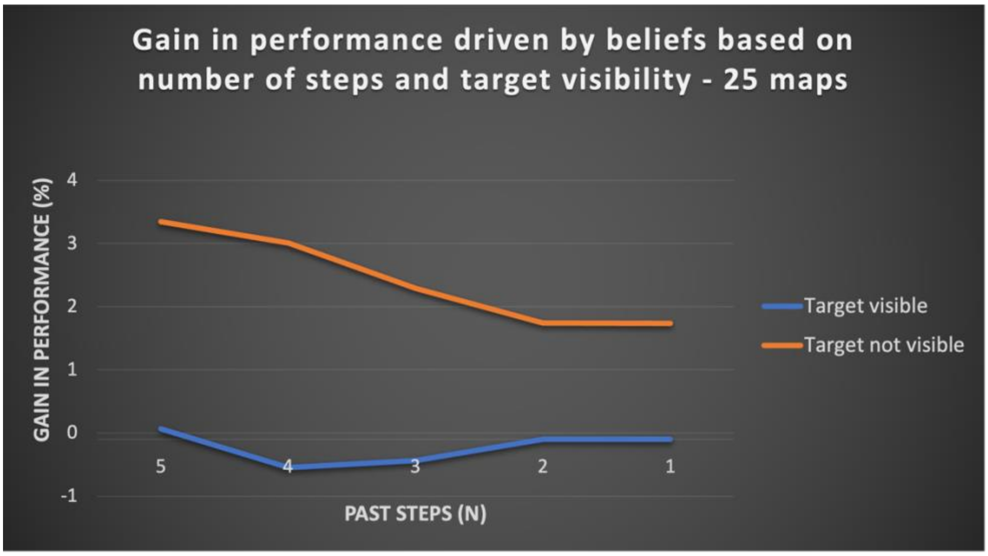


Figure 29. Visualisation of gain in performance driven by beliefs processing based on target visibility and number of past steps with the given target visibility, when considering 25 train maps.

Developmental analysis

A developmental trend can be observed in both performance and beliefs impact (when the target is *not visible*) on behaviour prediction, with the former improving with increasing experience and the latter decreasing at extensive experience (see Tables 23A vs 23B). These results resonate with those from my previous studies in that, while experience helps with target prediction, learning beliefs and the regularisation effect it introduces in this architecture becomes less informative at extensive experience.

In order to further assess learning through multi-task-induced regularisation, I compared the gain in performance between the above conditions and an “adaptive” network (120 train maps) (see Table 23B). In more detail, the “adaptive” architecture would activate (or not) the Belief *head* based on the characteristics of the environment identified in the steps-driven performance experiments. The Belief *head* would be activated and used to learn and predict target accuracy only when the target was *not visible* by the agent or *visible* from less than 3 steps. The results from the adaptive network (see column “adaptive” in Table 23B) are similar to those obtained from the architecture with the sole beliefs processing (*Beliefs* architecture), suggesting an absence of additional gain in performance with an adaptive, hybrid architecture. This may be due to (a) having selected a developmental phase for which the belief contribution was already strongly reduced, and (b) different distribution of the training and test datasets with regards to the number of samples including *visible* and *not visible* target. Nonetheless, this result, together with the steps-driven performance results above, suggests that the actual gain in performance seen for the *Beliefs* vs *NoBeliefs* architectures is driven by the learning and processing of beliefs *when they are not redundant*.

Exp. 3: Visual crowding when target is not visible and Beliefs processing

Next, I further explored the impact of beliefs on performance in the target *not visible* condition when a varying number of distractor objects (3, 2, 1, or 0) were *visible* by the actor. I call this condition “visual crowding”. The rationale for conducting this experiment was the following.

Having a high number of distractor objects *visible* by the actor (visual crowding) when the target is *not visible* may on the one hand result in (H1) confusion in the target prediction process, possibly due to an acquired “target-close-to-actor” bias. On the other hand, it may result in (H2) improved target prediction performance if relying on the actor’s behaviour who is seen neglecting such distractors. This latter hypothesis is driven by the fact that, even when considering the optimal target recognition process, only objects *visible* by the actor can be discarded by the observer and considered non-target on the basis of the actor's behaviour. In contrast, distractor objects that are *not visible* by the actor would not be discarded and would continue to confuse the observer.

In line with this, as seen in the previous experiments, the actor's navigation behaviour changes based on whether the target is *visible* or *not visible* by the actor. When the target is *not visible* by the actor, his exploratory behaviour may be a source of information that, if recognised and distinguished from goal-directed behaviour, can be used by the observer to discard objects currently *visible* by the actor as non-targets.

Given that, in the previous experiments, the impact of belief learning on target prediction was found to be stronger in the target *not visible* condition, thus pointing to a higher efficiency with exploratory behaviours, I expect the second hypothesis (H2) to be stronger in belief learning architectures (i.e. the *Beliefs* architecture).

To assess this, all trained neural networks (i.e. from both configurations and trained with different numbers of maps and initial weights) were tested for the following new conditions: (1) Target *not visible* + 3 Objects *visible*, (2) Target *not visible* + 2 Objects *visible*, (3) Target *not visible* + 1 Object *visible*, (4) Target *not visible* + No Objects *visible*. *Visible* and *not visible* distractor objects were generated and randomly located within and out of the actor's FOV, respectively. Results are summarised in Tables 24A-D.

Tables 24. Best target prediction accuracies for the *Beliefs* vs *NoBeliefs* architectures by number of training maps in the conditions with the target not visible by the actor and ***(A)*** 3, ***(B)*** 2, ***(C)*** 1, and ***(D)*** No distractor objects *visible.* Accuracies were calculated as averages across 18 initial network weights; the associated variances were reported.

A.

	Target *not visible* - 3 Objects *visible*					
	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	37.44	96.97	38.56	31.44	-1.11	.681
10	52.61	29.31	48.83	17.44	3.78	***.025***
15	58.17	27.09	51.11	16.69	7.06	***<.001***
20	56.94	8.29	54.17	12.85	2.78	***.015***
25	57.11	8.58	51.56	13.91	5.56	***<.001***
30	56.22	8.18	53.22	6.42	3.00	***.002***
60	59.39	19.08	54.44	12.73	4.94	***.001***
120	59.39	13.08	55.11	10.46	4.28	***.001***
300	61.61	5.19	63.11	4.34	-1.50	***.047***
				avg.	3.20	
				max.	7.06	

B.

	Target *not visible* – 2 Objects *visible*					
	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	32.22	50.65	34.33	14.59	-2.11	.278
10	41.83	9.91	40.67	9.88	1.17	.274
15	45.28	10.21	42.33	10.35	2.94	***.009***
20	44.89	5.99	44.61	10.37	0.28	.773
25	45.11	4.22	42.22	7.01	2.89	***.001***
30	44.11	3.87	43.56	4.26	0.56	.414
60	46.22	8.77	44.50	7.68	1.72	.080
120	47.33	5.18	44.44	6.14	2.89	***.001***
300	47.89	3.75	47.67	2.47	0.22	.708
				avg.	1.17	
				max.	2.94	

C.

	Target *not visible* - 1 Object *visible*					
	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	29.28	32.09	31.75	5.00	-2.47	.112
10	36.44	5.08	36.39	6.02	0.06	.944
15	39.22	5.01	38.00	4.94	1.22	.109
20	39.50	4.97	39.56	6.03	-0.06	.944
25	39.33	3.06	38.39	3.78	0.94	.135
30	38.72	2.57	39.17	2.74	-0.44	.418
60	40.11	4.93	40.11	4.69	0.00	1.000
120	41.11	2.46	39.67	3.65	1.44	***.018***
300	41.39	2.02	40.44	1.79	0.94	***.048***
				avg.	0.18	
				max.	1.44	

D.

	Target *not visible* - No Objects *visible*					
	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	27.33	22.59	31.11	11.52	-3.78	.102
10	33.22	2.18	34.11	4.22	-0.89	.146
15	35.72	3.27	35.56	2.73	0.17	.775
20	36.17	2.97	36.83	4.38	-0.67	.304
25	36.22	2.07	36.28	2.33	-0.06	.991
30	35.94	1.11	36.67	2.47	-0.72	.115
60	36.22	2.18	37.67	2.35	-1.44	***.007***
120	37.06	1.11	37.61	2.37	-0.56	.215
300	36.50	0.97	36.50	0.97	0.00	1.000
				avg.	-0.88	
				max.	0.17	

BEL: *Beliefs* architecture; NoBEL: *NoBeliefs* architecture

Within-architectures analysis

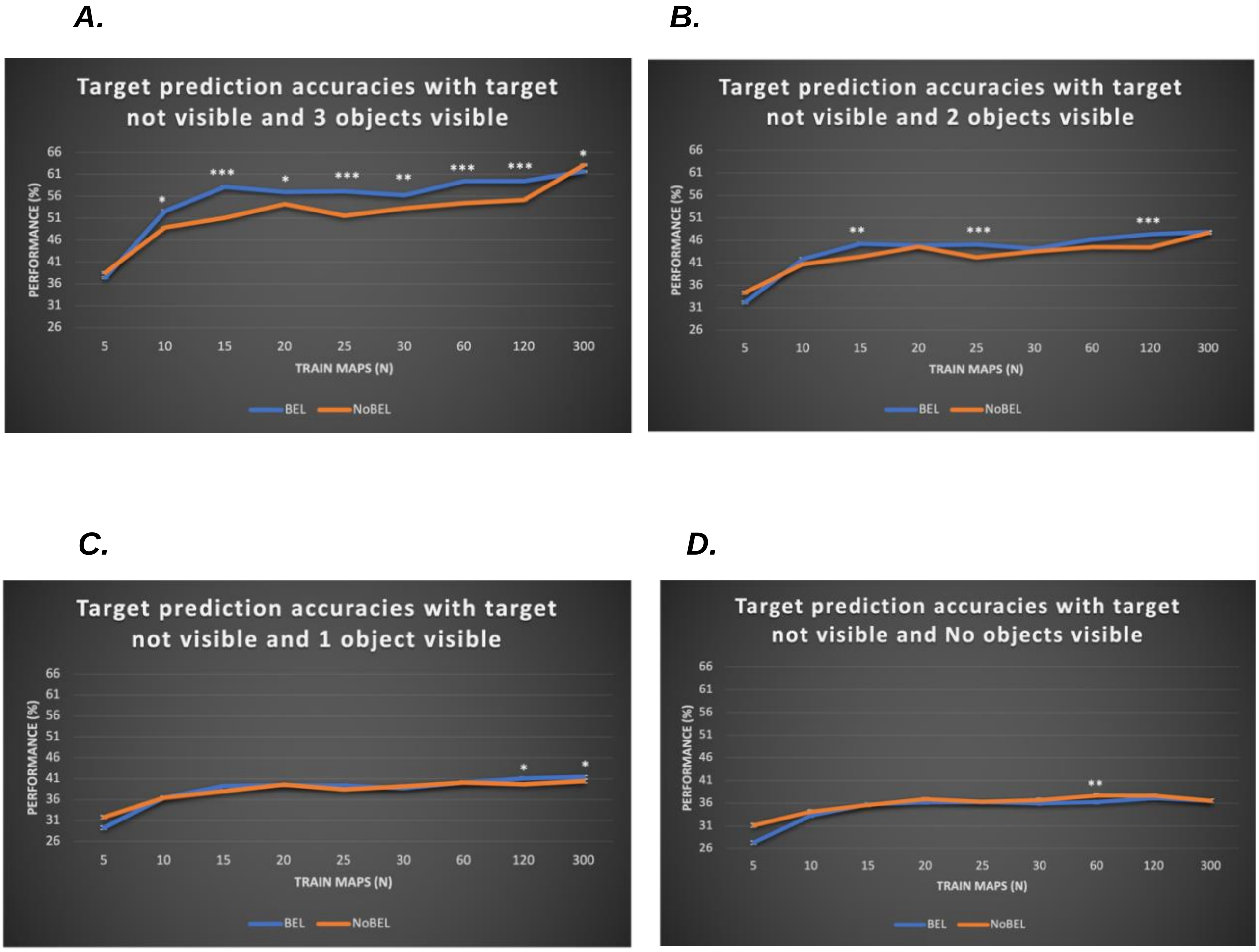
In Tables 24A-D, when the target is *not visible* by the actor, having an increasing number of objects in the actor's FOV (visual crowding) improves predictive performance, regardless of the architecture and the number of train maps or experience (see also Figures 30A-D for a visualisation of the effect of visual crowding on predictive performance when the target is *not visible*). In more detail, from the condition with No[6] to 3 objects *visible* by the actor and target *not visible*, performances increase by ~21% and ~15% in the *Beliefs* and *NoBeliefs* architectures, respectively (when considering 25 train maps).

This may be possibly driven by a relationship between the informativity of objects and actor's behaviours resulting in improved target prediction, supporting hypothesis (H2). Indeed, if objects were *visible* by the actor and one of them was his target, the actor would engage in a goal-directed behaviour. However, considering that in this experiment the actor has an exploratory behaviour (target *not visible* condition), an observer that recognises such behaviour can conclude that those distractor objects are non-target. Specifically, prediction of object neglect may be considered a very informative strategy when predicting the behaviour of an actor who does not hold all the information regarding an environment. The influence of object neglect strategy when predicting others' targets, as well as the relationship with beliefs processing and

[6] Condition (4) Target *not visible* + No Objects *visible* did not result in average 25% target prediction accuracy in neither of the architectures, as one would expect considering that the actor in this condition is not able to see any object nor the target. Indeed, the actor's behaviour should be considered random by the observer in this condition, and equal probabilities of being targets should be assigned to all objects. I reasoned that this may be a result of some of the distractor objects being in the past trajectory of the actor. This would mean that such objects were *visible* by the actor in previous steps and neglected, representing a valuable piece of information for the observer. I conducted a further experiment to validate this hypothesis. Specifically, I ensured that none of the objects would be in the actor's previous steps (thus *visible* in previous steps). As a result, equal probabilities of being targets (~25%) were assigned to all objects in this implementation (see supplementary material 2).

behaviour recognition, was further tested in Exp. 5 (“*Object Neglect and Beliefs Processing”*), where all objects were placed in the past trajectory of the actor.

Figures 30. Visualisation of the effect of visual crowding on predictive performance of the *Beliefs* and *NoBeliefs* architectures when the target is not visible by the actor. Specifically, lower target prediction accuracies are observed in both architectures from **(A)** 3 objects *visible* by the actor to **(B)** 2 objects *visible*, **(C)** 1 object *visible*, and **(D)** No objects *visible*, regardless of the number of maps (average target prediction accuracy at each number of train maps); *: significant at the .05 level (2-tailed); **: significant at the .01 level (2-tailed); ***: significant at the .001 level (2-tailed).



Between-architectures analysis

In Tables 24A-D, I continued to observe a role for beliefs in all conditions and number of maps (except for 5 maps). This seemed to increase (as well as statistical significance) with an increasing number of *visible* objects (i.e. higher effect during visual crowding). It instead reached similar or worse target prediction accuracies to the *NoBeliefs* architecture in the condition (4) Target *not visible* + No Objects *visible* (see Figure 32). However, statistical significance was observed to decrease with a decreasing number of objects, thus suggesting that the apparent negative effect of beliefs is not as relevant and that beliefs may at least not be harmful.

Overall, the role of beliefs for behaviour prediction when the target is *not visible* seems to be related to the recognition of the actor's exploratory behaviour, which in turn results in prediction of object neglect strategy mentioned earlier. Specifically, when the target is *not visible*, the actor's beliefs are spread over the parts of the environment that are yet to be seen by the actor. When distractor objects are *visible* and the actor does not engage in goal-directed behaviour, an observer can consider those objects as neglected by the actor, thus aiding target prediction. The variable impact of beliefs by number of neglected objects in this experiment can be explained by the fact that object neglect takes place only when distractor objects are *visible* by the actor. Therefore, any gain in prediction resulting from prediction of object neglect decreases with a decreasing number of *visible* distractors. Overall, these results confirm that, during visual crowding and when the target is *not visible* by the actor, the effect of beliefs through behaviour interpretation may be that of reducing a target-close-to-actor bias (see also Exp. 5 *"Object Neglect and Beliefs Processing"* below).

Developmental analysis

From a developmental perspective, better target prediction accuracy was found with increasing experience in both architectures in all conditions, once again confirming the role of experience for improved target prediction and extending this to visual crowding during exploratory behaviour.

In addition, these results are coherent with the developmental trend seen in the previous experiments with respect to the impact of beliefs on target prediction. Indeed, when considering the condition in which beliefs were most impactful (i.e. Target *not visible* and 3 Objects *visible*), the gain in performance driven by beliefs processing was highest at a medium level of experience. Overall, this suggests that the developmental trend extends to visual crowding when target is *not visible* by the actor.

Exp. 4: Visual crowding when target is visible and Beliefs processing

In this experiment, I assessed whether the above strategy is also utilised to predict the actor's intentions with target *visible*, as well as whether processing beliefs would result in different performance in this condition. The rationale for conducting this experiment was that the dynamics identified in the previous experiments (i.e. an advantage of processing beliefs, as well as an informative role of *visible* objects by the actor) may change when trying to predict the intentions of an actor that is in close proximity, and thus can see, his target.

Similarly to the previous experiment (Exp. 3), a valid hypothesis (H1) is that the presence of distractor objects *visible* by the actor may confuse the observer when the target is also *visible*. However, in contrast to the previous experiment, discarding such distractor objects as potential targets based on an exploratory behaviour of the actor (H2) is not possible in the current experiment. Indeed, in the condition here assessed the actor would instead show a goal-oriented behaviour, at least for the latest chunk of the observed trajectory. As a result, less information may be extracted from the proximity of distractor objects to the actor.

As in the previous experiment, I tested this by including a variable number of distractor objects in the environment in the actor's FOV; however, in this case, the target was also visible by the actor. This resulted in the following conditions: (1) Target *visible* + 3 Objects *visible*, (2) Target *visible* + 2 Objects *visible*, (3) Target *visible* + 1 Object *visible*, (4) Target *visible* + No Objects *visible*. *Visible* and *not visible* distractor objects were generated and randomly located within and out of the actor's FOV, respectively. Results are shown in Tables 25A-D below.

Tables 25. Best target prediction accuracies for the *Beliefs* vs *NoBeliefs* architectures by number of training maps in the conditions with the target visible by the actor and ***(A)*** 3, ***(B)*** 2, ***(C)*** 1, and ***(D)*** No distractor objects *visible*. Accuracies were calculated as averages across 18 initial network weights; the associated variances were reported.

A.

	Target *visible* - 3 Objects *visible*					
	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	79.33	48.59	81.17	2.97	-1.83	.286
10	85.50	0.97	84.89	0.58	0.61	***.045***
15	86.50	0.50	85.83	0.38	0.67	***.005***
20	87.11	0.22	86.56	0.26	0.56	***.002***
25	87.17	0.26	86.44	0.26	0.72	***<.001***
30	87.33	0.24	86.61	0.25	0.72	***<.001***
60	87.94	0.06	87.72	0.21	0.22	.080
120	88.28	0.21	88.00	0.00	0.28	***.020***
300	88.94	0.06	88.28	0.21	0.67	***<.001***
				avg.	0.29	
				max.	0.72	

B.

	Target *visible* - 2 Objects *visible*					
	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	80.56	55.08	82.50	3.32	-1.94	.288
10	86.94	1.00	86.61	0.72	0.33	.289
15	88.22	0.65	87.56	0.26	0.67	***.006***
20	89.06	0.17	88.22	0.18	0.83	***<.001***
25	89.17	0.15	88.28	0.33	0.89	***<.001***
30	89.44	0.26	88.53	0.26	0.92	***<.001***
60	90.06	0.29	89.89	0.10	0.17	.269
120	90.72	0.21	90.28	0.21	0.44	***.007***
300	91.28	0.21	90.94	0.06	0.33	***.010***
				avg.	0.29	
				max.	0.92	

C.

	Target *visible* – 1 Object *visible*					
	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	82.44	56.97	84.17	5.68	-1.72	.362
10	89.22	1.59	88.56	1.56	0.67	.120
15	90.39	1.66	89.94	0.53	0.44	.211
20	91.50	0.26	90.33	0.35	1.17	***<.001***
25	91.78	0.30	90.89	0.81	0.89	***.002***
30	92.17	0.38	91.24	0.32	0.93	***<.001***
60	92.78	0.54	92.83	0.15	-0.06	.788
120	93.56	0.38	93.33	0.24	0.22	.237
300	94.17	0.15	94.00	0.00	0.17	.743
				avg.	0.30	
				max.	1.17	

D.

	Target *visible* – No Objects *visible*					
	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	84.17	65.68	86.06	8.17	-1.89	.358
10	91.61	2.13	91.00	2.24	0.61	.233
15	92.89	2.10	92.56	0.73	0.33	.407
20	94.17	0.50	93.11	1.16	1.06	***<.001***
25	94.72	0.68	93.72	1.62	1.00	***.015***
30	95.33	0.47	94.24	0.32	1.10	***<.001***
60	95.78	0.89	96.00	0.35	-0.22	.403
120	96.39	0.60	96.78	0.30	-0.39	.093
300	97.06	0.29	97.22	0.18	-0.17	.312
				avg.	0.16	
				max.	1.10	

BEL: *Beliefs* architecture; NoBEL: *NoBeliefs* architecture

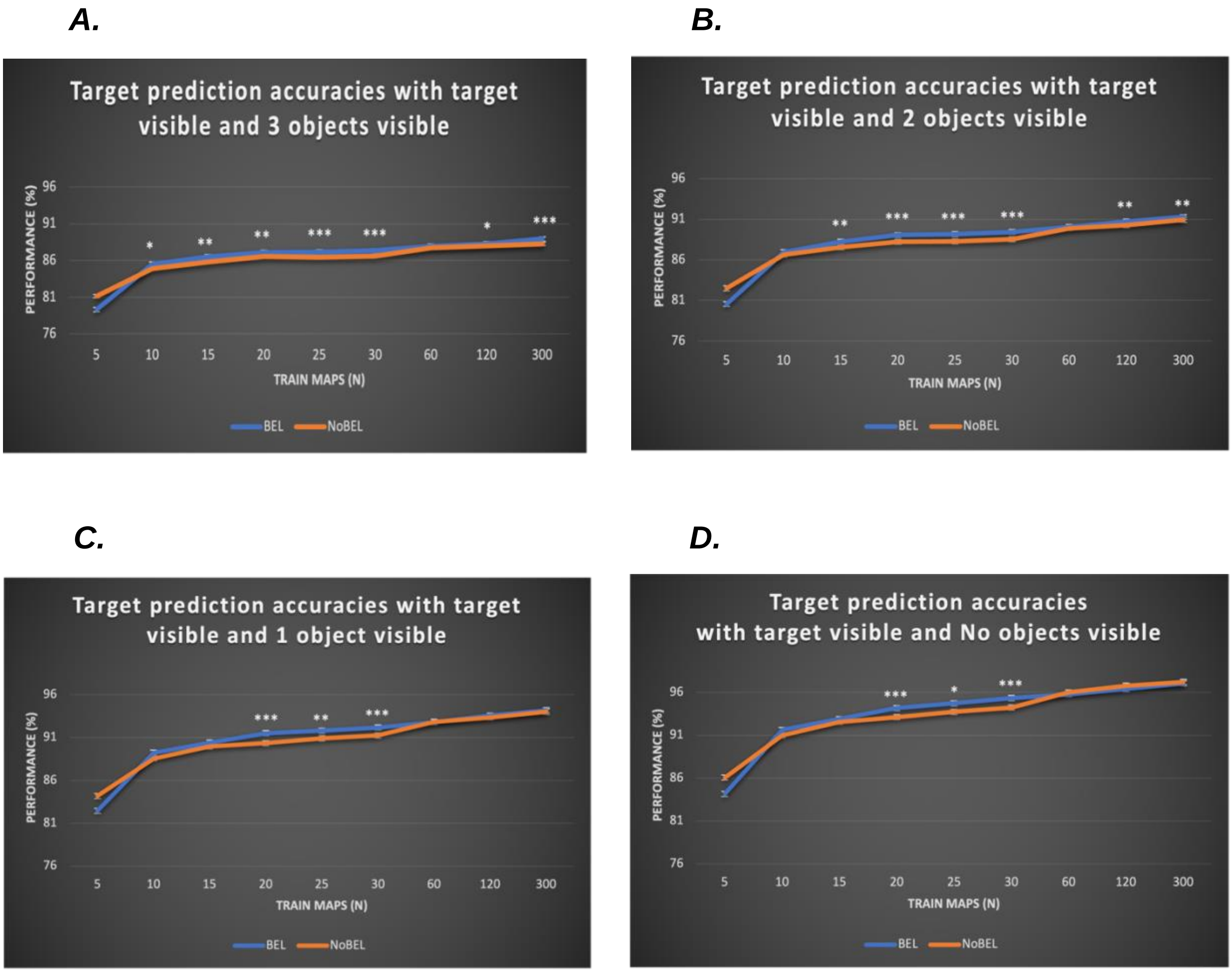
Within-architectures analysis

In Tables 25A-D, a general opposite trend is seen within-architectures compared to the previous experiment (Exp. 3). Specifically, when the target is *visible* by the actor, having less *visible* objects in the actor's FOV (no visual crowding) improves predictive performance, regardless of the architecture and the number of maps or experience (see also Figures 31A-D for a visualisation of the effect of visual crowding on predictive performance when the target is *visible*). In more detail, from the condition with 3 to No objects *visible* by the actor (and target *visible*), performances increase by ~7% in both the *Beliefs* and *NoBeliefs* architectures, when considering 25 train maps.

I hypothesised this to be a result of object *alignment* with the target, which could confuse an observer on which is the real target of an actor and thus lead to poorer performance, regardless of the number of maps or experience. Alignment is expected to be more relevant when the target is *visible* (i.e. during goal-directed behaviours), as approaching an object may look intentional and not casual. Overall, this interpretation would suggest that the observer was able to recognise the goal-directed behaviour of the actor, but that however was unsure on which was the actor's target, likely due to object-target alignment.

Furthermore, the opposite results seen here compared to the previous experiment (Exp. 3) may also be explained by a relationship between alignment and predicted object neglect. Specifically, when the target is *not visible* by the actor, thus during exploratory behaviours, aligned objects which are *visible* by the actor may be considered neglected by an observer (see Exp. 3). In contrast, when the target is *visible* by the actor, thus during goal-directed behaviours, the presence of aligned objects would be more ambiguous and render it more challenging to differentiate them from the target, given that these are on the actor's future efficient path (see current experiment). Therefore, it may be important for the observer to (a) distinguish between actor's goal-oriented behaviours (i.e. with *visible* target) and exploratory ones (i.e. when target is *not yet visible*) and (b) evaluate the target-approaching behaviours of the actor differently in the two cases. The hypothesis of a negative effect of object alignment with target on prediction of the actor's intentions, as well as the role of beliefs processing and behaviour prediction were further tested in Exp. 6 ("*Object Alignment and Beliefs Processing"*), where all objects were placed in the future actor's trajectory.

Figures 31. Visualisation of the effect of visual crowding on predictive performance of the *Beliefs* and *NoBeliefs* architectures when the target is visible by the actor. Specifically, higher target prediction accuracies are observed in both architectures from **(A)** 3 objects *visible* by the actor to **(B)** 2 objects *visible*, **(C)** 1 object *visible*, and **(D)** No objects *visible*, regardless of the number of maps. Average target prediction accuracy at each number of train maps; *: significant at the .05 level (2-tailed); **: significant at the .01 level (2-tailed); ***: significant at the .001 level (2-tailed).



Between-architectures analysis

Tables 25A-D show a statistically significant positive role for beliefs with all numbers of maps (except for 5 and 60 maps) when all the distractor objects were *visible* by the actor. This gain in significance, but not the magnitude, seemed to increase with the number of *visible* distractor objects (visual crowding) (see Tables 25A-D). However, this effect was less strong that the one observed in the previous experiment (Exp. 3) where the target was *not visible* (maximum gain in performance driven by beliefs processing: 1.17 vs 7.06%, respectively) (see Figure 32). Overall, these results suggest that the multi-task-induced regularisation effect earlier identified during visual crowding remains valid also when the target is *visible*, although this is much stronger when the target is *not visible*.

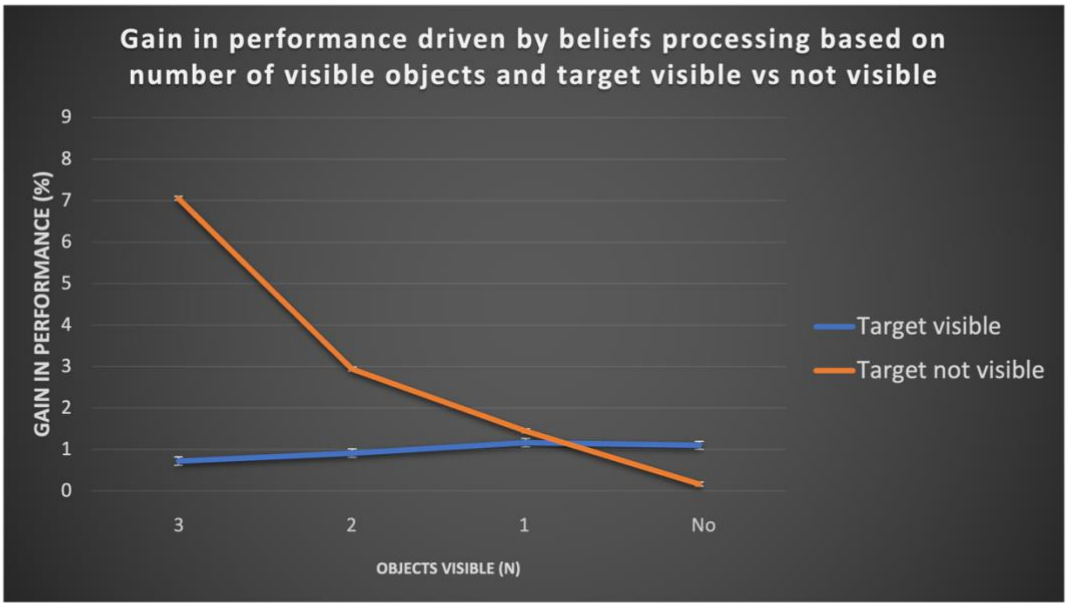


Figure 32. Gain in performance driven by beliefs processing based on the number of distractor objects visible by the actor when the target is *visible* vs *not visible*.

More in detail, when the target is *visible* by the actor, the gain in performance obtained from learning beliefs does not substantially differ when changing the number of visible distractor objects. This is in line with my previous findings of a limited impact of beliefs on goal-directed behaviours. However, beliefs show on average a higher significance when more distractor objects are *visible* (see Figures 31).

The reason behind these results may lie in the fact that, considering that the network input consists also of the actor's past steps and that these may include steps where the target was *not yet visible*, the observer may have seen the actor previously engaging in an explorative behaviour. My previous experiments indicate that belief processing is helpful for predicting explorative behaviours, allowing the observer to discard distractors close to the actor that can be considered neglected. Thus, the more objects are *visible* by the actor, the better the observer can predict them as neglected. This is likely due to a reduction of the harm of the "target-close-to-actor" bias which may occur with *visible* objects. Indeed, this bias is only helpful if the object next to the actor is the actual target. If this were the case, I would expect beliefs processing to be beneficial when the target is *visible* and objects are *visible* and when the aligned distractors are close enough to the actor to predict their neglect. When they are far, they would be difficult to discard based on the earlier-observed exploratory behaviour, as they were likely not in the actor's FOV. This hypothesis was further investigated in Exp. 6 below (*"Object Alignment and Beliefs Processing"*), where all objects were placed in the future trajectory of the actor, closer or farther away from the actor.

Developmental analysis

From a developmental perspective, better target prediction accuracy was found with increasing experience in both architectures in all conditions, once again confirming the role of experience for improved behaviour prediction and extending this to visual crowding during goal-directed behaviour.

Furthermore, these results are coherent with the developmental trend observed in previous experiments with respect to the impact of beliefs on target prediction. Indeed, when considering the condition in which beliefs were most impactful (i.e. Target *visible* and 3 Objects *visible*), the gain in performance driven by beliefs processing was highest at a medium level of experience, although this is limited (e.g. see Table 25D).

Exp. 5: Object Neglect and Beliefs Processing

In order to further investigate the relationship between beliefs processing, recognition of actor's behaviour, and detection of object neglect for target prediction, I conducted a separate experiment in an environment built for assessing object neglect. Specifically, a varying number of the three objects present in the environment, in addition to the target object, were placed in the last past steps of the actor. The objects being in the actor's past trajectory meant that such objects were seen and neglected by the actor during previous steps, thus providing clear additional information to the observer with regards to the actor's target identity. This experiment investigated the ability of the architectures to exploit this information. Furthermore, to determine whether detection of object neglect may have a role also when the target is *visible*, which may have influenced my previous results, this condition was also assessed.

Therefore, the following conditions were included in this experiment: (1) Target *visible* + 3 Objects neglected, (2) Target *visible* + 2 Objects neglected, (3) Target *visible* + 1 Object neglected, (4) Target *not visible* + 3 Objects neglected, (5) Target *not visible* + 2 Objects neglected, (6) Target *not visible* + 1 Object neglected. As above, this experiment was conducted for two sets of maps only, i.e. 25 and 120 maps. Distractor objects not placed in the past trajectory (i.e. conditions 2, 3, 5, 6) were randomly located outside the past trajectory and out of the actor's FOV. Results are shown in Tables 26A-D below.

Tables 26. Best target prediction accuracies for the *Beliefs* vs *NoBeliefs* architectures in the conditions with varying number of neglected objects and target *visible* by the actor (25 maps ***(A)***, 120 maps ***(C)***) or target not *visible* by the actor (25 maps ***(B)***, 120 maps ***(C)***). Accuracies were calculated as averages across 18 initial network weights; the associated variances were reported.

A.

	Target *visible* – 25 maps					
	BEL		NoBEL			
Objects Neglected (N)	Avg Acc (%)	Var	Avg Acc (%)	Var	Bel-NoBEL	*p-value*
3	99.00	0.00	99.00	0.00	0	1
2	98.89	0.10	98.44	0.26	0.44	***.010***
1	97.72	0.21	96.67	0.59	1.06	***<.001***

B.

	Target *not visible* – 25 maps					
	BEL		NoBEL			
Objects Neglected (N)	Avg Acc (%)	Var	Avg Acc (%)	Var	Bel-NoBEL	*p-value*
3	98.56	0.51	97.33	0.84	1.22	***<.001***
2	94.17	0.38	91.78	2.07	2.39	***<.001***
1	85.28	0.68	81.89	6.46	3.39	***<.001***

C.

	Target *visible* – 120 maps					
	BEL		NoBEL			
Objects Neglected (N)	Avg Acc (%)	Var	Avg Acc (%)	Var	Bel-NoBEL	*p-value*
3	99.00	0.00	99.00	0.00	0	1
2	99.00	0.00	98.61	0.37	0.39	***.010***
1	98.44	0.26	96.83	1.44	1.61	***<.001***

D.

	Target *not visible* – 120 maps					
	BEL		NoBEL			
Objects Neglected (N)	Avg Acc (%)	Var	Avg Acc (%)	Var	Bel-NoBEL	*p-value*
3	98.94	0.24	98.72	0.46	0.22	.77
2	94.67	0.24	90.78	13.48	3.89	***<.001***
1	86.22	0.77	79.67	33.65	6.56	***<.001***

BEL: *Beliefs* architecture; NoBEL: *NoBeliefs* architecture

Within-architectures analysis

In both architectures, an increasing number of objects neglected (both when the target was or was not *visible*) leads to better performance, regardless of the number of maps or experience. This effect was most obvious for the conditions in which the target was *not visible* (4-6), although the *NoBeliefs* architecture was found to overfit in this condition. Specifically, up to ~2% improved performance was seen when target was *visible* vs ~14% when target was *not visible* (when discarding cases with high variance). These results support my interpretation of the previous data, in that detection of object neglect is a factor guiding target prediction in both goal-directed and exploratory behaviours, although with a bigger impact on the latter. Furthermore, these results suggest that the neural networks were able to detect object neglect and use this information for guiding target prediction.

Interestingly, when comparing the present results with those from the experiments of visual crowding with target *not visible* (Exp. 3) and *visible* (Exp. 4), I could further confirm the hypothesised effect of detection of object neglect on target prediction, especially when the target is *not visible* by the actor (see Table 27 below).

Table 27. Gain in performance in the *Beliefs* and NoBeliefs architectures resulting from one object *neglected* compared to *No* or *one object* in the actor's FOV, both by target visibility and number of maps.

	One Object Neglected vs ***No*** Objects *visible*				One Object Neglected vs ***One*** Object *visible*			
	BEL		NoBEL		BEL		NoBEL	
Target	25 maps	120 maps	25 maps	120 maps	25 maps	120 maps	25 maps	120 maps
Visible	3%	2%	3%	0%	6%	5%	6%	4%
Not Visible	49%	49%	46%	42%	46%	45%	44%	40%

BEL: *Beliefs* architecture; NoBEL: *NoBeliefs* architecture

In more details, I observed an improvement in performance when comparing the condition of *one neglected object* in the actor's past trajectory (current experiment) and the condition with *no* distractor objects in the actor's FOV with target *visible* and *not visible* (Exps. 3-4). Specifically, when considering the *Beliefs* architecture and 25 maps as an example, a ~49% improvement in performance was seen when including only *one* neglected object in the actor's past trajectory and the target was *not visible* (thus during exploratory behaviours). An improvement in performance was also seen when the target was *visible* by the actor (thus during goal-directed behaviours), although this was much smaller (~3%). In the *NoBeliefs* architecture, an improvement in performance of ~46% and ~3% was also achieved, respectively (see Figure 33). Similar results were obtained for 120 maps, although the gain in performance was lower (see Table 27 for more details).

Even more strikingly, I observed improved performance also when comparing the condition of *one neglected object* in the actor's past trajectory (current experiment) and the condition with *one* distractor objects randomly placed in the actor's FOV with target *visible* and *not visible* (Exps. 3-4). Specifically, when considering the *Beliefs* architecture and 25 maps as an example, a ~46% gain in performance was observed when including only *one* neglected object in the actor's past trajectory and the target was *not visible* (thus during exploratory behaviours). An improvement in performance was also seen when the target was *visible* by the actor (thus during goal-directed behaviours), although this was much smaller (~6%). In the *NoBeliefs* architecture, an improvement in performance of ~44% and ~6% was achieved, respectively (see Figure 33). Similar results were obtained for 120 maps, although the gain in performance was lower (see Table 27 for more details).

Finally, the above analyses indicate an increasing gain in performance driven by the condition "One object *neglected* and *one visible* distractor" vs "One object *neglected* and *No visible* distractor" when the target is *visible*; while the opposite is valid when the target is *not visible*. These results once again support a different informative nature of objects based on the recognised types of behaviours (i.e. during exploratory behaviours, a visible object can be easily predicted as neglected).

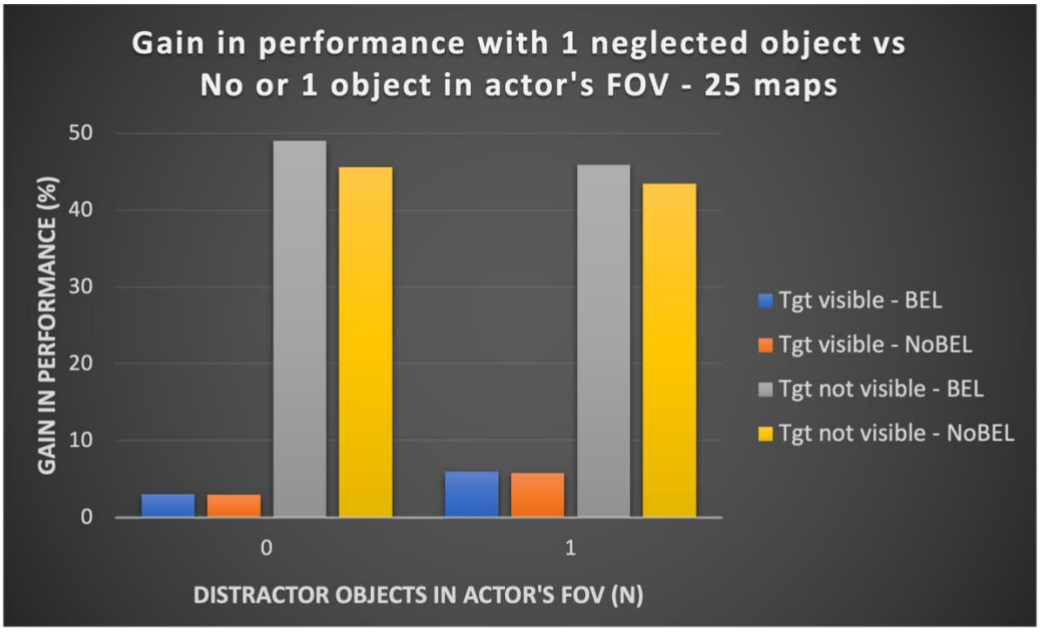


Figure 33. Visualisation of gain in performance in the *Beliefs* and *NoBeliefs* architectures resulting from one object neglected vs No or one object in the actor's FOV, both by target visibility (25 train maps).

Overall, my results confirm the strong contribution of detection of neglected objects towards target prediction, especially when the target is *not visible* by the actor, who engages in an exploratory behaviour. Furthermore, findings indicate that the contribution of *visible* distractor objects to target prediction increases when they are clearly neglected (e.g. on the actor's past trajectory and not just in the actor's FOV). Thus, a neglected distractor object is much more informative than a randomly positioned distractor. Finally, my results suggest that the informative nature of objects is influenced by type of behaviour (goal-directed vs exploratory); therefore, recognition of such behaviours is informative and allows to accurately use the information derived from objects in the environment to achieve better target prediction performance.

Between-architectures analysis

In Tables 26A-D, I continued to observe a statistically significant better performance in the *Beliefs* vs the *NoBeliefs* architecture for most conditions, except for conditions (1) and (4, 120 maps), i.e. when all 3 distractor objects were neglected. Furthermore, this statistically significant difference in performance between the two architectures increased as the number of neglected objects decreased. Finally, data suggest that beliefs have a stronger impact on performance in conditions in which the target is *not visible*. See Figure 34 below for a visualisation of this data.

These results indicate that detection of object neglect may not interact with beliefs, considering the limited impact that beliefs processing has on target prediction when all objects have been neglected. Instead, these results indicate that beliefs processing may help disambiguate the remaining objects in the environment which have not been neglected in the past trajectory, but which are not approached with an efficient trajectory. This is in concordance with the finding that the impact of beliefs processing in this experiment was higher in the target *not visible* condition, indicating that beliefs may indeed help recognise exploratory behaviours and in turn reduce the "target-close-to-actor" bias through predicted object neglect, ultimately improving target prediction. This is further investigated in the next experiment (Exp. 6). The impact that beliefs have on the target *visible* condition, thus in goal-directed behaviours, may be due to beliefs resulting in better detection of the actor's explorative behaviour in the early phases of the observed behavioural chunk. During that phase, the observer may have assumed the predicted object neglect strategy to consider some objects as non-targets, e.g. when objects are aligned on the trajectory to the target which is however not yet visible. The relationship between object alignment with target, recognition of actor's behaviour, and beliefs processing was further investigated in the next experiment (Exp. 6).

Overall, these findings confirm that detection of object neglect results in improved performance and suggest that this happens regardless of the presence of beliefs processing. However, they also support previous results indicating that that learning to explicitly represent others' beliefs aids the recognition of explorative behaviours and, in turn, the disambiguation of distractor objects, possibly using a predicted object neglect strategy. Therefore, we may hypothesise the existence of a relationship between beliefs processing, recognition of actor's behaviour, and prediction of object neglect which was better investigated in the next experiment (Exp. 6).

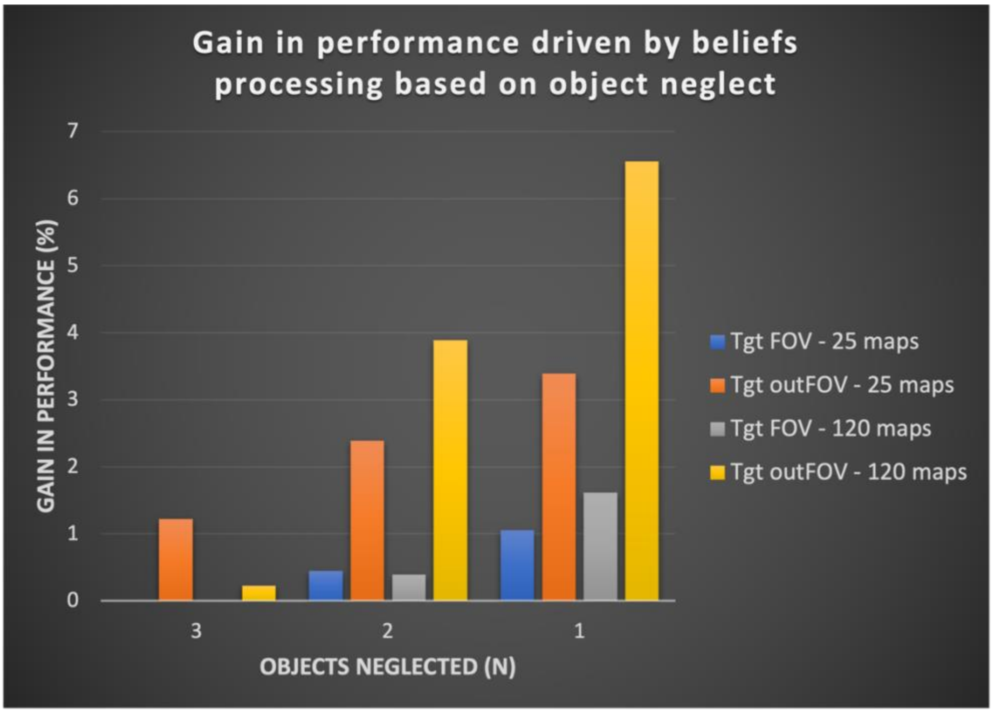


Figure 34. Visualisation of gain in performance driven by beliefs processing based on the number of objects neglected, by target visibility and number of maps.

Developmental analysis

Although the developmental trend was not systematically assessed here as in previous experiments, in concordance to my previous findings, target prediction performance, when including neglected objects, improved with experience in all conditions. An increasing role of beliefs towards target prediction, when including neglected objects, was found with increasing experience[7]. This may be a result of the limited percentage of samples made available during training showing perfect neglect and, therefore, the need for further experience to achieve the maximum associated gain.

Exp. 6: Object Alignment and Beliefs Processing

Finally, to investigate the relationship between object alignment with target, belief processing, recognition of behaviour, and prediction of object neglect, I conducted a separate experiment in an environment built for assessing object alignment. Specifically, a varying number of three objects present in the environment, in addition to the target object, were placed in the next steps of the actor's trajectory.

In conditions with target *visible* by the actor, I expect all steps to be efficient towards the target, as the actor is following a goal-directed trajectory[8]. In contrast, in conditions with target *not visible*, steps may not to be necessarily efficient. Therefore,

[7] Please note that both target prediction accuracy and impact of beliefs were seen to mostly increase with experience in the *NoBeliefs* condition. However, the high variance associated with this condition may indicate overfit of the model, thus I cannot provide any conclusive remarks on this regard.

[8] Based on a follow-up analysis of steps statistics (see supplementary material 1), in my simulated trajectories, the actor generally takes a direct path to the target once *visible*, reaching the target within 3 steps. As a result, if distractor objects are placed in the actor's next steps, and given that the target is generally reachable or visible within a few steps, this condition would result object-target alignment. Please note that, as the third step was, in most simulations, the location of the target, the third object was not placed exactly on the third step but in the next cell on the grid; thus, it was in this case not directly aligned with the target but next to the target instead.

I defined a distractor object *aligned* with respect to the future actor's behaviour, rather than with the actual fastest path to the target. Crucially, predicting the actor's next steps, in this case, is different from predicting the best-informed path to the target.

As a result, the included conditions in this experiment were the following: (1) Target *visible* + 3 Objects aligned, (2) Target *visible* + 2 Objects aligned, (3) Target *visible* + 1 Object aligned, (4) Target *not visible* + 3 Objects aligned, (5) Target *not visible* + 2 Objects aligned, (6) Target *not visible* + 1 Object aligned. As above, this experiment was conducted for only the 25 and 120 maps. Distractor objects not placed in the future trajectory (i.e. conditions 2, 3, 5, 6) were randomly placed outside the future trajectory and out of the actor's FOV. Results are shown in Tables 28A-D below.

Tables 28. Target prediction accuracies for the *Beliefs* vs *NoBeliefs* architectures in the conditions with varying number of aligned objects and target visibility. ***(A)*** Target *visible*, 25 maps; ***(B)*** Target *not visible*, 25 maps; ***(C)*** Target *visible,* 120 maps; and ***(D)*** Target *not visible*, 120 maps. Accuracies were calculated as averages across 18 initial network weights; the associated variances were reported.

A.

	Target *visible* - 25 maps					
	BEL		NoBEL			
Objects Aligned (N)	Avg Acc (%)	Var	Avg Acc (%)	Var	Bel-NoBEL	*p-value*
3	43.50	12.85	36.22	12.65	7.28	***<.001***
2	40.94	19.70	31.61	15.78	9.33	***<.001***
1	51.44	18.97	45.78	28.07	5.67	***.001***

B.

	Target *not visible* - 25 maps					
	BEL		NoBEL			
Objects Aligned (N)	Avg Acc (%)	Var	Avg Acc (%)	Var	Bel-NoBEL	*p-value*
3	16.61	14.02	11.56	8.38	5.06	***<.001***
2	17.56	23.91	10.83	10.50	6.72	***<.001***
1	42.56	33.08	34.83	37.44	7.72	***<.001***

C.

	Target *visible* - 120 maps					
	BEL		NoBEL			
Objects Aligned (N)	Avg Acc (%)	Var	Avg Acc (%)	Var	Bel-NoBEL	*p-value*
3	49.33	6.82	43.22	14.18	6.11	***<.001***
2	46.61	3.01	36.83	4.05	9.78	***<.001***
1	56.67	15.88	43.83	12.97	12.83	***<.001***

D.

	Target *not visible* - 120 maps					
	BEL		NoBEL			
Objects Aligned (N)	Avg Acc (%)	Var	Avg Acc (%)	Var	Bel-NoBEL	*p-value*
3	15.39	16.02	12.11	16.34	3.28	***.020***
2	14.94	20.17	10.89	16.22	4.06	***.007***
1	46.28	33.27	35.06	31.35	11.22	***<.001***

BEL: *Beliefs* architecture; NoBEL: *NoBeliefs* architecture

Within-architectures analysis

Tables 28A-D indicate the presence of high variances throughout, which may be a result of the limited percentage of samples made available during training showing perfect alignment, and model overfit. Therefore, while I did attempt data interpretation given the statistically significant results, care should be taken when interpretating results.

In both architectures, worse performance is seen when 3 vs 1 objects are aligned with the target (both with target *visible* and *not visible*). These results support my interpretation of the previous data, in that object alignment with actor's future behaviour is a factor *negatively* affecting target prediction, both during goal-directed and exploratory behaviours.

Counterintuitively, (in most cases) performance is seen to increase when including 3 vs 2 objects aligned with the target. This may be explained through prediction of object neglect as follows. According to the prediction of object neglect strategy, during exploratory behaviours, near distractors are easier to consider as neglected by the actor, thus non-targets. Therefore, when three aligned objects are present, one of them must be close to the actor, the second less close and the third farther away. In contrast, when 2 aligned objects are only present, there will be one third of the cases in which the closest objects will be absent, thus removing the prediction of object neglect gain that we know is strong. While this can explain the results seen in the target *not visible* condition here observed, it can also explain results seen for the target *visible* condition. Indeed, while in this condition the target is *visible* in the current step, it may not have been in previous steps; therefore, the actor may have had an exploratory behaviour that the observer could have used to predict object neglect, which is most likely when 3 objects are present.

To further investigate the impact of aligned objects in goal-directed and exploratory behaviours, I compared the present results with those from the experiments of visual crowding with target *not visible* (Exp. 3) and *visible* (Exp. 4). As a result, I could further confirm the general *negative* effect of object-target alignment on target prediction when the target is *visible* by the actor, thus during goal-directed behaviours. However, the *Beliefs* architecture seemed to be able to find it informative when the target is *not visible*, thus during exploratory behaviours (see Table 29 below).

Table 29. Gain in performance (%) in the *Beliefs* and *NoBeliefs* architectures resulting from one object *aligned* compared to *No* or *one object* in the actor's FOV, both by target visibility and number of maps.

	One Object Aligned vs ***No*** Objects *visible*				One Object Aligned vs ***One*** Object *visible*			
	BEL		NoBEL		BEL		NoBEL	
Target	25 maps	120 maps	25 maps	120 maps	25 maps	120 maps	25 maps	120 maps
Visible	-43%	-40%	-48%	-53%	-40%	-37%	-45%	-50%
Not Visible	6%	9%	-1%	-3%	3%	5%	-4%	-5%

BEL: *Beliefs* architecture; NoBEL: *NoBeliefs* architecture

In more details, I observed an improvement in performance (~6%) by the *Beliefs* architecture when the target *was not visible* (25 maps)*,* when comparing the condition of *one aligned object* in the actor's future trajectory (current experiment) and the condition with *no* distractor objects in the actor's FOV (Exps. 3-4). In contrast, a drop in performance (~43% decrease) was observed when the target was *visible* by the actor. In the *NoBeliefs* architecture, a drop in performance was seen both when the target was *not visible* and *visible* by the actor, although such a decrease in performance was much more negative in the latter condition (~1 vs 48% decrease, respectively) (see Figure 35). These results thus indicate that object alignment can be informative for the interpretation of exploratory behaviours, only when beliefs are considered as well. Specifically, these results support previous findings on a role for beliefs in recognising exploratory behaviours and, in turn, disambiguating objects in the environment (possibly through the predication of object neglect).

Furthermore, I continued to observe an improvement in performance (~3%) by the *Beliefs* architecture when target was *not visible* (25 maps), when comparing *one* aligned distractor object in the actor's future trajectory (current experiment) with the condition with *one* distractor object randomly placed in the actor's FOV (Exps. 3-4). However, this gain was lower than that observed in the previous paragraph when *no* distractor objects were visible. These results support findings in the previous paragraph and further highlight that having one *visible* object still helps target prediction during exploratory behaviours, through prediction of object neglect. This can be explained by the fact that aligned distractors are usually closer to the agent than randomly placed *visible* objects. In turn, the closer to the agent the distractor object is, the easier it is to detect it as neglected, i.e. not being reached efficiently by the agent after being spotted. This neglect event can be predicted and contribute to the recognition of the actor's intention, while the other distractor objects may not be spotted and act as harmful distractors.

In contrast, I continued to observe a drop in performance in both the *Beliefs* architecture with target *visible*, and in the *NoBeliefs* architecture both with *visible* and *not visible* target. Only the drops observed when the target was *visible* were less negative than those observed in the previous paragraph when *no* distractor objects were visible (see Figure 35), confirming that object alignment or *visibility* is particularly negative for predicting goal-directed behaviours and if not considering beliefs.

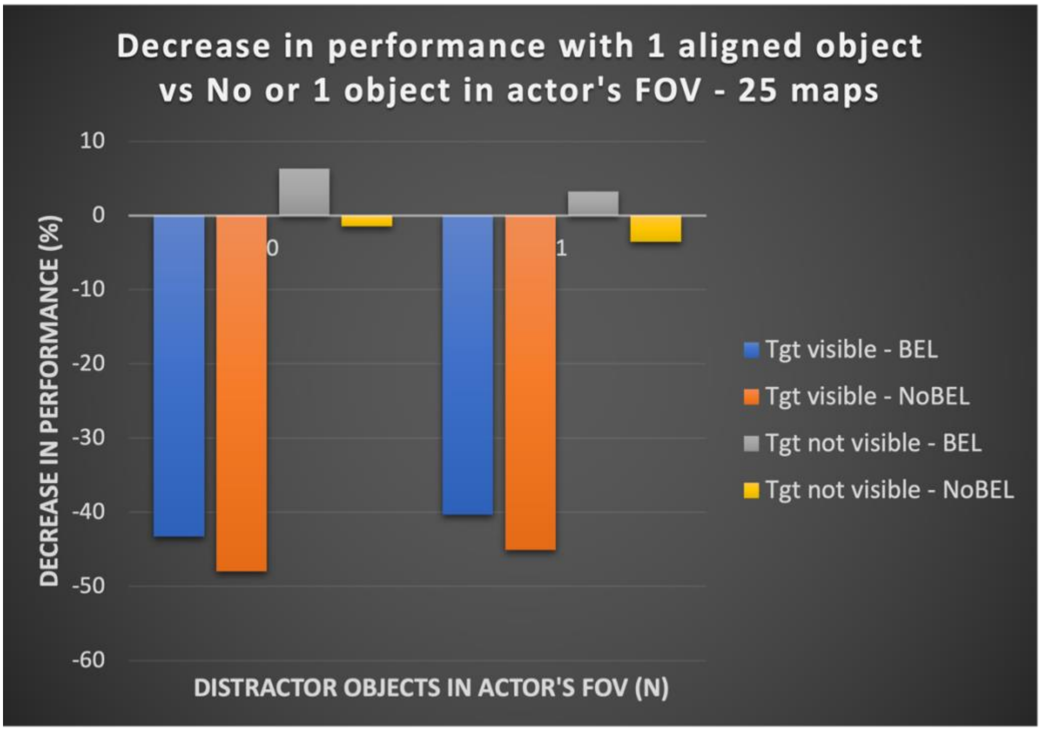


Figure 35. Visualisation of decrease in performance in the *Beliefs* and *NoBeliefs* architectures resulting from *one* object aligned vs *No* or *one* object in the actor's FOV, both by target visibility (25 train maps).

Overall, these results confirm the *negative* impact of object alignment with target towards target prediction, regardless of the actor's behaviour. However, when comparing object alignment with conditions of one *visible* object or absence of objects in the environment, results suggest a positive contribution of object alignment towards target prediction only when the target is *not visible* and beliefs processing is included. These results may thus support my previous findings indicating that beliefs processing aids the recognition of exploratory behaviours, which is crucial in conditions of object alignment, and in turn the disambiguation of objects in the environment, possibly through prediction of object neglect. To conclude, these results suggest that the informative nature of aligned objects is influenced by a crucial ability to interpret actor's behaviours in this condition, which is driven by beliefs processing.

Between-architectures analysis

In Tables 28A-D, I continued to observe a statistically significant better performance in the *Beliefs* vs *NoBeliefs* architecture for all conditions. Furthermore, this statistically significant difference in performance between the two architectures increased as the number of aligned objects decreased (except for Target *visible* + 2 vs 1 object aligned, 25 maps). See Figure 36 below for a visualisation of this data.

I interpret these results to indicate that object alignment ultimately challenges both architectures towards target prediction; however, they point to a better performance of the *Beliefs* architecture in this condition. This may be driven by the strong contribution of beliefs in exploratory behaviour recognition and thus better prediction of object neglect, overall resulting in better target prediction. This interpretation is supported by the fact that object alignment leads to better performance compared to No or one *visible* objects only in the exploratory condition in the *Beliefs* architecture. Finally, the decrease in the impact of beliefs with an increasing number of aligned objects along the actor's trajectory is likely due to the fact that some objects may not have been visible at all in the actor's previous steps, thus decreasing the gain induced by beliefs processing on recognition of exploratory behaviour and following prediction of distractor object neglect.

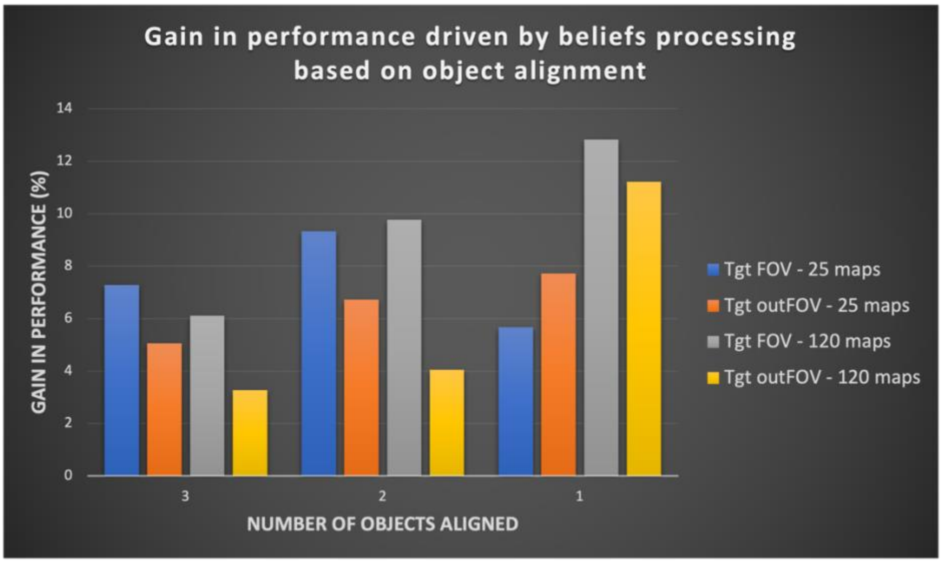


Figure 36. Visualisation of gain in performance driven by beliefs processing based on the number of objects aligned, by target visibility and number of maps.

Developmental analysis

In accordance with my previous studies, increasing experience was shown to improve target prediction performance also in the object-target alignment condition. In contrast, I here failed to identify a linear developmental trend for the impact of beliefs on target prediction when considering object alignment, as the role of beliefs varied based on the conditions of the environment. Once again, these interpretations should be considered with caution given the high variance observed in all conditions.

Conclusions

In summary, results from study 2 indicate a role for beliefs towards others' behaviour prediction, which leads to faster learning and improved performance of others' intentions predictions. Furthermore, study 2 sheds light into the conditions which benefit from this relationship between beliefs processing and behaviour recognition the most (and the least). More details are provided below.

- Generally, better target accuracy was achieved in the target *visible* vs target *not visible* condition, suggesting that it may be easier for an observer to recognise and predict an actor's goal-directed behaviour, rather than exploratory behaviour.

  With regards to human behaviour, this may indicate that the ability to recognise and predict goal-directed, as opposed to exploratory, behaviours may be mastered earlier in development; however, they both improve with experience.

- Learning to explicitly represent beliefs was seen to compensate for the delayed learning and prediction of exploratory behaviours, rendering it faster and resulting in better prediction of others' intentions. This positive impact of beliefs processing on behaviour and target prediction followed the developmental trend identified in study 1. Beliefs were instead shown to be mostly redundant during goal-directed behaviours.

  With regards to human behaviour, results support the early development of beliefs representation. In addition, they indicate that beliefs processing can be used to aid the recognition and interpretation of others' behaviours, especially when actors have partial knowledge about the environment and engage in exploratory behaviours. This is true in particular when the observer has had some experience with similar situations. Ultimately, given that exploratory behaviours require actors to rely more strongly on their beliefs, if an observer is aware of them (i.e. has a ToM), a substantial gain in performance is obtained.

- Visual crowding results in improved performance when the target is *not visible* (i.e. exploratory behaviour), while the opposite is valid when the target is *visible* (i.e. goal-directed behaviour). Beliefs were shown to aid the recognition of exploratory behaviours during visual crowding and, in turn, the disambiguation of objects in the environment (likely due to prediction of object neglect), resulting in improved prediction of others' intentions. Beliefs were shown to be mostly redundant in visual crowding conditions during goal-directed behaviours. However, their impact increased with increased visual crowding (likely due to a relationship between object-target alignment and prediction of object neglect).

  With regards to human behaviour, this suggests that recognising (goal-directed vs exploratory) behaviours is useful for the discrimination of *visible* objects as non-targets. Beliefs processing supports behaviour recognition (through faster learning), which ultimately results in better discrimination of objects in the environment and improved prediction of others' intentions. Furthermore, it indicates that observers rely on the informativity of visible objects in the environment for others' intentions prediction.

- Detection of object neglect is a factor guiding target prediction, regardless of the actor's observed behaviour, although it is mostly important in exploratory behaviours. Furthermore, object neglect represents a more informative source of information than having a randomly positioned *visible* object. However, I here identified that the actor's behaviour influences the informativity of objects in the environment for the observer. Beliefs aid the recognition of the correct behaviour and, in turn, the disambiguation of objects in the environment (likely through prediction of object neglect and reduced "target-close-to-actor" bias).

With regards to human behaviour, this extends previous results by highlighting the importance of recognising an actor's correct behaviour to accurately use information derived from objects in the environment to achieve better target prediction performance. Furthermore, it shows that this ability is supported by beliefs processing, following a developmental trend.

- Object alignment negatively affects target prediction performance, regardless of the actor's observed behaviour. However, it is more informative than *visible* or absent objects in the environment when observing an actor engaging in exploratory behaviour, an effect mediated by beliefs processing. Therefore, I identified the informative nature of aligned objects to be influenced by a crucial ability to interpret actor's behaviours in this condition, which is driven by beliefs processing. These findings support the hypothesis of beliefs aiding the disambiguation of objects through prediction of object neglect and reduced "target-close-to-actor" bias.

  With regards to human behaviour, this supports previous findings by showing an advantageous role for beliefs processing towards others' behaviour recognition, which is crucial for disambiguating objects in the environment given challenging conditions (such as target-object alignment).

Overall, study 2 advances the previous findings by highlighting that, rather than being random, the multi-task-induced regularisation between beliefs and target processing is mostly beneficial for recognising others' behaviours and interpreting exploratory behaviours. This in turn allows better disambiguation of objects in the environment, ultimately resulting in improved prediction of others' intentions. While supporting early ToM emergence, this study also highlights the advantages of having a ToM (intended here as beliefs processing) for improved human behaviour prediction in several, and challenging, situations.

3.3 Generalisation of architectural choice using "like them" approach

In study 3, I explored whether the advantageous effect of learning to explicitly represent beliefs for others' behaviour prediction through a "like them" approach could generalise to observers and environments of varying complexity. Specifically, I investigated this by first varying the observer's complexity (i.e. varying neural network layers), while maintaining the same environment. Successively, I changed the environmental complexity (i.e. number of objects in the grid world), while maintaining the same observer's complexity. The aim of this study was to show whether the architectural choices made in this project can withstand changes and support behaviour prediction also in other situations.

Exp. 1: Observer's complexity

First, I investigated whether learning to explicitly represent others' beliefs generalises to other observers. Specifically, I changed the layers in both the *Beliefs* and *NoBeliefs* neural networks to investigate the impact of the observer's neural complexity on performance. While two layers were originally used in studies 1 and 2, the following number of layers were investigated in this experiment: 1, 5, and 8. Given the previously identified impact of beliefs on target prediction on 25, 120 and 300 maps, such maps were used for training in these new conditions. The original environmental complexity was used for this experiment. The results of these runs are summarised in Tables 30A-D below.

Tables 30. Target prediction accuracies for the *Beliefs* vs *NoBeliefs* architectures in the conditions with varying complexity of the observer. Neural network layers: ***(A)*** *1* (simpler observer); ***(B)*** *2* (original observer); ***(C)*** *5* (more complex observer); and ***(D)*** *8* (more complex observer). Accuracies were calculated as averages across 18 initial network weights; the associated variances and learning rates were reported.

A.

Neural network layers: *1*

	BEL			NoBEL			BEL-NoBEL	*p-value*
Train Maps (N)	LR	Avg Acc (%)	Var	LR	Avg Acc (%)	Var		
25	0.001	68.64	1.26	0.001	67.61	0.86	1.03	***.003***
120	0.00075	71.19	0.69	0.0005	70.31	0.59	0.88	***.002***
300	0.00015	72.18	0.27	0.00015	71.51	0.42	0.66	***.001***
	avg	70.67				0.63	0.86	
	var	3.34				0.05		

B.

Neural network layers: *2 - Original*

	BEL			NoBEL			BEL-NoBEL	*p-value*
Train Maps (N)	LR	Avg Acc (%)	Var	LR	Avg Acc (%)	Var		
25	0.001	69.45	1.63	0.001	67.57	1.44	1.87	***<.001***
120	0.00075	71.47	0.87	0.0005	70.46	0.89	1.00	***<.001***
300	0.00015	72.43	0.28	0.00015	71.59	0.43	0.84	***<.001***
	avg	71.12				0.92	1.24	
	var	2.32				0.25		

C.

Neural network layers: *5*

	BEL			NoBEL			BEL-NoBEL	*p-value*
Train Maps (N)	LR	Avg Acc (%)	Var	LR	Avg Acc (%)	Var		
25	0.001	68.46	1.40	0.001	67.77	0.91	0.69	.061
120	0.00075	71.01	0.54	0.0005	69.87	0.72	1.14	***<.001***
300	0.00015	71.49	0.38	0.00015	71.00	0.46	0.49	***.008***
	avg	70.32				69.88	0.77	
	var	2.64				4.30		

D.

Neural network layers: *8*

	BEL			NoBEL			BEL-NoBEL	*p-value*
Train Maps (N)	LR	Avg Acc (%)	Var	LR	Avg Acc (%)	Var		
25	0.001	68.34	1.53	0.001	67.58	1.47	0.76	.062
120	0.00075	70.81	0.93	0.0005	69.96	1.19	0.86	***.003***
300	0.00015	71.34	0.57	0.00015	70.78	0.43	0.56	***.009***
	avg	70.16			avg	1.03	0.73	
	var	2.57			var	0.29		

LR: learning rate; BEL: *Beliefs* architecture; NoBEL: *NoBeliefs* architecture

Within-architectures analysis

Tables 30A-D show that averaged accuracies across maps generally result in similar accuracies across different observers. Furthermore, compared to the original observer's complexity used in studies 1 and 2 (i.e. 2 layers), similar averaged accuracies for both architectures were here observed, although they were slightly higher in the original condition. See Figure 37 for an example visualisation of this data. Overall, these results may suggest that the observer's complexity may not extremely change overall performance towards predicting others' intentions.

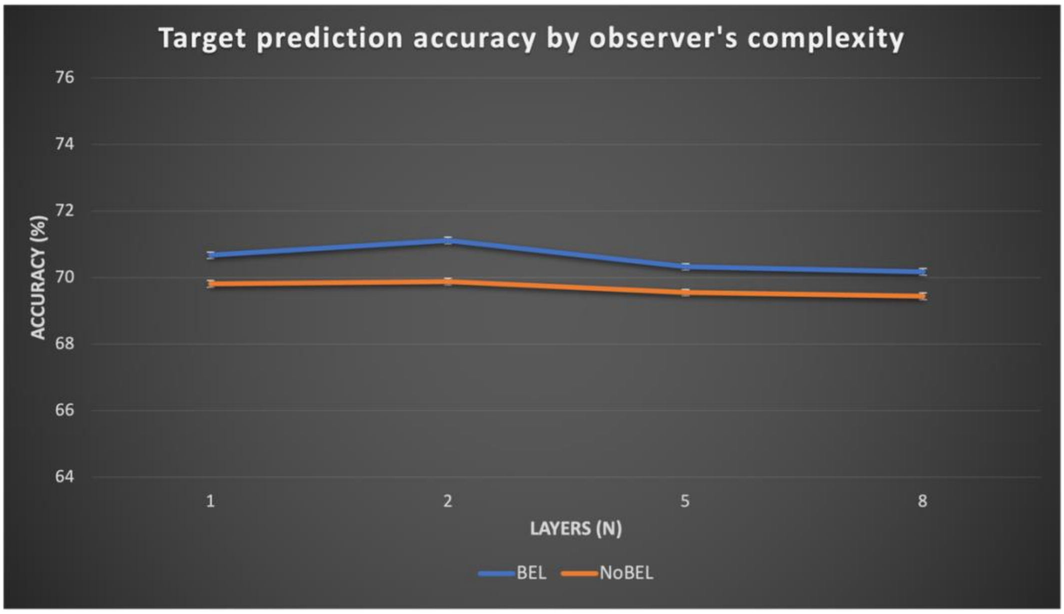


Figure 37. Visualisation of target prediction accuracy by the *Beliefs* and *NoBeliefs* architectures by observer's complexity (simpler vs more complex observers). Average target prediction accuracies across train maps for each observer are reported.

Between-architectures analysis

With regards to the role of beliefs for target prediction, the above results indicate similar positive differences between the *Beliefs* and *NoBeliefs* architectures among all observers. Nonetheless, only the 1-layer observer resulted in statistical significance on all three numbers of maps, thus replicating the results seen in the original study 1. Furthermore, higher statistical significance was observed for the condition with 1 layer versus the conditions with 5 and 8 layers. However, a linear trend in the impact of beliefs on performance was not observed. See Figure 38 for a visualisation of this gain driven by beliefs processing. Overall, these results indicate that the multi-task-induced regularisation hypothesis maintains with this study and that generalisation to different observers' complexities is possible. In other words, from an applicative point of view, these results indicate that introducing beliefs processing is a good strategy for achieving improved target prediction performance, regardless of the observer's level of complexity.

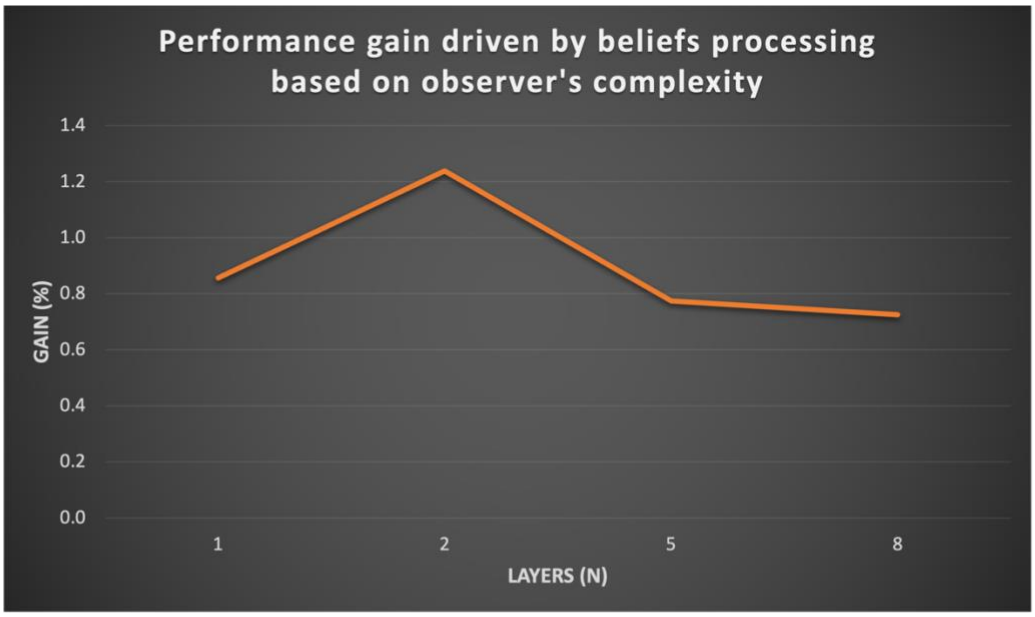


Figure 38. Gain in performance driven by beliefs processing based on observer's complexity (simpler vs more complex observers).

Developmental analysis

Regardless of the number of layers, a steady trend for achieving better performance was seen with an increasing number of maps made available during training, thus experience (see Tables 30A-D). With respect to the impact of beliefs processing on target prediction, slightly different trends were seen between observers. More in detail, similarly to the original study 1, the gain in performance driven by beliefs seemed to peak at 25 maps in the observer with 1 layer. In contrast, such peak seemed to be delayed (occurring at 120 maps) in observers with an increasing number of layers (more complex observers) (see Figure 39). Overall, this may indicate that while experience improves performance regardless of the complexity of the observer, the contribution of beliefs processing on target prediction may extend for a longer time during development in more complex observers. Nevertheless, my results show that the actual dynamics of the impact of learning explicit belief representations on target prediction may be influenced by the complexity of the observers.

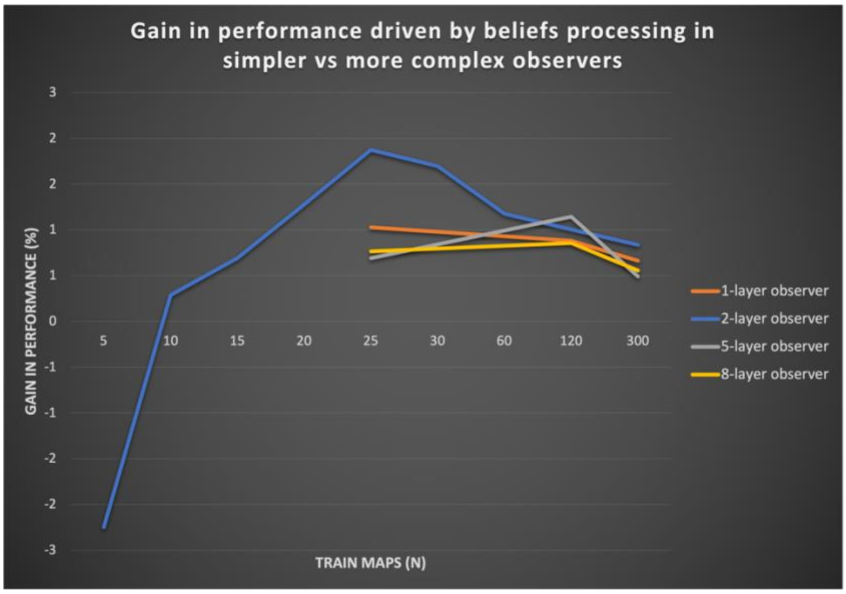


Figure 39. Delayed peak of impact of beliefs processing on target prediction accuracy in more complex vs simpler observers, by number of train maps.

Exp. 2: Environmental complexity

Next, I trained both architectures with simulations which had 8 distractor objects in the environment, as opposed to 4, and investigated their performance in predicting the actor's target. The distractor objects were still randomly placed within the 11 x 11 grid world. This way I assessed whether learning to explicitly represent beliefs can also generalise to other environments of higher complexity. The following neural networks from both configurations (and all seeds) were trained for this new condition: 25, 60, 90, 120, and 300 maps. The original observer's complexity was used for this experiment. The results of these runs are summarised in Tables 31 below.

Tables 31. Target prediction accuracies for the *Beliefs* vs *NoBeliefs* architectures in the conditions with varying complexity of the environment. Number of objects in the environment: ***(A)*** *4* (Original); and ***(B)*** *8* (more complex environment). Accuracies were calculated as averages across 18 initial network weights; the associated variances were reported.

A.

4 objects - *Original*

	BEL		NoBEL		BEL-NoBEL	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
25	69.45	1.63	67.57	1.44	1.87	***<.001***
30	69.49	1.60	67.79	1.76	1.69	***<.001***
60	70.49	1.23	69.32	1.33	1.17	***.001***
120	71.47	0.87	70.46	0.89	1.00	***<.001***
300	72.43	0.28	71.59	0.43	0.84	***<.001***
avg	70.67		69.35		1.32	
var	1.67		2.96			

B.

8 objects

	BEL		NoBEL		BEL-NoBEL	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
25	46.46	0.56	46.50	1.16	-0.03	.913
30	48.53	0.58	48.43	1.19	0.10	.761
60	49.12	0.81	48.50	0.85	0.62	***.038***
120	51.17	1.36	50.27	1.05	0.91	***.010***
300	52.79	0.19	52.29	0.19	0.50	***.001***
avg	49.61		49.20		0.42	
var	5.96		4.77			

BEL: *Beliefs* architecture; NoBEL: *NoBeliefs* architecture

Within-architectures analysis

Tables 31 summarise performances across differing number of training maps for both the *Beliefs* and *NoBeliefs* architectures when considering a more complex vs the original, simpler environment. Compared to the original study with only 4 distractor objects (study 1), lower accuracies are generally observed in both the *Beliefs* and *NoBeliefs* architectures (see Figure 40 for a visualisation of this comparison for the *Beliefs* architecture). These results indicate worse predictive performance in more complex environments.

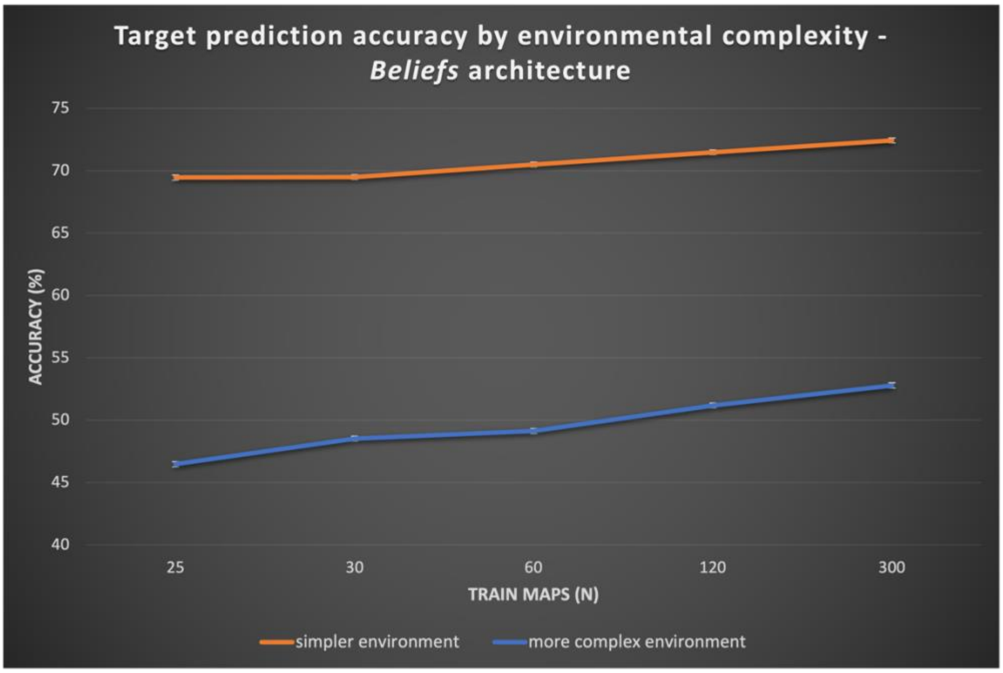


Figure 40. Example visualisation of target prediction accuracy by the *Beliefs* architecture by environmental complexity (simpler vs more complex environment). Average target prediction accuracies at each number of train maps are reported.

Between-architectures analysis

With regards to the role of beliefs for target prediction in the more complex environment, the above results indicate better performance (highly statistically significant in most cases) by the architecture including the Belief *head* (see Table 31). Specifically, performance improved by up to 0.91% (*p* = .010) (90 maps) when allowing belief processing. Compared to the original study 1, the impact of beliefs on target accuracy is lower (1.87% maximum impact vs 0.91%, respectively). Furthermore, beliefs processing does not seem to be significantly beneficial in this experiment when providing 25 and 60 train maps (see Figure 41). Overall, these results support findings from the original study 1, in that beliefs are beneficial for predicting others' intentions, and extends them to more complex environments, although to a lesser degree.

Developmental analysis

From a developmental perspective, a steady trend in both the architectures was seen for achieving better performance with an increasing number of maps made available during training. In accordance with the previous studies, these results suggest that better target prediction accuracies can be achieved with increasing experience even in more complex environments.

With regards to the impact of belief processing on target prediction, the trend identified in Study 1 was here also reported, although the peak of the impact of beliefs on performance improvement was shifted as well (see Figure 41). Specifically, the impact of beliefs processing on performance here peaked at 90 maps made available during training, as opposed to the 25 train maps seen in the original study, and decreased again from 120 to 300 maps.

Overall, these results indicate that the multi-task-induced regularisation hypothesis maintains with this study. However, in more complex environments, the role of beliefs becomes more evident with increasing experience and extends for a longer period of time. Nevertheless, these results indicate that the actual dynamics of the impact of learning explicit belief representations on target prediction depends on the complexity of the environment.

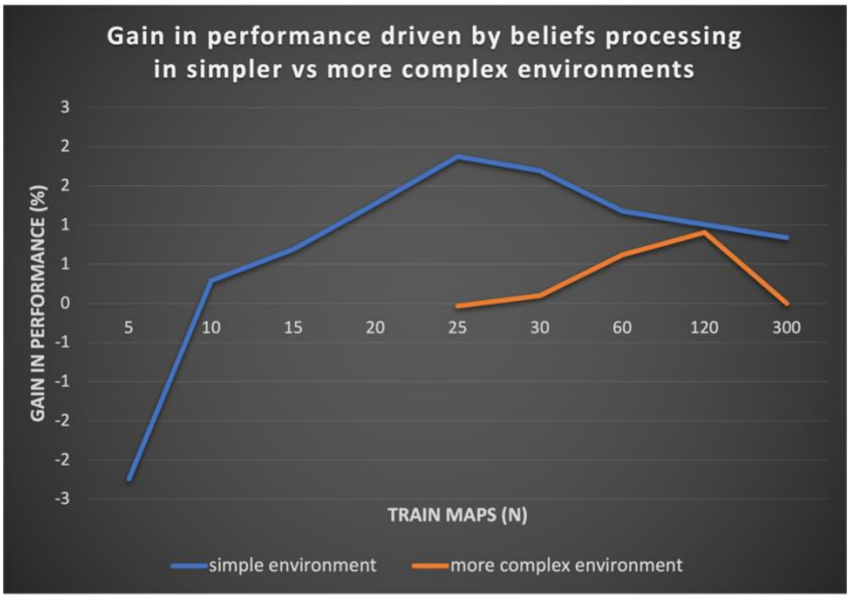


Figure 41. Delayed peak of impact of beliefs processing on target prediction accuracy in more complex vs simpler environments, by number of train maps.

Conclusions

- Overall, these studies provide evidence of the suitability of my architectural choices, supporting generalisation to other situations, including varying observers' and environmental complexities.
- While observers' characteristics do not extremely influence overall target prediction performances, environmental characteristics have a bigger impact.
- A significant impact of learning to explicitly represent beliefs towards predicting others' intentions and behaviours is visible in different observers and environments. However, this impact is lower in more complex observers and environments.
- From a developmental point of view, target prediction accuracy improves with experience both in more complex observers and environments. However, a delayed maximum impact of beliefs on performance is observed in more complex observers and environments. This indicates that the contribution of beliefs processing may extend for a longer period of time during development in these conditions.
- To conclude, these results confirm that the multi-task-induced regularisation hypothesis maintains across observers and that generalisation to observers and environments of different complexities is possible. However, the actual dynamics of the impact of learning to explicitly represent others' beliefs on intention prediction may be influenced by observers' and environmental complexity. Finally, these results show the suitability of this architectural choice for implementation in robotic systems.

3.4 Generalisation of the architecture

Finally, I explored in Study 4 the generalisation of the actual architecture by testing its ability to predict others' behaviours when observing actors with different (a) cognitive capabilities (i.e. varying max samples driving actor's behaviour), and (b) physical ability (i.e. varying speed for reaching target). The aim of this study was to show whether learning to explicitly represent others' beliefs can provide advantages for an observer towards predicting the behaviour of actors with new cognitive and physical abilities, with no (or little) experience with these behaviours. The investigation of this generalisation ability is important to understand the extent of the "like them" approach for understanding others who differ from the self. In these studies, I employed the same conditions used in the original studies, i.e. observer's complexity (2-layers neural network) and environmental complexity (4 objects in the environment).

Exp. 1: Varying actor's cognitive abilities

First, I conducted further testing on both the originally trained *Beliefs* and *NoBeliefs* architectures with simulations which saw actors with different cognitive capabilities (deliberation and planning as in Ognibene et al. (2019)). Specifically, the actors provided at testing differed in the amount of available cognitive resources determining the depth of navigation in their internal model of action-state relationships and associated rewards. Therefore, limited cognitive resources result in incorrect representation and action-sequence assessments, leading to suboptimal deliberation, planning and choices (Ognibene et al., 2019). While actors provided during training presented high cognitive capabilities (250 max samples), those at testing had lower cognitive capabilities, i.e. 150, 50, 25 and 5 max samples. This way I investigated whether learning to explicitly represent beliefs can also generalise to different actors with varying cognitive capabilities without the need to train on their specific behaviours or observe multiple behaviours as done in Rabinowitz et al. (2018). In other words, I assessed the extent to which the “like them” approach would be beneficial to understand actors whose behaviour highly differs from that produced by the observer without previous exposure to their behaviour or if a personal bias would hinder accurate predictions. All trained neural networks (i.e. from both configurations and trained with different numbers of maps and seeds) were tested for these new conditions. The results of these runs are summarised in Tables 32A-E below.

Tables 32. Target prediction accuracies for the *Beliefs* vs NoBeliefs architectures in the conditions with varying complexity of the actor's cognitive capabilities. Number of max samples underlying actor's behaviour: ***(A)*** *250* (Original); ***(B)*** *150;* ***(C)*** *50;* ***(D)*** *25;* ***(E)*** *5*. Accuracies were calculated as averages across 18 initial network weights; the associated variances were reported.

A.

250 max samples – *Original*

	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	59.26	18.84	61.51	20.30	-2.25	***.043***
10	65.75	8.99	65.46	11.82	0.29	.488
15	67.54	2.79	66.85	1.84	0.69	.077
20	68.74	2.08	67.47	1.51	1.27	***.001***
25	69.45	1.63	67.57	1.44	1.87	***<.001***
30	69.49	1.60	67.79	1.76	1.69	***<.001***
60	70.49	1.23	69.32	1.33	1.17	***.001***
120	71.47	0.87	70.46	0.89	1.00	***<.001***
300	72.43	0.28	71.59	0.43	0.84	***<.001***
avg	68.29		67.56		0.73	
var	15.45		8.67		1.48	

B.

150 max samples

	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	59.61	14.49	61.39	2.60	-1.78	.077
10	65.72	3.04	65.06	1.82	0.67	.208
15	67.61	2.96	67.18	2.03	0.43	.423
20	67.83	2.15	67.17	2.03	0.67	.175
25	68.61	2.25	67.44	2.73	1.17	***.033***
30	68.61	1.90	67.72	2.09	0.89	.068
60	69.50	1.09	69.22	1.24	0.28	.445
120	70.94	1.47	70.39	0.96	0.56	.139
300	71.50	0.85	70.72	0.68	0.78	.012
avg	67.77		67.37		0.41	
var	12.40		8.14		0.74	

C.

50 max samples						
	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	57.83	12.97	60.17	8.26	-2.33	***.039***
10	64.83	4.15	64.22	1.95	0.61	.301
15	66.06	2.64	66.24	1.32	-0.18	.709
20	67.78	2.18	66.50	2.62	1.28	***.019***
25	68.22	1.59	66.78	1.59	1.44	***.002***
30	68.44	2.14	67.28	1.51	1.17	***.014***
60	69.11	2.46	68.94	1.82	0.17	.735
120	70.11	1.28	69.61	0.96	0.50	.165
300	70.50	1.21	70.50	0.50	0.00	1.000
avg	66.99		66.69		0.29	
var	15.02		9.67		1.30	

D.

25 max samples						
	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	56.94	17.00	59.06	3.58	-2.11	.057
10	62.89	2.10	63.22	3.48	-0.33	.553
15	64.72	2.45	64.12	0.74	0.60	.165
20	65.78	0.77	65.06	2.06	0.72	.079
25	65.83	1.21	65.00	1.18	0.83	***.028***
30	66.56	0.85	65.83	1.21	0.72	***.040***
60	67.61	1.43	67.78	0.89	-0.17	.645
120	68.50	0.74	68.61	1.31	-0.11	.744
300	68.72	0.80	68.61	1.08	0.11	.733
avg	65.28		65.25		0.03	
var	13.17		9.15		0.84	

E.

5 max samples

	BEL		NoBEL			
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var	BEL-NoBEL (%)	*p-value*
5	47.667	29.06	50.39	10.49	-2.72	.075
10	56.00	4.47	55.39	3.66	0.61	.370
15	57.67	4.12	57.76	1.94	-0.10	.869
20	57.83	4.15	58.22	3.12	-0.39	.545
25	57.78	2.07	58.72	3.15	-0.94	.088
30	58.33	2.59	59.06	2.88	-0.72	.199
60	58.72	2.21	59.89	1.28	-1.17	***.012***
120	59.11	1.40	60.39	1.78	-1.28	***.005***
300	61.06	1.00	60.28	0.57	0.78	***.012***
avg	57.13		57.79		-0.66	
var	14.42		10.07		1.13	

Within-architectures analysis

Compared to the original results (Table 32A), averaged accuracies across maps generally revealed similar target prediction accuracies for actors with varying cognitive capabilities, although these were lower for actors who were increasingly different from the observer (e.g. 5 max samples condition) (see Figure 42 for a visualisation of this data). Overall, these results suggest an ability to generalise of these architectures to actors whose behaviour (driven by their cognitive capabilities) varies and differs from the self. Although lower generalisation is seen in conditions that are increasingly different from those the neural networks were trained upon.

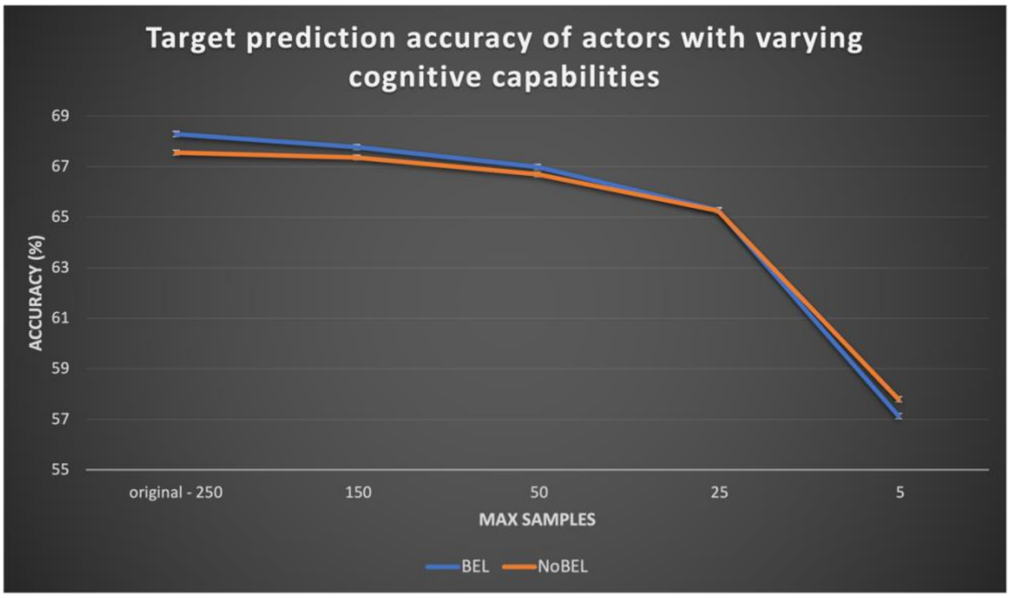


Figure 42. Target prediction accuracy by the *Beliefs* and *NoBeliefs* architectures bases on actors' cognitive capabilities (number of max samples). Average target prediction accuracies are reported.

Between-architectures analysis

The role of beliefs for target prediction remains evident in these results, although varying significance was observed. A lower gain in performance driven by beliefs processing can be globally reported when observing actors with lower cognitive capabilities. These results indicate once again successful multi-task-induced regularisation to be beneficial also for generalising predictive ability to actors whose behaviours (driven by cognitive capabilities) differ from the self. However, these results also suggest that such a gain in performance decreases with increasingly different actors.

Developmental analysis

From a developmental perspective, a steady trend in both the architectures was seen for achieving better performance with an increasing number of maps made available during training. In accordance with my previous studies, these results suggest that better target prediction accuracies can be achieved with increasing experience even when predicting the behaviour of actors with different cognitive capabilities.

With regards to the impact of belief processing on target prediction, I failed to identify a developmental trend. Specifically, while the impact of beliefs for actors with higher cognitive capabilities peaked at a medium level of experience (similarly to my previous studies), their significance varied across development, as well as in actors with lower cognitive capabilities.

Exp. 2: Varying actor's physical abilities

Finally, I further varied the complexity of the observed actor, however in this experiment changing the actor's physical abilities. Specifically, I tested both the originally trained architectures with simulations which saw actors moving around the grid world at different speeds throughout the task (i.e. at x0.75, x0.9, x1.1, and x1.25 the speed on which the nets were trained upon). The aim of this experiment was the same as the previous one, although from a different perspective. Specifically, I investigated whether the "like them" approach would be beneficial to understand actors whose behaviour highly differs from that produced by the observer or if bias would hinder accurate predictions. However, this time I considered actors' differing physical abilities (i.e. different speeds). This is another example of generalisation ability of my architecture. In a post-hoc analysis, I also exposed both originally trained architectures to a few behaviours of actors with such different physical abilities, through a brief re-training (5 maps). This was done in an attempt to determine if short exposure to their behaviour would enhance predictive performance. The results of these runs are summarised in Tables 33 below.

Tables 33. Target prediction accuracies for the *Beliefs* vs NoBeliefs architectures in the conditions with varying complexity of the actor's physical abilities. Speed underlying actor's behaviour: ***(A)*** *x0.75;* ***(B)*** *x0.90;* ***(C)*** *x1 (Original);* ***(D)*** *x1.1;* ***(E)*** *x1.25*. Left tables: *test* following original training. Right tables: *test* following re-training. Accuracies were calculated as averages across 18 initial network weights; the associated variances were reported.

A.

Speed: x*0.75*

	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	53.89	28.22	56.94	13.35	-3.06	.052
10	60.56	3.08	61.06	3.70	-0.50	.421
15	61.28	6.92	63.06	2.64	-1.78	***.021***
20	61.56	3.08	62.44	1.91	-0.89	.101
25	61.78	3.48	56.68	19.23	5.09	***.016***
30	62.44	3.91	63.22	2.54	-0.78	.202
60	63.50	3.21	63.89	1.87	-0.39	.469
120	64.50	2.15	64.72	0.80	-0.22	.586
300	66.11	0.81	65.56	0.50	0.56	***.047***
avg	61.73		61.95		-0.22	
var	11.70		10.14		5.02	

Re-training, Speed: x*0.75*

	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	63.92	1.19	64.00	0.68	-0.08	.759
10	64.98	0.59	64.69	0.60	0.29	.223
15	65.81	0.46	65.43	0.50	0.37	.105
20	66.08	0.48	65.81	0.81	0.27	.210
25	66.78	0.51	66.18	0.69	0.60	***.007***
30	67.12	0.60	66.36	0.77	0.76	***.004***
60	67.71	0.33	67.12	0.60	0.60	***.011***
120	68.16	0.42	67.96	0.62	0.19	.322
300	69.16	0.23	68.90	0.25	0.26	.087
avg	66.64		66.27		0.36	
var	2.65		2.38		0.06	

B.

Speed: x*0.9*

	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	56.39	21.78	58.94	7.94	-2.56	.055
10	62.89	1.99	63.39	2.96	-0.50	.347
15	63.94	2.64	64.56	2.50	-0.61	.261
20	65.06	1.82	64.78	2.65	0.28	.581
25	65.33	1.65	65.50	0.74	-0.17	.650
30	66.11	3.05	65.28	2.09	0.83	.128
60	66.83	1.32	66.00	1.18	0.83	***.032***
120	67.72	1.39	66.83	0.97	0.89	***.019***
300	69.11	0.69	68.28	1.04	0.83	***.010***
avg	64.82		64.84		-0.02	
var	13.56		6.84		1.27	

Re-training, Speed: x*0.9*

	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	64.47	0.93	64.44	1.34	0.03	.939
10	65.77	0.73	65.28	0.63	0.49	.085
15	66.90	0.68	66.32	0.92	0.58	***.020***
20	67.17	0.80	66.72	0.66	0.45	.064
25	67.98	0.97	67.26	1.22	0.73	***.018***
30	68.34	1.11	67.44	0.59	0.91	***.006***
60	68.90	0.45	68.24	0.48	0.66	***.005***
120	69.53	0.32	69.06	0.33	0.47	***.013***
300	70.30	0.22	70.25	0.36	0.06	.742
avg	67.71		67.22		0.48	
var	3.39		3.27		0.08	

C.

Speed: x*1 – Original*

	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	59.26	18.84	61.51	20.30	-2.25	***.043***
10	65.75	8.99	65.46	11.82	0.29	.488
15	67.54	2.79	66.85	1.84	0.69	.077
20	68.74	2.08	67.47	1.51	1.27	***.001***
25	69.45	1.63	67.57	1.44	1.87	***<.001***
30	69.49	1.60	67.79	1.76	1.69	***<.001***
60	70.49	1.23	69.32	1.33	1.17	***.001***
120	71.47	0.87	70.46	0.89	1.00	***<.001***
300	72.43	0.28	71.59	0.43	0.84	***<.001***
avg	68.29		67.56		0.73	
var	15.45		8.67		1.48	

D.

Speed: x*1.1*

	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	58.78	21.01	61.28	6.68	-2.50	.052
10	65.56	3.44	65.17	2.74	0.39	.551
15	67.33	1.65	67.06	1.94	0.28	.538
20	68.67	2.00	67.06	1.23	1.61	***.001***
25	69.56	2.03	67.56	1.44	2.00	***<.001***
30	69.50	1.79	67.94	1.11	1.56	***<.001***
60	70.56	1.67	68.94	1.82	1.61	***.001***
120	71.67	0.47	70.17	1.09	1.50	***<.001***
300	72.33	0.35	71.67	0.94	0.67	***.018***
avg	68.22		67.43		0.79	
var	16.86		8.91		1.90	

Re-training, Speed: x*1.1*

	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	64.82	1.09	64.83	1.12	-0.01	.978
10	66.30	0.99	65.83	0.55	0.48	.070
15	67.63	0.61	66.84	0.81	0.79	***.002***
20	68.08	0.68	67.21	0.54	0.87	***<.001***
25	68.74	1.08	67.69	1.46	1.05	***<.001***
30	69.13	0.83	67.89	0.56	1.24	***<.001***
60	69.89	0.75	68.87	0.70	1.02	***<.001***
120	70.53	0.42	69.84	0.38	0.69	***.002***
300	71.59	0.51	71.15	1.00	0.44	***.034***
avg	68.52		67.79		0.73	
var	4.42		3.82		0.15	

E.

Speed: *x1.25*

	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	59.11	17.75	61.00	5.29	-1.89	.104
10	65.22	3.59	64.72	2.80	0.50	.407
15	67.11	1.87	66.56	2.03	0.56	.241
20	68.11	2.58	66.56	1.44	1.56	***.002***
25	68.83	1.21	67.22	1.24	1.61	***<.001***
30	69.50	1.44	67.50	1.79	2.00	***<.001***
60	70.33	1.29	69.06	1.00	1.28	***.001***
120	71.33	1.29	69.94	1.23	1.39	***.001***
300	72.22	0.77	71.11	0.58	1.11	***<.001***
avg	67.98		67.07		0.90	
var	15.61		8.95		1.33	

Re-training, Speed: x*1.25*

	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	64.80	1.23	64.84	1.27	-0.04	.913
10	66.12	1.08	65.61	0.65	0.51	.108
15	67.23	0.82	66.98	0.86	0.25	.383
20	67.78	0.61	67.05	0.78	0.74	***.010***
25	68.46	0.96	67.39	1.18	1.06	***.001***
30	68.93	0.95	67.64	0.60	1.29	***<.001***
60	69.84	0.61	68.73	0.57	1.10	***<.001***
120	70.37	0.49	69.53	0.62	0.84	***<.001***
300	71.326	0.39	70.62	0.67	0.70	***.001***
avg	68.32		67.60		0.72	
var	4.33		3.31		0.18	

Within-architectures analysis

Compared to the original results (Table 33C), averaged accuracies at testing across maps generally revealed decreased accuracy with slower actors, while accuracies remained similar when predicting faster actors. These results may be driven by the fact that when observing slower actors, especially in the x0.75 condition, the actors are almost still for multiple steps, thus leading to less information and ultimately to lower performances. The opposite is valid with faster actors.

Furthermore, the same results were observed following a brief re-training which exposed the nets to a few of such different behaviours, although accuracies were more stable, also when predicting slower actors. See Figure 43 below for a visualisation of this data for the *Beliefs* architecture.

Overall, these results suggest an ability to generalise of these architectures to actors whose behaviour (driven by their speed during the task) varies and differs from self-experience. Although lower generalisation is seen in conditions that are increasingly different from personal experience, especially when interpreting slower actors.

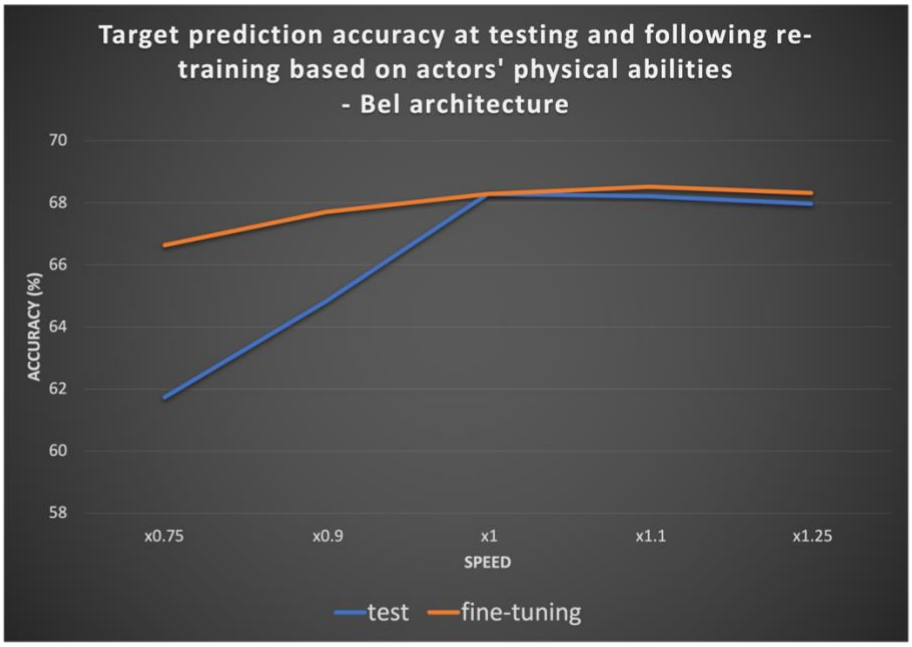


Figure 43. Example visualisation of target prediction accuracies achieved at test and following brief re-training by the *Beliefs* architecture based on actor's physical abilities (speed). Average target prediction accuracies are reported.

Between-architectures analysis

With regards to the impact of beliefs on target prediction, a similar interpretation to the above can be made. Specifically, when considering statistical significance, an advantage of beliefs processing towards target prediction was seen in most conditions. On average, the impact of beliefs was higher, as opposed to the original gain observed, when interpreting the behaviour of faster actors, while the opposite was valid for slower actors. However, when considering statistical significance, a big positive impact of beliefs was seen also when interpretating slower actors. Furthermore, following a brief adaptation to the newly observed behaviour, beliefs become beneficial also when predicting slower actors on average, stabilising this effect. See Figure 44 below for a visualisation of this data.

Overall, these results indicate once again successful multi-task-induced regularisation and generalisation to actors whose behaviours differ from the self in terms of physical ability (i.e. speed), especially after some short adaptation. This result is in concordance with my previous experiments suggesting the feasibility of the “like them” approach to beliefs towards the prediction of others’ behaviours. Furthermore, it highlights that, while personal bias does influence the ability to interpret others’ behaviours, this effect is only minimal and specific to certain conditions. Finally, these results indicate that this bias can be overcome by a short adaptation.

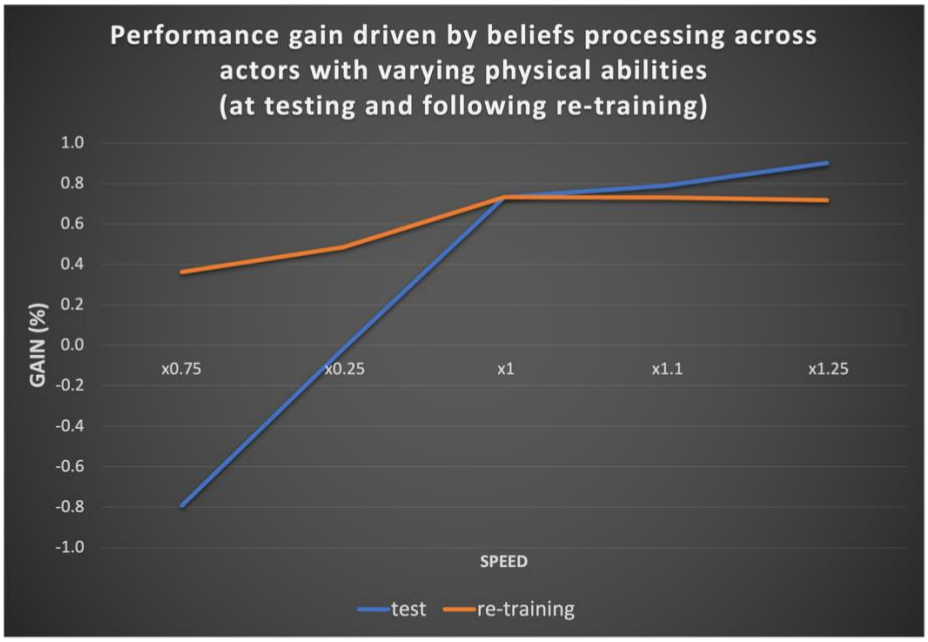


Figure 44. Gain in performance driven by beliefs processing based on actor's physical abilities capabilities (slower vs faster actors), both at testing and following brief re-training.

Developmental analysis

From a developmental perspective, a steady trend in both the architectures is seen for achieving better performance with an increasing number of maps made available during training. In accordance with my previous studies, these results suggest that better target prediction accuracies can be achieved with increasing experience even when predicting the behaviour of actors with different physical abilities.

With regards to the impact of belief processing on target prediction, this follows the same developmental (although less linear) trend previously identified, peaking at a medium level of experience, regardless of the actor's physical abilities. This is valid both at testing and following brief re-training.

Conclusions

- Overall, these studies provide evidence of the generalisation ability of my architecture when predicting behaviour of actors with differing cognitive and physical abilities from the self.
- The multi-task-induced regularisation effect was seen to be beneficial regardless of the actors' abilities, although the dynamics of the actual impact of learning to explicitly represent others' beliefs were found to be influenced by such factors. Specifically, the impact of beliefs processing on target prediction remained evident also when interpreting behaviours of actors who differ from the self (in terms of cognitive and physical abilities). However, this decreased when predicting actors with increasingly different cognitive and physical (mainly slower actors) abilities compared to the observer.
- To conclude, these findings suggest that the "like them" approach to beliefs, using multi-task-induced regularisation, is feasible and generally advantageous to interpret and predict behaviours and intentions of others who differ from the self. Furthermore, personal bias does not necessarily hinder performance, as seen through successful generalisation. However, actor-observer similarity is advised for improved predictive performance.

General discussion

In this series of studies, I investigated the “like them” approach to learning explicit beliefs representation for predicting others’ intentions and its developmental trajectory. This was done in an attempt to show the advantages of beliefs processing throughout development towards understanding others’ mental states. As a result, I identified through multi-task-induced regularisation a role for learning to explicitly represent beliefs towards predicting others’ intentions, generally developmental in nature. In addition, I further outlined specific conditions in which this regularisation between target and beliefs processing is more beneficial. These include situations in which the target is *not visible* by the actor (thus during exploratory behaviours) and when there is object neglect or target-object alignment. Indeed, my results indicated a relationship between beliefs processing, behaviour prediction, disambiguation of objects in the environment, and improved prediction of others’ intentions. Furthermore, I showed the generalisation of my architectural choices to different observers and environments with varying complexity, supporting this system implementation. Finally, I showed the generalisation ability of my architecture itself to actors with varying cognitive and physical abilities. These results indicated this approach to be advantageous for predicting the behaviour of others who differ from the self, without the need of extensive training on such different behaviours. Overall, these results outlined that taking a “like them” approach for beliefs is beneficial towards improved performance for predicting others’ intentions. Learning to explicitly represent beliefs does indeed increase the computational demands and complexity of a task. However, it ultimately helps others’ intentions prediction and beliefs-driven behaviour interpretation via self-observation-based training and its regularisation effect. This is valid when considering varying environments, observers, and actors.

Learning explicit beliefs representations

More in detail, this series of studies contributes to the literature by showing that including the learning of explicit belief representations is beneficial for improved performance in predicting others' intentions and beliefs-driven behaviours. Specifically, by using multi-task-induced regularisation between target and beliefs processing, I provided a viable approach to equip systems with a more accurate prediction of others' intentions given the same number of samples and faster learning. I deem this finding to be highly relevant for the AI literature considering that generally this approach is not taken when building systems for providing robots with a ToM or when creating computational models aimed at predicting others' mental states. Specifically, while previous studies implemented systems able to represent others' beliefs or predict others' beliefs-driven behaviours (e.g. Baker et al., 2017; Demiris & Khadhouri, 2006; Ramirez & Geffner, 2011), they generally did not include the learning of explicit beliefs representations. Rather, they focused on the final performance of intention prediction and assumed a learning free simulation-based approach. However, I here identified that learning explicit beliefs represents a valuable piece of information for the observer in several conditions and when interacting with different actors. Overall, the beneficial effect here observed of including the learning of explicit beliefs representation in ToM-related systems may intensify the focus in future artificial implementations on the learning of explicit mental states rather than sole predictive performance.

Only a few studies in the literature relied on the learning of explicit beliefs representations for predicting others' mental states (Breazeal et al., 2009; Kennedy et al., 2009; Rabinowitz et al., 2018). Of these, Rabinowitz et al. (2018) indicated this approach to be limited, due to (1) high computational demands connected with the complexity of beliefs representation and (2) the bias resulting from using own mental states as supervisory teaching signal to explicitly learn beliefs as accessed through meta-cognition. In this series of studies, I introduced the “like them” assumption, implemented through a deep learning architecture trained via self-observation, as a successful approach for addressing these points, while using a simple architecture.

Specifically, I showed that relying on self-representations of beliefs as supervisory signals for predicting others’ beliefs is computationally possible and beneficial from early development. Indeed, improved prediction of others’ mental states was driven by the multi-task-induced regularisation between target and beliefs processing, which resulted in faster learning with a lower number of training samples (i.e. more efficient learning). These results address concern (1) of Rabinowitz et al. (2018).

In addition, while I observed that this “like them” approach may lead to personal biases, I showed that these do not necessarily hinder performance (e.g. see experiments included in Part 3, chapter 3.3 of this thesis), as successful generalisation to actors differing from the observer was achieved (addressing concern (2) of Rabinowitz et al. (2018)). Furthermore, my results indicated that bias can be overcome by a short adaptation to the newly observed actor’s behaviour. These results support previous statements by Gilbert & Malone (1995) suggesting that biases are not always negative and can still be beneficial for understanding others’ behaviours even when the resulting prediction from this bias are not completely accurate. Nevertheless, in an attempt to reduce such biases and provide a system able to support more flexible forms of a shared representational framework assumption, the “like them” approach differs from the other two studies in the literature which implemented the learning of explicit beliefs representations for predicting others’ mental states (Breazeal et al., 2009; Kennedy et al., 2009). Indeed, these authors used simulation-inspired systems for learning explicit beliefs representations and for predicting others' beliefs, which could lead to increased biases due to the correspondence problem (Brass & Heyes, 2005; Nehaniv & Dautenhahn, 2002). Furthermore, given the limits of generative methods when applied to inputs that do not satisfy their assumptions, these systems may provide a limited solution for adaptive ToM. Therefore, I here provided a suitable approach, alternative to such implementations, using a simple architecture. From a human behaviour perspective, these results indicate that the "like them" approach can support ToM engagement also in cases of absent self-other similarity. Indeed, my findings suggested it to be a mechanism (operating on the "like me" shared representational framework) flexible enough to allow ToM engagement towards differing others, without the presence of personal biases hindering this cognitive ability. Overall, these results contribute to previous psychology literature debating early ToM acquisition, which mechanisms may underlie ToM, and if self-other similarity may be a pre-requisite of ToM ability and development (e.g. Frith & Frith, 2006b; Keysers & Gazzola, 2007; see also Parts 1-2 of this thesis).

Learning trajectory and impact of explicit beliefs representations on prediction of others' intentions

To my knowledge, this is the first study addressing the sample efficiency of learning to explicitly represent others' beliefs and the impact that this has on predictive performances. Indeed, no papers in the literature compare the learning trajectory of intention prediction models that explicitly learn to predict beliefs using their own mental states as teaching signal to that of prediction models that do not. As a result, I identified a general developmental trend of the impact of beliefs processing on prediction of others' intentions. The advantageous impact of multi-task-induced regularisation between target and beliefs processing started early during development and increased with experience. This then reached a plateau of maximum impact at a medium level of experience, which resulted in up to ~12% increase in predictive performance (see Exp. 6 above). Thereafter, the impact decreased, although still resulting in improved predictive performance. In addition, as mentioned earlier, I identified multi-task-induced regularisation between target and beliefs processing to result in faster learning with a lower number of training samples (i.e. more efficient learning). This is particularly important considering the use of deep convolutional networks that are known to be data hungry, the variability of social interactions, and the high ecological cost that may be associated with misinterpreting others' actions. Overall, these results support previous papers indicating multi-task-induced regularisation as an effective approach to reduce overfitting through shared representations and achieve faster learning in virtue of complementary information between related tasks (Crawshaw, 2020; Ruder, 2017). Furthermore, they extend it to the learning of beliefs for predicting others' intentions. Ultimately, from a developmental point of view on human behaviour, these results suggest that not only, computationally speaking, early belief representation is possible, but also that it may be beneficial for predicting others' intentions throughout development.

Beliefs for behaviour prediction and object disambiguation

As mentioned previously, beliefs processing positively impacted the prediction of others' intentions. My findings however further indicated that beliefs may be important also for discriminating between different types of behaviours, i.e. goal-directed vs exploratory behaviours. Indeed, while both the “ToM observer” and “simple observer” in my experiments were able to recognise several types of behaviours, a substantial impact of learning to explicitly represent beliefs was seen for the recognition and interpretation of explorative behaviours. This positive impact was seen with little experience / since early development. This suggests that beliefs processing can be used to aid the interpretation of others’ behaviours, especially when actors have partial knowledge about the environment and engage in exploratory behaviours, by significantly accelerating the learning process. This result has important implications as it shows that including learning of explicit beliefs representation in a system may be a better strategy than developing more complex architectures or creating bigger datasets for training models. This is especially resourceful for conditions in which specific data are not highly available, due to variability or volatility of the environments or agents (e.g. searching and rescuing during disasters or helping in construction sites).

These results are interesting from a human behaviour perspective as well. Indeed, considering that during exploratory conditions actors need to rely more strongly on their beliefs, my results highlight that, if an observer is aware of them (i.e. is endowed with a ToM), a substantial gain in performance may be obtained. In relation to this, my results may also support beliefs processing for successful false-belief tasks in infancy. Specifically, my findings indicated that architectures endowed with beliefs processing are faster at learning and predicting beliefs-driven behaviours and outperform architectures without this ability, given the same number of samples for training the model. My architectures (both the *Beliefs* and *NoBeliefs* architectures) were adapted from the ToMnet developed by Rabinowitz et al. (2018). While the "simple observer" is an adaptation of their ToMnet without beliefs processing, the "ToM observer" is an adaptation of their ToMnet with beliefs processing. Considering that their ToMnet (without beliefs processing) passed a false-belief task, and given the advantage of the *Beliefs* architecture in the above series of studies, we can assume that the "ToM observer" would also pass and outperform the "simple observer" in a false-belief task, by resulting in improved performance with less data. As this effect is evident from early development, outperformance in the false-beliefs task can be envisaged also early in infancy. Ultimately, these results contrast previous studies indicating that infants cannot rely on beliefs for passing false-belief studies due to an inability to process beliefs, given the complexity of developing this skill (Apperly & Butterfill, 2009; Heyes, 2014a, 2014b; Perner & Ruffman, 2005). On the contrary, my results support beliefs representation and reasoning in infants, also indicating that this ability may be associated with advantages towards success at false-belief tasks where agents have partial knowledge of the environment. Furthermore, they indicate that if beliefs processing ability is developed together with other abilities (e.g. others' target prediction), it results in more efficient learning and better understanding of others' minds (see multi-task-induced regularisation seen in this series of experiments). Therefore, there are no apparent reasons as to why, *computationally*, beliefs processing should not be supported in infancy.

Finally, my results indicated a relationship between beliefs processing and disambiguation of objects in the environment (as non-targets), mediated by discrimination of behaviour. Specifically, the above findings suggested that objects in the environment can be informative for an observer with respect to predicting others' intentions. However, the informative nature of such objects was seen to be influenced by the observer's recognition of goal-directed vs exploratory behaviours. More in detail, while objects close to the actor and aligned with the actor's target would negatively impact target prediction when a goal-directed behaviour was recognised by the architectures, objects were shown to be useful when an exploratory behaviour was instead considered. Crucially, only the architecture which included beliefs processing managed to exploit the informativity of aligned objects and better utilise that of other visible objects in the environment, by allowing the observer to better recognise exploratory behaviours. Ultimately, my findings indicated an advantage of architectures endowed with beliefs processing towards the discrimination of exploratory vs goal-directed behaviours, which in turn resulted in improved disambiguation of objects in the environment as non-targets, through a predicted object neglect strategy and reduced "target-close-to-actor" bias. Overall, this finding provides another reason for considering the implementation of learning explicit beliefs representation in an intelligent system for improved social skills, behaviour prediction, and interpretation of the informativity of the elements in the environment. In other words, beliefs processing allows an active discrimination between behaviours which in turn results in active disambiguation of objects in ambiguous environments and situations. To conclude, beliefs processing better accommodates for challenges seen in real-world scenarios; it would be interesting to validate these findings and the permanence of the identified effect with human data.

Limitations and future work

While these experiments resulted in several interesting insights, their limitations will now be outlined. First, as for the original study which inspired these experiments (Rabinowitz et al., 2018), the extension of these results to conditions of multi-agent interaction is not immediate. Indeed, the training signal for beliefs was here generated by self-training in solo interaction with the environment (Ognibene & Demiris, 2013; Ognibene et al., 2019). This limit related to the presence of and interaction with additional agents in the environment poses several other challenges to this model (and the original one), including (a) how does the brain simultaneously represent intentions and beliefs of multiple agents and (b) how this may differ from prediction of multi-inanimate object motion.

In addition, similarly to the original study (Rabinowitz et al., 2018), I made the strong assumption about the observer's complete and exact knowledge of the environment, which does not reflect real-life scenarios. Another assumption made is the ability of the observer to represent both self- and other-behaviours in an allocentric space, an ability which has its own developmental trajectory. Finally, objects differentiation was not considered, nor the impact that object appearance has on action prediction (e.g. Ambrosini et al., 2011; Rabinowitz et al., 2018).

Compared to Rabinowitz et al. (2018), I focused in this thesis on identifying the impact of learning explicit belief representation on the prediction of others' intentions; thus I limited the exploration of the architecture's generalization capabilities over diverse actors' populations. It would be interesting to study in the future if the performance gain here identified is preserved when observing other classes of agents, as in Rabinowitz et al. (2018). Furthermore, future studies investigating the extent to which beliefs are implicitly learnt by the observer, without explicit beliefs learning, are warranted. Nevertheless, the observed difference in performance and its evolution over time point to a limited ability to implicitly learn an accurate belief representation.

Future studies should explore the role that other forms of belief estimation, such simulation- or probabilistic-based estimations, could play in the developmental trajectory of these beliefs-aware social architectures. In addition, it would be important to investigate whether the performance gain driven by beliefs processing here identified would persist when belief representations are not learnt from the perspective of an optimal state observer, but rather from that of an autonomous observer such as the state of RL for POMDP (Ognibene & Baldassare, 2015), who may actually discard useful information. Finally, the ability to generalize over substantially diverse tasks, that is prevented by the simple setup of the grid world here utilised, should be examined.

Conclusions

Overall, the findings from this series of studies have implications on the ToM debate which directly interest (at least) two disciplines, i.e. artificial intelligence and psychology. On the one hand, I envisage that my approach of equipping robots with beliefs processing based on the "like them" assumption and using multi-task learning may extend their application to several scenarios. Indeed, this approach partly overcomes the high requirement of data of deep networks resulting in faster training with less samples and improves the prediction of others' intentions and behaviours, with flexibility and during social interactions of varying complexity. On the other hand, this series of studies contributes to the psychology debate on ToM emergence by indicating that beliefs processing is not only computationally possible and efficient from an early age, but also that it is beneficial for improved prediction of others' mental states and beliefs-driven behaviours. Furthermore, my findings suggest that this is possible through a shared representational framework between self and others. This result has implications in the interpretation of false-belief tasks results as well, supporting ToM competence in infancy. To conclude, my results are informative and useful both to further our understanding of human ToM and for implementing robotic architectures to improve robots' social skills and HRIs.

Conclusions

General discussion

The purpose of this section is twofold. It first proposes a unified discussion of the (developmental) psychology and computational modelling experimental contributions that were presented in this thesis. These are organised by sections representing my original research questions on ToM. Second, this section presents future research directions that the thesis may inspire.

Overview and contributions of the thesis

The main contribution of this thesis is the study of ToM using a mixed approach involving different disciplines and methodologies to achieve a more complete understanding of this cognition. On the one hand, this led to empirical psychology findings that contributed to our knowledge of ToM ability and development, and which can inspire future computational models and robotic architectures for improving HRI and robots' application to everyday scenarios. On the other hand, this approach led to an artificial system able to provide insights into the importance of ToM for behaviour and intention prediction from a developmental point of view, which can be validated in human studies and implemented in robotic systems. This thesis has thus addressed and contributed to research on ToM on several levels, which are summarised and discussed further below.

ToM emergence

It is debated in the developmental psychology literature when ToM emerges (Apperly & Butterfill, 2009; Heyes, 2014b; Onishi & Baillargeon, 2005; Surian et al., 2007; Yott & Poulin-Dubois, 2012). While several infant research studies have been conducted to address this question, contrasting results prevent conclusive remarks

(Kampis et al., 2020; Kulke et al., 2018; Kulke & Rakoczy, 2018; Powell et al., 2018; Southgate et al., 2007). In this thesis, I contributed to this debate from a new perspective. Specifically, I used computational modelling to determine the feasibility of learning and representing others' beliefs from an early age. As a result, my computational modelling studies supported early ToM ability by showing an advantage of having a ToM during observation of social interactions, which was developmental in nature. Specifically, my results suggested that there is an interplay between the impact of learning explicit belief representations on others' intentions prediction and development. My findings indicated that beliefs start playing an important role for understanding others' intentions and behaviours from early development. Overall, these results suggested that, in the computational setting selected, early beliefs representation is *not only computationally possible and efficient, but also advantageous for predicting others' actions, mental states and behaviours*.

The *developmental trend* identified indicated that the positive impact of beliefs processing increases with increasing experience, until reaching a plateau of maximum experience after which beliefs gradually become less useful for predicting others' intentions and behaviours, although still representing a good source of information. From a human development perspective, this may explain the variable performances seen throughout development in ToM-related tasks. Nonetheless, these computational results remain to be validated on human data. For example, a study could examine the age at which belief processing is more likely to impact intention and behaviour prediction, when this effect peaks and if it decreases with age. To note, the actual dynamics depend on the complexity of the environment, agents and observers; therefore, it would be interesting to investigate if these results persist when interpreting human development and in which conditions.

An additional takeaway from my computational studies with respect to ToM emergence is the fact that beliefs processing, and its early emergence, was shown to be most impactful for *exploratory behaviours*, where the observed actor has partial knowledge of the environment. This is particularly interesting for the debate surrounding ToM emergence as it could be a means to support infants' ability to understand others' false beliefs and thus predict their beliefs-driven behaviours in false-belief tasks. More in detail, false-belief tasks depict agents with partial knowledge of their environment, which results in them having a false belief. To be successful at these tasks, infants need to correctly interpret agents' beliefs-driven behaviours, even if contrasting infants' own perspective. My computational modelling results indicated faster learning of the recognition of such beliefs-driven behaviours in virtue of beliefs processing, supporting the idea that beliefs representation may aid false-belief tasks performance, earlier rather than later in development.

In this thesis, I was able to *support early ToM emergence* also with my experimental studies comparing ToM ability in individuals with congenital vs acquired limb difference vs controls. More in detail, individuals with limb difference reported enhanced ToM ability compared to the general population, an effect which was driven by the congenital limb difference group. Indeed, only individuals with congenital limb difference scored on average significantly higher than controls at several ToM tasks. These findings suggested the existence of a critical window during early development for achieving improved ToM in people with limb difference, an effect which however seem to persist throughout adulthood, thus showing some flexibility. The candidate mechanisms underlying this effect were discussed in Part 2, chapter 2.2 of this thesis and are briefly summarised below (in "*Mechanisms underlying ToM ability and development"* section). Interestingly, these results seem to partially be in line with my computational findings and the developmental trend outlined. Specifically, my findings in the limb difference population partly support early ToM emergence and its critical impact from an early age (congenital limb difference results), as well as their decreased impact with increasing age (acquired limb difference results), reflecting extensive experience in the artificial neural network. However, in contrast to the computational findings, my results from the limb difference population indicated that ToM development may be mostly critical at birth (congenital limb difference results), rather than at a medium level of experience (acquired limb difference results). Nonetheless, and as mentioned earlier, I cannot determine through these studies the developmental age corresponding to the artificial age implemented in the neural network. Future research is needed to shed light onto the matter.

Finally, while the above findings point to an early ToM development, my *infant study may not support* this stance. More in detail, I introduced in this thesis a new false-belief task for infants, in an attempt to investigate ToM ability in 18-month-olds. My findings may suggest an absence of ToM ability at such a young age, thus possibly contributing to the literature not supporting early ToM development. These results may partly be explained by my computational results. Indeed, while my computational studies highlighted ToM presence from an early age, it remains to be determined whether “early age” in the computational model corresponds to 18 months of human age or earlier or later development. Nevertheless, findings from my infant study did not necessarily provide evidence against early ToM emergence. Indeed, as indicated in my infant methodological contribution section (Part 1, chapter 2), solely relying on looking behaviour as a measure of infant ToM ability may be limited and may result in failed replications and contrasting findings (Kampis et al., 2020; Kulke et al., 2018; Kulke & Rakoczy, 2018; Powell et al., 2018; Southgate et al., 2007). Furthermore, given the outlined limitations of my study and the possible limitations of the new methodology implemented for this investigation, future studies with a larger population and with complementary measures of ToM are warranted to validate these results.

Mechanisms underlying ToM ability and development

It is debated in the literature which computational mechanisms may underlie ToM ability and development. Specifically, while the association (Csibra & Gergely, 2007), simulation (Rizzolatti et al., 2001) and teleological (Csibra & Gergely, 2007) mechanisms have been indicated to underlie (or be precursors of) ToM, a separate mentalising mechanism for ToM (Frith & Frith, 2003) has also been proposed. Neuroimaging studies have identified separate cortical areas to underlie such different mechanisms. For example, simulation has been associated with the activation of the mirror neuron system, which includes parietal and premotor cortical areas (de Lange et al., 2008); whereas the ToM network, including the medial prefrontal cortex, temporoparietal junction, superior temporal sulcus and the temporal poles (Amodio & Frith, 2006; Frith & Frith, 2003), has been associated with mentalising. Similarly, computational modelling generally relies on different implementations of ToM models inspired from such mechanisms. For example, Baker et al. (2017) developed a teleological-based model of ToM (see also Hamlin et al., 2013). Others used association- (Rabinowitz et al. 2018) or simulation-based (Breazeal et al., 2009; Kennedy et al., 2009) models of ToM. Nonetheless, neuroimaging studies have also evidenced an absence of such a clear separation between systems supporting each mechanism as seen through brain activity in response to ToM-related tasks. For example, Marsh et al. (2014) suggested that the mPFC and TPJ (brain areas associated with mentalising) may be associated with rationality resolution and mentalising about the reasons and intentions underlying an unusual behaviour (teleological mechanism). de Lange et al. (2008) evidenced the activation of both simulation- and mentalising-related brain areas for understanding action intentions. Furthermore, successful computational models of ToM including both simulation- and teleological-based principles have been developed (e.g. Asakura & Inui, 2016). Keysers and Gazzola (2007) instead brought forward the idea of a continuum between simulation and ToM. It has therefore been suggested that such mechanisms may overlap and coexist, and that they may collaborate for achieving ToM (Asakura & Inui, 2016; de Lange et al., 2008; Keysers & Gazzola, 2007).

In accordance with the cited articles supporting a "like me" approach to ToM, the computational and limb difference studies conducted in this thesis also supported the suitability of the "like me" approach to ToM. This approach can be considered more abstract than the above-mentioned mechanisms as it does not make specific assumptions on the type of mechanism using the implicated shared representational code; it thus is considered to engulf all such mechanisms (Meltzoff, 2007a). Specifically, my computational studies indicated that a simple architecture, based on the "like them" assumption, presents advantages towards predicting others' intentions and behaviours, withstanding differences between self and others. My limb difference studies also pointed to a shared representational framework between individuals with sensorimotor impairments and the general population, thus indicating the appropriateness of the "like me" approach for engaging in ToM, regardless of self-other similarity.

These findings, supporting the "like me" approach, therefore highlight that shared representations are at the hallmark of social cognition, rather than the result of complex developmental and inferential processes driven by an initially solitary representation of the self, as suggested in other contrasting theories (Piaget, 1952, 1954). These results could in turn also contribute to the debate in the literature surrounding the development of first- and third-person perspectives (representations of the world). Specifically, some studies indicated that infants first develop a first-person perspective which hinders false-beliefs understanding, as they fail to suppress it in favour of a third-person perspective. Others proposed that infants may have an altercentric view since birth and only later in development develop a first-person perspective (Happé, 2003; Kampis & Kovács, 2022; Southgate, 2020). My results supporting the "like me" assumption from early development for understanding others may point to the latter as a better alternative. This is true especially when considering my computational results, as they indicate that it is *computationally advantageous to have a shared representational framework for understanding others*. Nonetheless, future studies are warranted to further elucidate these mechanisms in line with studies by Kampis and Kovács (2022) and Yeung et al. (2022).

In addition, my studies contributed to previous research with respect to the identification of the candidate mechanisms using the "like me" approach for ToM. Indeed, in this thesis I discussed association, simulation and teleological for mentalising as candidate mechanisms underlying ToM ability in my studies. To briefly recapitulate my findings to this respect, my computational studies supported an *associative*-based mechanism for ToM (encompassing beliefs processing) by indicating an advantage in the introduction of the "like them" architectural approach towards predicting others' intentions and behaviours. Specifically, by relying on associations between explicit belief representations learnt through self-experience and consequent behaviours, the "like them" approach makes them available through self-observation as supervisory teaching signals for predicting others' behaviours.

Overall, this approach was shown to pose architectural demands that are simpler than previously believed, contribute to speeding-up the acquisition of socio-cognitive prediction skills, strongly improve the interpretation of beliefs-driven behaviours, and increase the generalisation ability to predict behaviours of others acting in different environments and presenting different cognitive and physical abilities.

On the contrary, I could not determine at this stage the exact mechanism underlying the enhanced ToM ability seen in the limb difference population; however, I proposed *simulation* and *teleological for mentalising* as alternative mechanisms, although not mutually exclusive, underlying such an effect. First, the simulation mechanism was justified given its relation to cognitive embodiment and the differing embodiments observed in the limb difference vs general populations. Furthermore, several measures used to assess ToM ability in individuals with limb difference, whose scores had been previously positively associated with the simulation mechanism (e.g. perspective taking, emotional reactivity), were found to be enhanced in individuals with limb difference compared to controls. (See next section "*Factors implicated in ToM*" for a discussion on perspective taking for ToM). While it may seem counterintuitive that individuals with sensorimotor impairment had enhanced simulation ability, my results were interpreted in light of previous research pointing to the presence of sensorimotor representations and associated simulation ability in individuals with limb difference as resulting from (a) in-utero spontaneous muscular activity and proprioceptive feedback (Price, 2006), (b) observation learning (Aziz-Zadeh et al., 2012; Brugger et al., 2000; Price, 2006), (c) motor imagery (Gandola et al., 2019; Malouin & Richards, 2010; Saruco et al., 2019). This additional process that people with limb difference require for engaging in simulation in order to interpret the world from the perspective of the majority of the population was proposed to lead to a relative strength, resulting in the observed enhanced ToM. Second, a teleological for mentalising mechanism was proposed to support ToM findings in the limb difference population, pointing to the non-motor, more deliberative component of ToM. Specifically, the teleological for mentalising mechanism suggests that mentalising can be seen as a rationalisation of behaviour for understanding others' unusual behaviours (Brass et al., 2007; Marsh et al., 2014). It was indicated in this thesis that this mechanism may result in enhanced ToM in individuals with limb difference considering that they engage more often in this rationalisation of behaviour when interacting with environments designed for typical embodiments and with the general population, given their differing physical characteristics. In other words, they are more trained in this non-motor and deliberative component of ToM. Future studies are warranted to shed further light into these mechanisms.

Given the above, my findings from the computational and limb difference populations may be considered as pointing to different mechanisms which may seem hard to interpret for a global understanding of the mechanisms underlying ToM development and ability. However, it should be taken into consideration that these studies involved different experimental environments and populations, which may have influenced the dynamics of the mechanisms underlying ToM ability. Specifically, my computational studies were developed in virtual and relatively simple environments, with the assumption that the observer had complete knowledge of the world. Therefore, my results may indicate that association may be a mechanism to engage in ToM when observing individuals in simple environments, with limited factors influencing others' behaviours. In contrast, the simulation or teleological for mentalising mechanisms may be more adequate for interactions in more complex environments, or more challenging scenarios in which the observer has partial knowledge of the world, e.g. during multi-agent interactions, as well as when considering other factors that may influence ToM (e.g. agent and observers' embodiments, perspective taking, mental rotation). In addition, while I suggested the mentioned mechanisms to underlie enhanced ToM in the limb difference population, I cannot exclude that these may be compensatory mechanisms for the lack of sensorimotor experiences in this population and that other mechanisms may underlie ToM in the general population. Nonetheless, previous studies suggested simulation and teleological reasoning as candidate mechanisms behind ToM, thus I do not believe this to be the case. To summarise, my findings do not necessarily contrast each other and can be interpreted as providing testable hypotheses to be addressed in future studies to identify whether different mechanisms can be supporting ToM in different scenarios.

In line with this view, my findings may be interpreted as supporting the *coexistence of these mechanisms and their collaboration for achieving successful ToM in different scenarios* (see discussion at the start of this section; but also Cook et al., 2010; de Lange et al., 2008; Heyes, 2010). To be more specific, I will now briefly present an interpretation which can be drawn by merging such findings for ToM. Please note that my aim is neither to support nor criticise the following interpretation, but only to put into perspective the contributions of these mixed findings to the literature and their yet valuable relevance for our discussion. Briefly, as mentioned previously, simulation is considered one of the candidate mechanisms underlying ToM (possibly supported also by my findings with the limb difference population). However, findings from my computational studies point to an association mechanism for ToM. Nonetheless, previous literature proposed that mirror neurons, which are assumed to be responsible for the simulation mechanism, are a by-product of associative learning deriving from sensorimotor experiences (Heyes, 2010). This proposal interestingly connects the association and simulation mechanisms which were proposed to underlie ToM in the computational and limb difference studies, respectively. Specifically, the coexistence of the two mechanisms may be suggested, with the association mechanism being at the heart of others' understanding, thus supporting my computational findings. According to Heyes (2010), sensorimotor associations may thereafter support the development of the mirror neuron system (and simulation mechanism); thus once considering an embodied agent (limb difference population) compared to an agent in a simulative environment (computational model). If we followed Heyes (2010)'s proposal and considered the sensorimotor impairment in the limb different population, we could speculate that such sensorimotor associations may be hindered in this population. As a result, simulation mechanisms would not be able to develop in the limb difference population, unless compensatory mechanisms came into play, such as sensorimotor representations resulting from observation learning or teleological for mentalising. Ultimately, this would support findings with the limb difference population as well. Future research addressing this interpretation is warranted to shed light into the associative or motor simulation components of mirror neurons for ToM, as well as the role of sensorimotor experience for this cognitive ability.

To conclude, the findings in this thesis contributed to the debate surrounding the mechanisms underlying ToM by suggesting that different mechanisms (including association, simulation, and teleological for mentalising) may coexist and collaborate to achieve successful ToM in different scenarios.

Factors implicated in ToM

With this thesis, I also contributed to the factors which may be implicated in ToM ability and development. Specifically, I addressed embodiment, multisensory integration, perspective taking, mental rotation, and self-other similarity.

First, I found evidence of *cognitive embodiment for ToM* through my studies on the limb difference population. More in detail, my experimental studies with the limb difference population indicated enhanced ToM ability in individuals with limb difference vs controls; an effect which was driven by the congenital limb difference subgroup. Overall, these results suggested that different sensorimotor experiences may impact ToM ability throughout development, and that the interaction between sensorimotor-driven embodiment and ToM may be crucial since birth (given the result with the congenital subgroup). These results support previous research (e.g. Chasiotis et al., 2006; Dyck et al., 2006; Leonard & Hill, 2014) highlighting a rich relationship between sensorimotor experiences and ToM (see also Parts 1-2 of this thesis). Indeed, either normal or reduced ToM ability has been previously indicated in individuals with other sensory impairments, such as visual (e.g. Anghel, 2012; Koster-Hale et al., 2014; Peterson et al., 2000) or hearing (e.g. Figueras-Costa & Harris, 2001; Marschark et al., 2019) impairments. Considering that my findings suggested sensorimotor impairment to result in enhanced ToM, they possibly highlighted a role for the motor component of such experiences towards ToM. While I am unable in this thesis to determine which is the mechanism behind this effect, I presented two alternative explanations, that are (a) enhanced motor component of such experiences for ToM (through compensatory mechanisms for simulation) or (b) enhanced non-motor, more deliberative component of ToM (see Part 2, chapter 3.1 of this thesis for more details). Nevertheless, future studies are warranted to shed some more light on the role of motor (and general sensorimotor) experiences for ToM. A final comment on the role of embodiment for ToM; as mentioned earlier, differing artificial and biological embodiments, in the form of simulation actors and limb difference population vs controls, may result in different mechanisms underlying ToM. It would be interesting to ultimately transfer my computational architectures to a robotic system, to also include the sensorimotor component in the architecture. This would allow to further shed light into (a) the role of sensorimotor experience for ToM, (b) the specific focus on the motor component of ToM, as well as (c) the impact of sensorimotor impairment for ToM. Indeed, robots allow to conduct 'impairment' studies in a way that is not possible in the human population, while maintaining their characteristic as embodied agents. Ultimately, these studies would further inform my findings and validate my interpretations.

Second, I investigated *multisensory integration* in infants for the development of ToM through an innovative multisensory false-belief task conducted with 18-month-olds. The task relied on two modalities (i.e. vision and touch) for false belief induction in the observed agent. As a result, this task critically required infants to integrate multisensory information from the agent's perspective in order to successfully infer the other agent's belief. My preliminary results indicated that infants may not have this ability, as they did not show longer looking when the agent had a false belief about the world. Considering that previous studies have identified ToM ability at this age (e.g. Moriguchi et al., 2018; Senju et al., 2011; Yott & Poulin-Dubois, 2012), these results may point to infants' inability to integrate multisensory information, which in turn hindered ToM. However, we cannot exclude a role of multisensory integration for beliefs understanding in 18-month-olds, given previous indirect evidence of this ability in infants (Forgács et al., 2019; Scott et al., 2015, 2017; Scott et al., 2010; Träuble et al., 2010). Ultimately, I could not determine in this thesis whether this study's findings were driven by (a) its online implementation and/or the newly introduced paradigm, (b) the methodology not relying on complimentary measures of ToM such as neuroimaging or computational modelling, or (c) an absence of infants' ability to integrate multisensory information for ToM. However, I believe to have shown ways to address this in future research and contributed to the psychological debate on ToM emergence and multisensory integration for ToM development in infants.

Third, I identified *perspective taking* as a factor influencing ToM ability and development, considering my findings from the limb difference population. More in detail, individuals with limb difference were found to score significantly higher than the general population in measures of perspective taking ability. This effect was driven by the congenital limb difference subgroup, indicating that this factor may be increasingly critical in early development for enhanced ToM. The same discussion provided above about the potential mechanisms underlying the enhanced effect seen in the limb difference population is valid for perspective taking. Furthermore, I advanced in this thesis the possibility of differences in self-other control between the limb difference and general population, which may result in enhanced perspective taking and ToM ability in the former (see Part 2, chapter 3.1 of this thesis). Ultimately, these findings indicated that perspective taking is a factor implicated in ToM ability and development, supporting previous research (e.g. Xie et al., 2018). Unfortunately, I was not able to achieve a sufficiently large sample size for my study directly addressing the role of perspective taking for false belief understanding in individuals with limb difference vs controls (Part 2, chapter 3.2). The same is valid for *mental rotation*, which I indicated in Part 2, chapter 1 of this thesis may also be a factor implicated in ToM ability. Future studies are warranted to shed light on this matter. Finally, I did not investigate perspective taking through computational modelling. However, it would be interesting to introduce this component in my computational studies to investigate whether perspective taking is essential to understand others' intentions (e.g. on the line of Chen et al. (2021)'s experiments), as well as the extent to which beliefs processing may impact this factor.

Fourth and last, I further assessed the requirement of *self-other similarity* for engaging in ToM. This factor has been previously discussed (see "*Mechanisms underlying ToM ability and development"* section); therefore, I will not digress on it here as well. Briefly, both my computational and experimental studies with the limb difference population pointed to a shared representational framework to understand others in relation to the self. Therefore, these findings may suggest that individuals need to share this "representational code" for understanding each other. Indeed, this is an ability that has been previously indicated to help the understanding of conspecifics (Decety & Chaminade, 2003) and not extend to individuals belonging to other species (e.g. Buccino et al., 2004). Such general shared representational framework may associate all individuals, while more stringent similarity between self and other may not be a requirement of ToM, as indicated by my studies. Indeed, the "ToM observer" in my computational studies was able to generalise its ability to predict others' behaviour also to actors with substantially different cognitive and physical capabilities. Similarly, my experimental studies with the limb difference population clearly pointed to humans' ability to understand and predict mental states and behaviours of others who differ from the self.

Future work

While future work associated with each study was already reported in each chapter and above, I will here briefly discuss the principal research directions that the main contributions of this thesis may inspire from a global perspective.

This thesis identified a role for sensorimotor-driven embodiment towards enhanced ToM ability and development, and discussed different underlying mechanisms. Future studies are warranted to shed further light onto this relationship, the role of cognitive embodiment for ToM, as well as ToM emergence and the underlying mechanisms. This research can take the following forms.

- First, additional experimental studies on the limb difference population should be conducted, possibly including neuroimaging methods, to investigate the activation of brain areas associated with simulation (parietal and premotor cortical areas) and/or teleological for mentalising (mPFC, TPJ) during the tasks here introduced (Aziz-Zadeh et al., 2012; Brass et al., 2007; Cusack et al., 2012; Marsh et al., 2014).
- Second, future studies should investigate the ability of individuals from the general population to engage in ToM when observing and interacting with others who differ from the self, e.g. individuals with limb difference. Indeed, previous studies (e.g. Aziz-Zadeh et al., 2012; Cusack et al., 2012) indicated different mechanisms (simulation vs teleological vs mentalising) in individuals with limb difference when understanding and imitating actions of others with vs without limb difference. However, it remains to be determined which mechanisms underlie mental state understanding and prediction in individuals from the general population towards others who show atypical bodily characteristics. This would increase our understanding on whether the

simulation / teleological for mentalising mechanisms are compensatory mechanisms specific to the impairment of sensorimotor experiences or mechanisms that are likely to be critical for ToM in the general population, too.

- Third, studies on children with limb difference could be performed to further address the hypothesis of perceptual narrowing resulting from sensorimotor experiences (e.g. see Kelly et al., 2007 for an example investigation of perceptual narrowing during infancy with respect to facial processing). This in turn could aid the understanding of the mechanisms underlying ToM ability and the developmental stages behind sensorimotor involvement in this cognition.
- Fourth, it would be interesting to compare performances of individuals with limb difference who are and are not prosthesis users in ToM tasks. This would indeed help determine the extent to which prosthesis use can impact the enhancing effect seen in this series of studies and shed further light on possible mechanisms involved, including self-other control and egocentric bias. Indeed, previous studies have identified prosthesis use to build representations (in individuals with congenital limb difference, e.g. Fritsch et al. 2021; Price, 2006) or preserve / adapt representations (in individuals with acquired limb difference, e.g. Guo et al., 2017; Mayer et al., 2008) of the missing limb. Thus, it would be interesting to investigate how prosthesis use affects the impact of sensorimotor-driven embodiment on ToM.
- Fifth, these experimental studies could be further addressed using a multidisciplinary approach and conducting computational modelling of this effect. For instance, future studies could add sensorimotor inputs to the here developed artificial architectures and assess how “impairing” the motor component may impact the prediction of others’ intentions. This would be

interesting to investigate in both the “ToM observer”, who can engage in beliefs processing, and the “simple observer” to compare differences in performance driven by beliefs processing. Additionally, the same approach could be taken with such artificial neural networks implemented in robotic systems. Robots represent artificial embodied agents with their own sensorimotor experiences, which are thus more relatable to human experiences. In line with developmental robotics (Sandini et al., 2021), robot as embodied agents represent a great tool to investigate the role of embodiment for the development of different cognitions. Indeed, they allow the investigation of such phenomena in real-world scenarios and through manipulations (e.g. “impairment” studies) that are not possible in human research.

Furthermore, this thesis supported through computational modelling an early ToM ability. My findings indicated that beliefs processing from an early age is not only computationally possible and efficient, but also beneficial for predicting others' intentions and behaviours in several scenarios. Future studies should validate such findings on human and robotics data.

- First, it would be interesting to identify whether the developmental trend seen in this thesis replicates in human participants and to determine the correlation between human age and the developmental steps found in the artificial implementation. While previous studies investigating beliefs processing in infants of various ages exist in the literature (e.g. Hamlin et al., 2013; Kampis et al., 2015; Luo, 2011; Moriguchi et al., 2018; Southgate & Vernetti, 2014; Surian et al., 2007; Träuble et al., 2010), this thesis provides a clear developmental trend which should be validated in human studies.

- Second, future research should explore if the multi-task-induced regularisation seen in my computational studies represents a benefit in humans as well. Multi-task learning has been previously suggested to be the most relatable type of learning to human learning, as humans are rarely presented with single tasks in isolation and they instead rely on information from different modalities to build their knowledge (Crawshaw, 2020). Therefore, the validation of the computational advantage resulting from this learning approach with human data would not be surprising.
- Third, it would be interesting to adapt the architecture here developed to actual robotic systems for the study of ToM emergence. Indeed, in line with the developmental robotics field (Sandini et al., 2021), robotic systems may lead to unexplored insights which can then be validated through human studies. The architecture here implemented was indeed quite simple and rather task-independent, allowing it to be easily adapted to actual robotic architectures (although different perceptual and motor conditions would need to be considered). I envisage this approach to extend the application of social robots to several scenarios, given their improved ability to understand humans' beliefs-driven behaviours adaptively.

Conclusion

A recurrent, yet still incredibly interesting, debate surrounds the development of human and machine ToM. In this thesis, I used a mixed approach involving different disciplines and methodologies to achieve a more complete understanding of this cognition. I specifically contributed to previous literature on ToM emergence, the mechanisms underlying this cognitive ability, as well as the factors that may be implicated in its development. Overall, I believe to have provided new insights into these topics by using this multidisciplinary approach. Ultimately, experimental findings in this thesis can be interpreted (and furthered in future investigations) to understand human cognition, the computational findings to inform artificial cognition, and the knowledge obtained from both disciplines to create increasingly social robotic systems and discover unexplored directions.

While this approach is valid in this particular context, I believe that it will also be beneficial if applied to other human cognitive abilities. In this time of advancements and innovations in various research fields, we should take advantage of the collaboration between sectors. Investigating the same cognitive ability from different perspectives and using different methodologies can achieve a more complete understanding of such cognition, improve our general knowledge and extend this to additional applicative scenarios in innovative ways.

References

Abate, K. H. (2013). Gender Disparity in Prevalence of Depression Among Patient Population: A Systematic Review. *Ethiopian Journal of Health Sciences*, *23*(3), 283–288.

Abeyasinghe, N. L., de Zoysa, P., Bandara, K. M. K. C., Bartholameuz, N. A., & Bandara, J. M. U. J. (2012). The prevalence of symptoms of Post-Traumatic Stress Disorder among soldiers with amputation of a limb or spinal injury: A report from a rehabilitation centre in Sri Lanka. *Psychology, Health & Medicine*, *17*(3), 376–381. https://doi.org/10.1080/13548506.2011.608805

Abu-Akel, A., & Shamay-Tsoory, S. (2011). Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia*, *49*(11), 2971–2984. https://doi.org/10.1016/j.neuropsychologia.2011.07.012

Abubshait, A., & Wiese, E. (2017). You Look Human, But Act Like a Machine: Agent Appearance and Behavior Modulate Different Aspects of Human–Robot Interaction. *Frontiers in Psychology*, *8*, 1393. https://doi.org/10.3389/fpsyg.2017.01393

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., & Damasio, A. R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *20*(7), 2683–2690.

Alami, R., Clodic, A., Montreuil, V., Sisbot, E., & Chatila, R. (2005). *Task planning for human-robot interaction*. 81–85. https://doi.org/10.1145/1107548.1107574

Albrecht, S. V., & Stone, P. (2018). Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence*, *258*, 66–95. https://doi.org/10.1016/j.artint.2018.01.002

Alterman, A. I., McDermott, P. A., Cacciola, J. S., & Rutherford, M. J. (2003). Latent structure of the Davis Interpersonal Reactivity Index in methadone maintenance patients. *Journal of Psychopathology and Behavioral Assessment*, *25*(4), 257–265. https://doi.org/10.1023/A:1025936213110

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

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews. Neuroscience*, *7*(4), 268–277. https://doi.org/10.1038/nrn1884

Anghel, D. (2012). The development of theory of mind in children with congenital visual impairments. *The Journal of Academic Librarianship*, *4*, 229.

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

Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., Ogino, M., & Yoshida, C. (2009). Cognitive Developmental Robotics: A Survey. *IEEE Transactions on Autonomous Mental Development*, *1*(1), 12–34. https://doi.org/10.1109/TAMD.2009.2021702

Asakura, N., & Inui, T. (2016). A Bayesian Framework for False Belief Reasoning in Children: A Rational Integration of Theory-Theory and Simulation Theory.

Frontiers in Psychology, *7*. https://www.frontiersin.org/article/10.3389/fpsyg.2016.02019

Ask, M., & Reza, M. (2016). Computational models in neuroscience: How real are they? A critical review of status and suggestions. *Austin Neurology & Neurosciences*, *1*(2), 1008.

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*(1), 48–53. https://doi.org/10.1111/j.1467-7687.2007.00563.x

Azhari, A., Truzzi, A., Neoh, M. J. Y., Balagtas, J. P. M., Tan, H. H., Goh, P. P., Ang, X. A., Setoh, P., Rigo, P., Bornstein, M. H., & Esposito, G. (2020). A decade of infant neuroimaging research: What have we learned and where are we going? *Infant Behavior & Development*, *58*, 101389. https://doi.org/10.1016/j.infbeh.2019.101389

Aziz-Zadeh, L., Sheng, T., Liew, S. L., & Damasio, H. (2012). Understanding Otherness: The Neural Bases of Action Comprehension and Pain Empathy in a Congenital Amputee. *Cerebral Cortex*, *22*(4), 811–819. https://doi.org/10.1093/cercor/bhr139

Bae, D. S., Canizares, M. F., Miller, P. E., Waters, P. M., & Goldfarb, C. A. (2018). Functional impact of congenital hand differences: Early results from the Congenital Upper Limb Differences (CoULD) Registry. *The Journal of Hand Surgery*, *43*(4), 321–330.

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

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 1–10. https://doi.org/10.1038/s41562-017-0064

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349. https://doi.org/10.1016/j.cognition.2009.07.005

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2021). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*, *26*(3), 295–314. https://doi.org/10.1037/met0000337

Balconi, M., & Bortolotti, A. (2013). The “simulation” of the facial expression of emotions in case of short and long stimulus duration. The effect of pre-motor cortex inhibition by rTMS. *Brain and Cognition*, *83*(1), 114–120. https://doi.org/10.1016/j.bandc.2013.07.003

Bar, M. (2009). Predictions: A universal principle in the operation of the human brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1181–1182. https://doi.org/10.1098/rstb.2008.0321

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or asperger syndrome. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *38*(7), 813–822. https://doi.org/10.1111/j.1469-7610.1997.tb01599.x

Baron-Cohen, S., Leslie, A. M., & 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

Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., & Wheelwright, S. (2003). The systemizing quotient: An investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1430), 361–374. https://doi.org/10.1098/rstb.2002.1206

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. https://doi.org/10.1023/b:jadd.0000022607.19833.00

Barone, P., & Gomila, A. (2021). Infants' performance in the indirect false belief tasks: A second-person interpretation. *Wiley Interdisciplinary Reviews. Cognitive Science*, *12*(3), e1551. https://doi.org/10.1002/wcs.1551

Baumard, J., & Osiurak, F. (2019). Is Bodily Experience an Epiphenomenon of Multisensory Integration and Cognition? *Frontiers in Human Neuroscience*, 316.

Bauminger-Zviely, N. (2013). False-Belief Task. In F. R. Volkmar (Ed.), *Encyclopedia of Autism Spectrum Disorders* (pp. 1249–1249). Springer. https://doi.org/10.1007/978-1-4419-1698-3_91

Beaumont, R. B., & Sofronoff, K. (2008). A new computerised advanced theory of mind measure for children with Asperger syndrome: The ATOMIC. *Journal of Autism and Developmental Disorders*, *38*(2), 249–260.

Bedny, M., Pascual-Leone, A., & Saxe, R. R. (2009). Growing up blind does not change the neural bases of Theory of Mind. *Proceedings of the National Academy of Sciences*, *106*(27), 11312–11317.

Bekrater-Bodmann, R., Schredl, M., Diers, M., Reinhard, I., Foell, J., Trojan, J., Fuchs, X., & Flor, H. (2015). Post-amputation pain is associated with the recall of an impaired body representation in dreams—Results from a nation-wide survey on limb amputees. *PLoS One*, *10*(3), e0119552.

Bhat, A. A., Mohan, V., Sandini, G., & Morasso, P. (2016). Humanoid infers Archimedes' principle: Understanding physical relations and object affordances through cumulative learning experiences. *Journal of The Royal Society Interface*, *13*(120), 20160310. https://doi.org/10.1098/rsif.2016.0310

Bianco, F., Lecce, S., & Banerjee, R. (2016). Conversations about mental states and theory of mind development during middle childhood: A training study. *Journal of Experimental Child Psychology*, *149*, 41–61. https://doi.org/10.1016/j.jecp.2015.11.006

Bianco, F., & Ognibene, D. (2019). *Transferring Adaptive Theory of Mind to Social Robots: Insights from Developmental Psychology to Robotics* (pp. 77–87). https://doi.org/10.1007/978-3-030-35888-4_8

Bianco, F., & Ognibene, D. (2020). From Psychological Intention Recognition Theories to Adaptive Theory of Mind for Robots: Computational Models. *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 136–138. https://doi.org/10.1145/3371382.3378364

Bicchi, A., & Tonietti, G. (2004). Fast and 'soft-arm' tactics. Dealing with the safety-performance tradeoff in robot arms design and control. *IEEE Robot. Autom. Mag.*, *11*, 1070–9932.

Blohm, G., Kording, K. P., & Schrater, P. R. (2020). A How-to-Model Guide for Neuroscience. *ENeuro*, *7*(1). https://doi.org/10.1523/ENEURO.0352-19.2019

Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, *77*(1), B25-31. https://doi.org/10.1016/s0010-0277(00)00096-2

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel'touch that eyes see. *Nature*, *391*(6669), 756–756.

Bradford, E. E., Jentzsch, I., & Gomez, J.-C. (2015). From self to social cognition: Theory of mind mechanisms and their relation to executive functioning. *Cognition*, *138*, 21–34.

Brandt, S., Buttelmann, D., Lieven, E., & Tomasello, M. (2016). Children's understanding of first- and third-person perspectives in complement clauses and false-belief tasks. *Journal of Experimental Child Psychology*, *151*. https://doi.org/10.1016/j.jecp.2016.03.004

Brass, M., & Heyes, C. (2005). Imitation: Is cognitive neuroscience solving the correspondence problem? *Trends in Cognitive Sciences*, *9*, 489–495. https://doi.org/10.1016/j.tics.2005.08.007

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: Inferential processes versus action simulation. *Current Biology: CB*, *17*(24), 2117–2121. https://doi.org/10.1016/j.cub.2007.11.057

Breazeal, C., Gray, J., & Berlin, M. (2009). An Embodied Cognition Approach to Mindreading Skills for Socially Intelligent Robots. *The International Journal of Robotics Research*, *28*(5), 656–680. https://doi.org/10.1177/0278364909102796

Bremner, A. J., Lewkowicz, D. J., & Spence, C. (2012). *The multisensory approach to development.*

Brooks, R., & Meltzoff, A. N. (2015). Connecting the dots from infancy to childhood: A longitudinal study connecting gaze following, language, and explicit theory of mind. *Journal of Experimental Child Psychology*, *130*, 67–78. https://doi.org/10.1016/j.jecp.2014.09.010

Brugger, P., Kollias, S. S., Müri, R. M., Crelier, G., Hepp-Reymond, M.-C., & Regard, M. (2000). Beyond re-membering: Phantom sensations of congenitally absent limbs. *Proceedings of the National Academy of Sciences*, *97*(11), 6167–6172.

Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., Porro, C. A., & Rizzolatti, G. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: An FMRI study. *Journal of Cognitive Neuroscience*, *16*(1), 114–126. https://doi.org/10.1162/089892904322755601

Butterfill, S. A., & Apperly, I. A. (2013). How to Construct a Minimal Theory of Mind. *Mind & Language*, *28*(5), 606–637. https://doi.org/10.1111/mila.12036

Butterworth, G., & Grover, L. (1989). Social cognition in infancy: Joint visual attention, manual pointing and the origins of referential communication. *Revue Internationale de Psychologie Sociale*, *2*(1), 9–22.

Cangelosi, A., & Schlesinger, M. (2018). From Babies to Robots: The Contribution of Developmental Robotics to Developmental Psychology. *Child Development Perspectives*, *12*(3), 183–188. https://doi.org/10.1111/cdep.12282

Carr, A., Slade, L., Yuill, N., Sullivan, S., & Ruffman, T. (2018). *Minding the children: A longitudinal study of mental state talk, theory of mind and behavioural adjustment from age 3 to age 10.* https://doi.org/10.1111/SODE.12315

Caruana, R. (1998). Multitask Learning. In S. Thrun & L. Pratt (Eds.), *Learning to Learn* (pp. 95–133). Springer US. https://doi.org/10.1007/978-1-4615-5529-2_5

Cavallo, F., Esposito, R., Limosani, R., Manzi, A., Bevilacqua, R., Felici, E., Nuovo, A. D., Cangelosi, A., Lattanzio, F., & Dario, P. (2018). Robotic Services Acceptance in Smart Environments With Older Adults: User Satisfaction and Acceptability Study. *Journal of Medical Internet Research*, *20*(9), e9460. https://doi.org/10.2196/jmir.9460

Chan, A. W.-Y., Bilger, E., Griffin, S., Elkis, V., Weeks, S., Hussey-Anderson, L., Pasquina, P. F., Tsao, J. W., & Baker, C. I. (2019). Visual responsiveness in sensorimotor cortex is increased following amputation and reduced after mirror therapy. *NeuroImage: Clinical*, *23*, 101882. https://doi.org/10.1016/j.nicl.2019.101882

Chasiotis, A., Kiessling, Winter, & Hofer, V. (2006). Sensory motor inhibition as a prerequisite for theory-of-mind: A comparison of clinical and normal preschoolers differing in sensory motor abilities. *International Journal of Behavioral Development*, *30*, 178–190. https://doi.org/10.1177/0165025406063637

Chen, B., Vondrick, C., & Lipson, H. (2021). Visual behavior modelling for robotic theory of mind. *Scientific Reports*, *11*(1), 424. https://doi.org/10.1038/s41598-020-77918-x

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A. (2015). Radical Predictive Processing. *The Southern Journal of Philosophy*, *53*(S1), 3–27. https://doi.org/10.1111/sjp.12120

Cliffordson, C. (2001). Parents' judgments and students' self-judgments of empathy: The structure of empathy and agreement of judgments based on the interpersonal reactivity index (IRI). *European Journal of Psychological Assessment*, *17*(1), 36–47. https://doi.org/10.1027/1015-5759.17.1.36

Conson, M., Mazzarella, E., Esposito, D., Grossi, D., Marino, N., Massagli, A., & Frolli, A. (2015). "Put Myself Into Your Place": Embodied Simulation and Perspective Taking in Autism Spectrum Disorders. *Autism Research*, *8*(4), 454–466. https://doi.org/10.1002/aur.1460

Cook, R., Press, C., Dickinson, A., & Heyes, C. (2010). Acquisition of automatic imitation is sensitive to sensorimotor contingency. *Journal of Experimental Psychology. Human Perception and Performance*, *36*(4), 840–852. https://doi.org/10.1037/a0019256

Corradi-Dell'Acqua, C., & Tessari, A. (2010). Is the body in the eye of the beholder?: Visual processing of bodies in individuals with anomalous anatomical sensory and motor features. *Neuropsychologia*, *48*(3), 689–702.

Cowan, N. (1998). Five Enigmas Regarding LaBerge's (1997) Triangular-Circuit Theory of Attention and Self-Referential Theory of Awareness. *Psyche*, *4*, 9.

Crawshaw, M. (2020). Multi-Task Learning with Deep Neural Networks: A Survey. *ArXiv:2009.09796 [Cs, Stat]*. http://arxiv.org/abs/2009.09796

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

Csibra, G., & Gergely, G. (2007). ‘Obsessed with goals’: Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, *124*(1), 60–78. https://doi.org/10.1016/j.actpsy.2006.09.007

Cusack, W. F., Cope, M., Nathanson, S., Pirouz, N., Kistenberg, R., & Wheaton, L. (2012). Neural activation differences in amputees during imitation of intact versus amputee movements. *Frontiers in Human Neuroscience*, *6*, 182.

Datavyu Team. (2014). *Datavyu: A Video Coding Tool. Databrary Project*. http://datavyu.org

Datta, D., Selvarajah, K., & Davey, N. (2004). Functional outcome of patients with proximal upper limb deficiency–acquired and congenital. *Clinical Rehabilitation*, *18*(2), 172–177. https://doi.org/10.1191/0269215504cr716oa

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85.

De Beni, R., Pazzaglia, F., & Gardini, S. (2006). The role of mental rotation and age in spatial perspective-taking tasks: When age does not impair perspective-taking performance. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(6), 807–821.

De Bruin, L. C., & Newen, A. (2012). An association account of false belief understanding. *Cognition*, *123*(2), 240–259. https://doi.org/10.1016/j.cognition.2011.12.016

de Klerk, C. C. J. M., Southgate, V., & Csibra, G. (2016). Predictive action tracking without motor experience in 8-month-old infants. *Brain and Cognition*, *109*, 131–139. https://doi.org/10.1016/j.bandc.2016.09.010

de Lange, F. P., Spronk, M., Willems, R. M., Toni, I., & Bekkering, H. (2008). Complementary systems for understanding action intentions. *Current Biology: CB*, *18*(6), 454–457. https://doi.org/10.1016/j.cub.2008.02.057

De Pisapia, N., Barchiesi, G., Jovicich, J., & Cattaneo, L. (2019). The role of medial prefrontal cortex in processing emotional self-referential information: A combined TMS/fMRI study. *Brain Imaging and Behavior*, *13*(3), 603–614. https://doi.org/10.1007/s11682-018-9867-3

Decety, J., & Chaminade, T. (2003). When the self represents the other: A new cognitive neuroscience view on psychological identification. *Consciousness and Cognition*, *12*(4), 577–596. https://doi.org/10.1016/S1053-8100(03)00076-X

Demetriou, C., Ozer, B. U., & Essau, C. A. (2015). Self-Report Questionnaires. In *The Encyclopedia of Clinical Psychology* (pp. 1–6). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118625392.wbecp507

Demiris, Y. (2007). Prediction of intent in robotics and multi-agent systems. *Cognitive Processing*, *8*(3), 151–158. https://doi.org/10.1007/s10339-007-0168-9

Demiris, Y., & Khadhouri, B. (2006). Hierarchical attentive multiple models for execution and recognition of actions. *Robotics and Autonomous Systems*, *54*(5), 361–369. https://doi.org/10.1016/j.robot.2006.02.003

den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*, 548. https://doi.org/10.3389/fpsyg.2012.00548

Devaine, M., Hollard, G., & Daunizeau, J. (2014). The social Bayesian brain: Does mentalizing make a difference when we learn? *PLoS Computational Biology*, *10*(12), e1003992.

Devin, S., & Alami, R. (2016). An Implemented Theory of Mind to Improve Human-Robot Shared Plans Execution. *The Eleventh ACM/IEEE International Conference on Human Robot Interation*, 319–326. https://doi.org/10.1109/HRI.2016.7451768

Devine, R. T., & Hughes, C. (2016). Measuring theory of mind across middle childhood: Reliability and validity of the Silent Films and Strange Stories tasks. *Journal of Experimental Child Psychology*, *149*, 23–40. https://doi.org/10.1016/j.jecp.2015.07.011

Devine, R. T., White, N., Ensor, R., & Hughes, C. (2016). Theory of mind in middle childhood: Longitudinal associations with executive function and social competence. *Developmental Psychology*, *52*(5), 758–771. https://doi.org/10.1037/dev0000105

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020). Shall I Trust You? From Child–Robot Interaction to Trusting Relationships. *Frontiers in Psychology*, *11*, 469. https://doi.org/10.3389/fpsyg.2020.00469

Diaconescu, A. O., Mathys, C., Weber, L. A., Daunizeau, J., Kasper, L., Lomakina, E. I., Fehr, E., & Stephan, K. E. (2014). Inferring on the intentions of others by hierarchical Bayesian learning. *PLoS Computational Biology*, *10*(9), e1003810.

Dick, F., Lloyd-Fox, S., Blasi, A., Elwell, C., Mills, D., & Elwell, C. (2014). Neuroimaging methods. *Educational Neuroscience*, 13–45.

Dionne-Dostie, E., Paquette, N., Lassonde, M., & Gallagher, A. (2015). Multisensory integration and child neurodevelopment. *Brain Sciences*, *5*(1), 32–57.

Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Research*, *1242*, 136–150.

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

Dyck, M. J., Piek, J. P., Hay, D., Smith, L., & Hallmayer, J. (2006). Are abilities abnormally interdependent in children with autism? *Journal of Clinical Child and Adolescent Psychology*, *35*(1), 20–33.

Elkholi, S. M. A., Abdelwahab, M. K., & Abdelhafeez, M. (2021). Impact of the smell loss on the quality of life and adopted coping strategies in COVID-19 patients. *European Archives of Oto-Rhino-Laryngology*, *278*(9), 3307–3314.

Emberson, L. L., Boldin, A., Riccio, J. E., Guillet, R., & Aslin, R. N. (2017). Deficits in Top-Down Sensory Prediction in Infants At-Risk Due to Premature Birth. *Current Biology : CB*, *27*(3), 431–436. https://doi.org/10.1016/j.cub.2016.12.028

Ephraim, P. L., Dillingham, T. R., Sector, M., Pezzin, L. E., & MacKenzie, E. J. (2003). Epidemiology of limb loss and congenital limb deficiency: A review of the literature. *Archives of Physical Medicine and Rehabilitation*, *84*(5), 747–761.

Erle, T. M., & Topolinski, S. (2015). Spatial and empathic perspective-taking correlate on a dispositional level. *Social Cognition*, *33*(3), 187–210.

Erle, T. M., & Topolinski, S. (2017). The grounded nature of psychological perspective-taking. *Journal of Personality and Social Psychology*, *112*(5), 683.

Figueras-Costa, B., & Harris, P. (2001). Theory of mind development in deaf children: A nonverbal test of false-belief understanding. *Journal of Deaf Studies and Deaf Education*, *6*(2), 92–102. https://doi.org/10.1093/deafed/6.2.92

Forgács, B., Parise, E., Csibra, G., Gergely, G., Jacquey, L., & Gervain, J. (2019). Fourteen-month-old infants track the language comprehension of communicative partners. *Developmental Science*, *22*(2), e12751. https://doi.org/10.1111/desc.12751

Fraser, C. M. (1998). Laterality, gender and age differences in estimated frequency and actual registration of people with congenital upper limb absences. *Prosthetics and Orthotics International*, *22*(3), 224–229.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2018). Does predictive coding have a future? *Nature Neuroscience*, *21*(8), 1019–1021. https://doi.org/10.1038/s41593-018-0200-7

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, *6*(4), 187–214. https://doi.org/10.1080/17588928.2015.1020053

Frith, C. D., & Frith, U. (1999). Interacting Minds—A Biological Basis. *Science*, *286*(5445), 1692–1695. https://doi.org/10.1126/science.286.5445.1692

Frith, C. D., & Frith, U. (2006a). How we predict what other people are going to do. *Brain Research*, *1079*(1), 36–46. https://doi.org/10.1016/j.brainres.2005.12.126

Frith, C. D., & Frith, U. (2006b). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

Frith, C., & Frith, U. (2005). Theory of mind. *Current Biology*, *15*(17), R644–R645.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1431), 459–473. https://doi.org/10.1098/rstb.2002.1218

Fritsch, A., Lenggenhager, B., & Bekrater-Bodmann, R. (2021). Prosthesis embodiment and attenuation of prosthetic touch in upper limb amputees—A proof-of-concept study. *Consciousness and Cognition*, *88*, 103073. https://doi.org/10.1016/j.concog.2020.103073

Funk, M., Shiffrar, M., & Brugger, P. (2005). Hand movement observation by individuals born without hands: Phantom limb experience constrains visual limb perception. *Experimental Brain Research*, *164*(3), 341–346.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493–501. https://doi.org/10.1016/S1364-6613(98)01262-5

Gandola, M., Zapparoli, L., Saetta, G., De Santis, A., Zerbi, A., Banfi, G., Sansone, V., Bruno, M., & Paulesu, E. (2019). Thumbs up: Imagined hand movements counteract the adverse effects of post-surgical hand immobilization. Clinical, behavioral, and fMRI longitudinal observations. *NeuroImage: Clinical*, *23*, 101838. https://doi.org/10.1016/j.nicl.2019.101838

Garfield, J. L., Peterson, C. C., & Perry, T. (2001). Social Cognition, Language Acquisition and The Development of the Theory of Mind. *Mind & Language*, *16*(5), 494–541. https://doi.org/10.1111/1468-0017.00180

Gergely, G., & Csibra, G. (1997). Teleological reasoning in infancy: The infant's naive theory of rational action. A reply to Premack and Premack. *Cognition*, *63*(2), 227–233. https://doi.org/10.1016/s0010-0277(97)00004-8

Giese, M. A., & Rizzolatti, G. (2015). Neural and Computational Mechanisms of Action Processing: Interaction between Visual and Motor Representations. *Neuron*, *88*(1), 167–180. https://doi.org/10.1016/j.neuron.2015.09.040

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*(1), 21–38. https://doi.org/10.1037/0033-2909.117.1.21

Goldman, A. I. (2012, January 18). *Theory of Mind*. The Oxford Handbook of Philosophy of Cognitive Science. https://doi.org/10.1093/oxfordhb/9780195309799.013.0017

Gordon, R. M. (2021). Simulation, Predictive Coding, and the Shared World. In M. Gilead & K. N. Ochsner (Eds.), *The Neural Basis of Mentalizing* (pp. 237–255). Springer International Publishing. https://doi.org/10.1007/978-3-030-51890-5_12

Görür, O., Rosman, B., Hoffman, G., & Albayrak, S. (2017, March 6). *Toward Integrating Theory of Mind into Adaptive Decision- Making of Social Robots to Understand Human Intention*.

Groen, Y., Fuermaier, A. B. M., Den Heijer, A. E., Tucha, O., & Althaus, M. (2015). The Empathy and Systemizing Quotient: The Psychometric Properties of the Dutch Version and a Review of the Cross-Cultural Stability. *Journal of Autism and Developmental Disorders*, *45*(9), 2848–2864. https://doi.org/10.1007/s10803-015-2448-z

Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2017). Implicit and explicit false belief development in preschool children. *Developmental Science*, *20*(5). https://doi.org/10.1111/desc.12445

Guo, X., Lin, Z., Lyu, Y., Bekrater-Bodmann, R., Flor, H., & Tong, S. (2017). The effect of prosthesis use on hand mental rotation after unilateral upper-limb amputation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(11), 2046–2053.

Haas, B. W., Anderson, I. W., & Filkowski, M. M. (2015). Interpersonal reactivity and the attribution of emotional reactions. *Emotion (Washington, D.C.)*, *15*(3), 390–398. https://doi.org/10.1037/emo0000053

Hamlin, J. K., Ullman, T., Tenenbaum, J., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science*, *16*(2), 209–226. https://doi.org/10.1111/desc.12017

Happé, F. (2003). Theory of mind and the self. *Annals of the New York Academy of Sciences*, *1001*(1), 134–144.

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

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-Inspired Artificial Intelligence. *Neuron*, *95*(2), 245–258. https://doi.org/10.1016/j.neuron.2017.06.011

He, H., Boyd-Graber, J., Kwok, K., & Daumé III, H. (2016). Opponent Modeling in Deep Reinforcement Learning. *ArXiv:1609.05559 [Cs]*. http://arxiv.org/abs/1609.05559

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

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, *32*(2), 175–191.

Heikkinen, M., Saarinen, J., Suominen, V. P., Virkkunen, J., & Salenius, J. (2007). Lower limb amputations: Differences between the genders and long-term survival. *Prosthetics and Orthotics International*, *31*(3), 277–286.

Heims, H. C., Critchley, H. D., Dolan, R., Mathias, C. J., & Cipolotti, L. (2004). Social and motivational functioning is not critically dependent on feedback of autonomic responses: Neuropsychological evidence from patients with pure autonomic failure. *Neuropsychologia*, *42*(14), 1979–1988. https://doi.org/10.1016/j.neuropsychologia.2004.06.001

Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews*, *34*(4), 575–583. https://doi.org/10.1016/j.neubiorev.2009.11.007

Heyes, C. (2014a). Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *9*(2), 131–143. https://doi.org/10.1177/1745691613518076

Heyes, C. (2014b). False belief in infancy: A fresh look. *Developmental Science*, *17*(5), 647–659. https://doi.org/10.1111/desc.12148

Heyes, C., & Catmur, C. (2022). What Happened to Mirror Neurons? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *17*(1), 153–168. https://doi.org/10.1177/1745691621990638

Hiatt, L., Harrison, A., & Trafton, J. (2011). *Accommodating Human Variability in Human-Robot Teams through Theory of Mind.* 2066–2071. https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-345

Hirai, M., Muramatsu, Y., Mizuno, S., Kurahashi, N., Kurahashi, H., & Nakamura, M. (2013). Developmental changes in mental rotation ability and visual perspective-taking in children and adults with Williams syndrome. *Frontiers in Human Neuroscience*, *7*, 856.

Hirsh, A., Dillworth, T., Ehde, D., & Jensen, M. (2009). Sex Differences in Pain and Psychological Functioning in Persons With Limb Loss. *The Journal of Pain : Official Journal of the American Pain Society*, *11*, 79–86. https://doi.org/10.1016/j.jpain.2009.06.004

Hoffmann, M., Chinn, L. K., Somogyi, E., Heed, T., Fagard, J., Lockman, J. J., & O’Regan, J. K. (2017). Development of reaching to the body in early infancy: From experiments to robotic models. *2017 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 112–119. https://doi.org/10.1109/DEVLRN.2017.8329795

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(5), 849–878. https://doi.org/10.1017/S0140525X01000103

Hooker, C. I., Bruce, L., Lincoln, S. H., Fisher, M., & Vinogradov, S. (2011). Theory of mind skills are related to gray matter volume in the ventromedial prefrontal cortex in schizophrenia. *Biological Psychiatry*, *70*(12), 1169–1178. https://doi.org/10.1016/j.biopsych.2011.07.027

Hughes, C., & Leekam, S. (2004). What are the Links Between Theory of Mind and Social Relations? Review, Reflections and New Directions for Studies of Typical and Atypical Development. *Social Development*, *13*(4), 590–619. https://doi.org/10.1111/j.1467-9507.2004.00285.x

Hyde, D., Simon, C., Ting, F., & Nikolaeva, J. (2018). Functional Organization of the Temporal-Parietal Junction for Theory of Mind in Preverbal Infants: A Near-Infrared Spectroscopy Study. *The Journal of Neuroscience*, *38*, 0264–17. https://doi.org/10.1523/JNEUROSCI.0264-17.2018

Hynes, C. A., Baird, A. A., & Grafton, S. T. (2006). Differential role of the orbital frontal lobe in emotional versus cognitive perspective-taking. *Neuropsychologia*, *44*(3), 374–383.

Inagaki, H., Meguro, K., Shimada, M., Ishizaki, J., Okuzumi, H., & Yamadori, A. (2002). Discrepancy between mental rotation and perspective-taking abilities in normal aging assessed by Piaget’s three-mountain task. *Journal of Clinical and Experimental Neuropsychology*, *24*(1), 18–25.

Jackson, P. L., Brunet, E., Meltzoff, A. N., & Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia*, *44*(5), 752–761. https://doi.org/10.1016/j.neuropsychologia.2005.07.015

James, M. A., Bagley, A. M., Brasington, K., Lutz, C., McConnell, S., & Molitor, F. (2006). Impact of prostheses on function and quality of life for children with unilateral congenital below-the-elbow deficiency. *The Journal of Bone and*

Joint Surgery. American Volume, *88*(11), 2356–2365. https://doi.org/10.2106/JBJS.E.01146

James, W. (1890). *The principles of psychology, Vol I.* (pp. xii, 697). Henry Holt and Co. https://doi.org/10.1037/10538-000

Jauniaux, J., Khatibi, A., Rainville, P., & Jackson, P. L. (2019). A meta-analysis of neuroimaging studies on pain empathy: Investigating the role of visual information and observers' perspective. *Social Cognitive and Affective Neuroscience*, *14*(8), 789–813.

Johnson, M. H. (2001). Functional brain development in humans. *Nature Reviews. Neuroscience*, *2*(7), 475–483. https://doi.org/10.1038/35081509

Johnson, M. H., Halit, H., Grice, S. J., & Karmiloff-Smith, A. (2002). Neuroimaging of typical and atypical development: A perspective from multiple levels of analysis. *Development and Psychopathology*, *14*(3), 521–536. https://doi.org/10.1017/s0954579402003073

Johnson, M. R., & Demiris, Y. K. (2005, September 30). Perspective Taking Through Simulation. *Towards Autonomous Robotic Systems (TAROS)*. http://spiral.imperial.ac.uk/handle/10044/1/12692

Kalbe, E., Grabenhorst, F., Brand, M., Kessler, J., Hilker, R., & Markowitsch, H. J. (2007). Elevated emotional reactivity in affective but not cognitive components of theory of mind: A psychophysiological study. *Journal of Neuropsychology*, *1*(1), 27–38. https://doi.org/10.1348/174866407x180792

Kampis, D., Karman, P., Csibra, G., Southgate, V., & Hernik, M. (2020). *A two-lab direct replication attempt of Southgate, Senju, & Csibra (2007)*. PsyArXiv. https://doi.org/10.31234/osf.io/gzy26

Kampis, D., & Kovács, Á. M. (2022). Seeing the world from others' perspective: 14-month-olds show altercentric modulation effects by others' beliefs. *Open Mind*, 1–19.

Kampis, D., Parise, E., Csibra, G., & Kovács, Á. M. (2015). Neural signatures for sustaining object representations attributed to others in preverbal human infants. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1819), 20151683. https://doi.org/10.1098/rspb.2015.1683

Kayhan, E., Meyer, M., O'Reilly, J. X., Hunnius, S., & Bekkering, H. (2019). Nine-month-old infants update their predictive models of a changing environment. *Developmental Cognitive Neuroscience*, *38*, 100680. https://doi.org/10.1016/j.dcn.2019.100680

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The Other-Race Effect Develops During Infancy. *Psychological Science*, *18*(12), 1084–1089. https://doi.org/10.1111/j.1467-9280.2007.02029.x

Kennedy, W., Bugajska, M., Harrison, A., & Trafton, J. (2009). "Like-Me" Simulation as an Effective and Cognitively Plausible Basis for Social Robotics. *I. J. Social Robotics*, *1*, 181–194. https://doi.org/10.1007/s12369-009-0014-6

Kessler, K., & Rutherford, H. (2010). The Two Forms of Visuo-Spatial Perspective Taking are Differently Embodied and Subserve Different Spatial Prepositions. *Frontiers in Psychology*, *1*. https://www.frontiersin.org/article/10.3389/fpsyg.2010.00213

Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: From self to social cognition. *Trends in Cognitive Sciences*, *11*(5), 194–196. https://doi.org/10.1016/j.tics.2007.02.002

King, M., Nazareth, I., Levy, G., Walker, C., Morris, R., Weich, S., Bellón-Saameño, J. A., Moreno, B., Svab, I., Rotar, D., Rifel, J., Maaroos, H.-I., Aluoja, A., Kalda, R., Neeleman, J., Geerlings, M. I., Xavier, M., de Almeida, M. C., Correa, B., & Torres-Gonzalez, F. (2008). Prevalence of common mental disorders in general practice attendees across Europe. *The British Journal of Psychiatry: The Journal of Mental Science*, *192*(5), 362–367. https://doi.org/10.1192/bjp.bp.107.039966

Klimecki, O. M., Leiberg, S., Lamm, C., & Singer, T. (2013). Functional neural plasticity and associated changes in positive affect after compassion training. *Cerebral Cortex (New York, N.Y.: 1991)*, *23*(7), 1552–1561. https://doi.org/10.1093/cercor/bhs142

Koay, K., Sisbot, E., Syrdal, D. S., Walters, M., Dautenhahn, K., & Alami, R. (2007). *Exploratory Study of a Robot Approaching a Person in the Context of Handing Over an Object.* 18–24.

Kominis, F., & Geffner, H. (2015). Beliefs in Multiagent Planning: From One Agent to Many. *Proceedings of the International Conference on Automated Planning and Scheduling*, *25*(1), 147–155.

Kominis, F., & Geffner, H. (2017). Multiagent Online Planning with Nested Beliefs and Dialogue. *Proceedings of the International Conference on Automated Planning and Scheduling*, *27*, 186–194.

Kooijman, C. M., Dijkstra, P. U., Geertzen, J. H. B., Elzinga, A., & van der Schans, C. P. (2000). Phantom pain and phantom sensations in upper limb amputees: An epidemiological study. *Pain*, *87*(1), 33–41. https://doi.org/10.1016/S0304-3959(00)00264-5

Koster-Hale, J., Bedny, M., & Saxe, R. (2014). Thinking about seeing: Perceptual sources of knowledge are encoded in the theory of mind brain regions of sighted and blind adults. *Cognition*, *133*(1), 65–78. https://doi.org/10.1016/j.cognition.2014.04.006

Koster-Hale, J., Richardson, H., Velez, N., Asaba, M., Young, L., & Saxe, R. (2017). Mentalizing regions represent distributed, continuous, and abstract dimensions of others' beliefs. *NeuroImage*, *161*, 9–18. https://doi.org/10.1016/j.neuroimage.2017.08.026

Koster-Hale, J., & Saxe, R. (2013). Theory of Mind: A Neural Prediction Problem. *Neuron*, *79*(5), 836–848. https://doi.org/10.1016/j.neuron.2013.08.020

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

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition*, *29*(5), 745–756. https://doi.org/10.3758/BF03200477

Kozhevnikov, M., Motes, M. A., Rasch, B., & Blajenkova, O. (2006). Perspective-taking vs. Mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology*, *20*(3), 397–417. https://doi.org/10.1002/acp.1192

Kraemer, M., Herold, M., Uekermann, J., Kis, B., Wiltfang, J., Daum, I., Dziobek, I., Berlit, P., Diehl, R. R., & Abdel-Hamid, M. (2013). Theory of mind and empathy in patients at an early stage of relapsing remitting multiple sclerosis. *Clinical Neurology and Neurosurgery*, *115*(7), 1016–1022. https://doi.org/10.1016/j.clineuro.2012.10.027

Krämer, U. M., Mohammadi, B., Doñamayor, N., Samii, A., & Münte, T. F. (2010). Emotional and cognitive aspects of empathy and their relation to social cognition—An fMRI-study. *Brain Research*, *1311*, 110–120. https://doi.org/10.1016/j.brainres.2009.11.043

Krane, E. J., & Heller, L. B. (1995). The prevalence of phantom sensation and pain in pediatric amputees. *Journal of Pain and Symptom Management*, *10*(1), 21–29. https://doi.org/10.1016/0885-3924(94)00062-P

Kulić, D., & Croft, E. A. (2005). Safe planning for human-robot interaction. *Journal of Robotic Systems*, *22*(7), 383–396. https://doi.org/10.1002/rob.20073

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

Kulke, L., von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is Implicit Theory of Mind a Real and Robust Phenomenon? Results From a Systematic Replication Study. *Psychological Science*, *29*(6), 888–900. https://doi.org/10.1177/0956797617747090

Kyberd, P. J., Beard, D. J., & Morrison, J. D. (1997). The population of users of upper limb prostheses attending the Oxford Limb Fitting Service. *Prosthetics and Orthotics International*, *21*(2), 85–91. https://doi.org/10.3109/03093649709164535

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*. https://doi.org/10.1017/S0140525X16001837

Lamm, C., Bukowski, H., & Silani, G. (2016). From shared to distinct self–other representations in empathy: Evidence from neurotypical function and socio-cognitive disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150083. https://doi.org/10.1098/rstb.2015.0083

Langdon, A., Botvinick, M., Nakahara, H., Tanaka, K., Matsumoto, M., & Kanai, R. (2022). Meta-learning, social cognition and consciousness in brains and machines. *Neural Networks*, *145*, 80–89. https://doi.org/10.1016/j.neunet.2021.10.004

Langdon, R., & Coltheart, M. (2001). Visual perspective-taking and schizotypy: Evidence for a simulation-based account of mentalizing in normal adults. *Cognition*, *82*(1), 1–26. https://doi.org/10.1016/S0010-0277(01)00139-1

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: Reliability and validity of the Empathy Quotient. *Psychological Medicine*, *34*(5), 911–919. https://doi.org/10.1017/s0033291703001624

Le, J. T., & Scott-Wyard, P. R. (2015). Pediatric limb differences and amputations. *Physical Medicine and Rehabilitation Clinics of North America*, *26*(1), 95–108. https://doi.org/10.1016/j.pmr.2014.09.006

Lecce, S., Bianco, F., Devine, R. T., & Hughes, C. (2017). Relations between theory of mind and executive function in middle childhood: A short-term longitudinal study. *Journal of Experimental Child Psychology*, *163*, 69–86. https://doi.org/10.1016/j.jecp.2017.06.011

Lemaignan, S., Warnier, M., Sisbot, E. A., Clodic, A., & Alami, R. (2017). Artificial cognition for social human–robot interaction: An implementation. *Artificial Intelligence*, *247*, 45–69. https://doi.org/10.1016/j.artint.2016.07.002

Leonard, H. C., & Hill, E. L. (2014). Review: The impact of motor development on typical and atypical social cognition and language: a systematic review. *Child and Adolescent Mental Health*, *19*(3), 163–170. https://doi.org/10.1111/camh.12055

Leslie, A. M. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In *Mapping the mind: Domain specificity in cognition and culture* (pp. 119–148). Cambridge University Press. https://doi.org/10.1017/CBO9780511752902.006

Lewkowicz, D. J., & Ghazanfar, A. A. (2011). Paradoxical psychological functioning in early child development. *The Paradoxical Brain*, 110–129. https://doi.org/10.1017/CBO9780511978098.008

Lin, Y., Ding, H., & Zhang, Y. (2020). Multisensory Integration of Emotion in Schizophrenic Patients. *Multisensory Research*, *33*(8), 865–901. https://doi.org/10.1163/22134808-bja10016

Liu, Y., Vannuscorps, G., Caramazza, A., & Striem-Amit, E. (2020). Evidence for an effector-independent action system from people born without hands. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(45), 28433–28441. https://doi.org/10.1073/pnas.2017789117

Longo, M. R., Schüür, F., Kammers, M. P. M., Tsakiris, M., & Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition*, *107*(3), 978–998. https://doi.org/10.1016/j.cognition.2007.12.004

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

Lungarella, M., Metta, G., Pfeifer, R., & Sandini, G. (2003). Developmental robotics: A survey. *Connection Science*, *15*(4), 151–190. https://doi.org/10.1080/09540090310001655110

Luo, Y. (2011). Do 10-month-old infants understand others' false beliefs? *Cognition*, *121*(3), 289–298. https://doi.org/10.1016/j.cognition.2011.07.011

Luo, Y., & Baillargeon, R. (2010). Toward a Mentalistic Account of Early Psychological Reasoning. *Current Directions in Psychological Science*, *19*, 301–307. https://doi.org/10.1177/0963721410386679

Lyu, Y., Guo, X., Bekrater-Bodmann, R., Flor, H., & Tong, S. (2017). An event-related potential study on the time course of mental rotation in upper-limb amputees. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *128*(5), 744–750. https://doi.org/10.1016/j.clinph.2017.02.008

MacIver, K., Lloyd, D. M., Kelly, S., Roberts, N., & Nurmikko, T. (2008). Phantom limb pain, cortical reorganization and the therapeutic effect of mental imagery. *Brain: A Journal of Neurology*, *131*(Pt 8), 2181–2191. https://doi.org/10.1093/brain/awn124

Mai, C. T., Isenburg, J. L., Canfield, M. A., Meyer, R. E., Correa, A., Alverson, C. J., Lupo, P. J., Riehle-Colarusso, T., Cho, S. J., Aggarwal, D., Kirby, R. S., & National Birth Defects Prevention Network. (2019). National population-based estimates for major birth defects, 2010-2014. *Birth Defects Research*, *111*(18), 1420–1435. https://doi.org/10.1002/bdr2.1589

Mainprice, J., Akin Sisbot, E., Jaillet, L., Cortes, J., Alami, R., & Simeon, T. (2011). Planning human-aware motions using a sampling-based costmap planner.

2011 IEEE International Conference on Robotics and Automation, 5012–5017. https://doi.org/10.1109/ICRA.2011.5980048

Malouin, F., & Richards, C. L. (2010). Mental practice for relearning locomotor skills. *Physical Therapy*, *90*(2), 240–251. https://doi.org/10.2522/ptj.20090029

Marschark, M., Edwards, L., Peterson, C., Crowe, K., & Walton, D. (2019). Understanding Theory of Mind in Deaf and Hearing College Students. *Journal of Deaf Studies and Deaf Education*, *24*(2), 104–118. https://doi.org/10.1093/deafed/eny039

Marsh, L. E., Mullett, T. L., Ropar, D., & Hamilton, A. F. de C. (2014). Responses to irrational actions in action observation and mentalising networks of the human brain. *NeuroImage*, *103*, 81–90. https://doi.org/10.1016/j.neuroimage.2014.09.020

Mayer, A., Kudar, K., Bretz, K., & Tihanyi, J. (2008). Body schema and body awareness of amputees. *Prosthetics and Orthotics International*, *32*(3), 363–382. https://doi.org/10.1080/03093640802024971

Mckechnie, P. S., & John, A. (2014). Anxiety and depression following traumatic limb amputation: A systematic review. *Injury*, *45*(12), 1859–1866. https://doi.org/10.1016/j.injury.2014.09.015

Meins, E., Fernyhough, C., Johnson, F., & Lidstone, J. (2006). Mind-mindedness in children: Individual differences in internal-state talk in middle childhood. *British Journal of Developmental Psychology*, *24*(1), 181–196. https://doi.org/10.1348/026151005X80174

Melchers, M., Montag, C., Markett, S., & Reuter, M. (2015). Assessment of empathy via self-report and behavioural paradigms: Data on convergent and discriminant validity. *Cognitive Neuropsychiatry*, *20*(2), 157–171. https://doi.org/10.1080/13546805.2014.991781

Meltzoff, A. N. (2007a). ‘Like me’: A foundation for social cognition. *Developmental Science*, *10*(1), 126–134. https://doi.org/10.1111/j.1467-7687.2007.00574.x

Meltzoff, A. N. (2007b). The ‘like me’ framework for recognizing and becoming an intentional agent. *Acta Psychologica*, *124*(1), 26–43. https://doi.org/10.1016/j.actpsy.2006.09.005

Meltzoff, A. N., & Borton, R. W. (1979). Intermodal matching by human neonates. *Nature*, *282*(5737), 403–404. https://doi.org/10.1038/282403a0

Melzack, R., Israel, R., Lacroix, R., & Schultz, G. (1997). Phantom limbs in people with congenital limb deficiency or amputation in early childhood. *Brain: A Journal of Neurology*, *120 (Pt 9)*, 1603–1620. https://doi.org/10.1093/brain/120.9.1603

Meristo, M., Falkman, K. W., Hjelmquist, E., Tedoldi, M., Surian, L., & Siegal, M. (2007). Language access and theory of mind reasoning: Evidence from deaf children in bilingual and oralist environments. *Developmental Psychology*, *43*(5), 1156–1169. https://doi.org/10.1037/0012-1649.43.5.1156

Meristo, M., Strid, K., & Hjelmquist, E. (2016). Early conversational environment enables spontaneous belief attribution in deaf children. *Cognition*, *157*, 139–145. https://doi.org/10.1016/j.cognition.2016.08.023

Milliez, G., Warnier, M., Clodic, A., & Alami, R. (2014). A framework for endowing an interactive robot with reasoning capabilities about perspective-taking and belief management. *IEEE RO-MAN 2014*, FrAT2.4. https://doi.org/10.1109/ROMAN.2014.6926399

Monroy, C., Gerson, S., & Hunnius, S. (2017). Infants' Motor Proficiency and Statistical Learning for Actions. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/article/10.3389/fpsyg.2017.02174

Montag, C., Neuhaus, K., Lehmann, A., Krüger, K., Dziobek, I., Heekeren, H. R., Heinz, A., & Gallinat, J. (2012). Subtle deficits of cognitive theory of mind in unaffected first-degree relatives of schizophrenia patients. *European Archives of Psychiatry and Clinical Neuroscience*, *262*(3), 217–226. https://doi.org/10.1007/s00406-011-0250-2

Montoya, P., Larbig, W., Grulke, N., Flor, H., Taub, E., & Birbaumer, N. (1997). The relationship of phantom limb pain to other phantom limb phenomena in upper extremity amputees. *Pain*, *72*(1–2), 87–93. https://doi.org/10.1016/s0304-3959(97)00004-3

Morash, V. S. (2016). Systematic Movements in Haptic Search: Spirals, Zigzags, and Parallel Sweeps. *IEEE Transactions on Haptics*, *9*(1), 100–110. https://doi.org/10.1109/TOH.2015.2508021

Moriguchi, Y., Ban, M., Osanai, H., & Uchiyama, I. (2018). Relationship between implicit false belief understanding and role play: Longitudinal study. *European Journal of Developmental Psychology*, *15*(2), 172–183. https://doi.org/10.1080/17405629.2017.1280022

Munakata, Y., Casey, B. J., & Diamond, A. (2004). Developmental cognitive neuroscience: Progress and potential. *Trends in Cognitive Sciences*, *8*(3), 122–128. https://doi.org/10.1016/j.tics.2004.01.005

Muncer, S. J., & Ling, J. (2006). Psychometric analysis of the empathy quotient (EQ) scale. *Personality and Individual Differences*, *40*(6), 1111–1119. https://doi.org/10.1016/j.paid.2005.09.020

Murray, C. E., Kelley-Soderholm, E. L., & Murray, T. L. (2007). Strengths, challenges, and relational processes in families of children with congenital upper limb differences. *Families, Systems, & Health*, *25*(3), 276–292. https://doi.org/10.1037/1091-7527.25.3.276

Murray, K., Johnston, K., Cunnane, H., Kerr, C., Spain, D., Gillan, N., Hammond, N., Murphy, D., & Happé, F. (2017). A new test of advanced theory of mind: The 'Strange Stories Film Task' captures social processing differences in adults with autism spectrum disorders. *Autism Research: Official Journal of the International Society for Autism Research*, *10*(6), 1120–1132. https://doi.org/10.1002/aur.1744

Nehaniv, C. L., & Dautenhahn, K. (2002). The correspondence problem. In *Imitation in animals and artifacts* (pp. 41–61). Boston Review. https://doi.org/10.7551/mitpress/3676.001.0001

Ng, A., & Jordan, M. (2002). On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive Bayes. *Advances in Neural Information Processing Systems*, *14*. https://proceedings.neurips.cc/paper/2001/hash/7b7a53e239400a13bd6be6c91c4f6c4e-Abstract.html

Oaksford, K., Frude, N., & Cuddihy, R. (2005). Positive Coping and Stress-Related Psychological Growth Following Lower Limb Amputation. *Rehabilitation Psychology*, *50*(3), 266–277. https://doi.org/10.1037/0090-5550.50.3.266

Ognibene, D., & Baldassare, G. (2015). Ecological active vision: four bioinspired principles to integrate bottom-up and adaptive top-down attention tested with a simple camera-arm robot. *IEEE Transations on Autonomous Mental Development*, 7, 3-25. https://doi.org/10.1109/TAMD.2014.2341351.

Ognibene, D., & Demiris, Y. (2013). *Towards active event recognition*. Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence.

Ognibene, D., Fiore, V. G., & Gu, X. (2019). Addiction beyond pharmacological effects: The role of environment complexity and bounded rationality. *Neural Networks : The Official Journal of the International Neural Network Society*, *116*, 269–278. https://doi.org/10.1016/j.neunet.2019.04.022

Omidshafiei, S., Pazis, J., Amato, C., How, J. P., & Vian, J. (2017). Deep Decentralized Multi-task Multi-Agent Reinforcement Learning under Partial Observability. *ArXiv:1703.06182 [Cs]*. http://arxiv.org/abs/1703.06182

Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science (New York, N.y.)*, *308*(5719), 255–258. https://doi.org/10.1126/science.1107621

Ornaghi, V., Brockmeier, J., & Gavazzi, I. G. (2011). The Role of Language Games in Children’s Understanding of Mental States: A Training Study. *Journal of Cognition and Development*, *12*(2), 239–259. https://doi.org/10.1080/15248372.2011.563487

Oudeyer, P.-Y. (2017). What do we learn about development from baby robots? *Wiley Interdisciplinary Reviews. Cognitive Science*, *8*(1–2). https://doi.org/10.1002/wcs.1395

Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016, July 9). *WebGazer: Scalable Webcam Eye Tracking Using User Interactions*.

Parker, J. L., & Robinson, C. W. (2018). Changes in multisensory integration across the life span. *Psychology and Aging*, *33*(3), 545–558. https://doi.org/10.1037/pag0000244

Patacchiola, M., & Cangelosi, A. (2016). A developmental Bayesian model of trust in artificial cognitive systems. *2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 117–123. https://doi.org/10.1109/DEVLRN.2016.7846801

Perner, J., & Ruffman, T. (2005). Infants' Insight into the Mind: How Deep? *Science*, *308*(5719), 214–216. https://doi.org/10.1126/science.1111656

Peterson, C. C., Peterson, J. L., & Webb, J. (2000). Factors influencing the development of a theory of mind in blind children. *British Journal of Developmental Psychology*, *18*(3), 431–447. https://doi.org/10.1348/026151000165788

Peterson, C. C., & Wellman, H. M. (2019). Longitudinal Theory of Mind (ToM) Development From Preschool to Adolescence With and Without ToM Delay. *Child Development*, *90*(6), 1917–1934. https://doi.org/10.1111/cdev.13064

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., & Spivey, M. J. (2011). The mechanics of embodiment: A dialog on embodiment and computational modeling. *Frontiers in Psychology*, *2*, 5. https://doi.org/10.3389/fpsyg.2011.00005

Pezzulo, G., Candidi, M., Dindo, H., & Barca, L. (2013). Action simulation in the human brain: Twelve questions. *New Ideas in Psychology*, *31*(3), 270–290. https://doi.org/10.1016/j.newideapsych.2013.01.004

Phillips, A. T., Wellman, H. M., & Spelke, E. S. (2002). Infants’ ability to connect gaze and emotional expression to intentional action. *Cognition*, *85*(1), 53–78. https://doi.org/10.1016/s0010-0277(02)00073-2

Piaget, J. (1952). *The origins of intelligence in children* (p. 419). W W Norton & Co. https://doi.org/10.1037/11494-000

Piaget, J. (1954). *The construction of reality in the child* (pp. xiii, 386). Basic Books. https://doi.org/10.1037/11168-000

Pino, M., Pettinelli, M., Clementi, D., Gianfelice, C., & Mazza, M. (2015). Improvement in cognitive and affective theory of mind with observation and imitation treatment in subjects with schizophrenia. *Clinical Neuropsychiatry*, *12*, 64–72.

Poli, F., Serino, G., Mars, R. B., & Hunnius, S. (2020). Infants tailor their attention to maximize learning. *Science Advances*, *6*(39), eabb5053. https://doi.org/10.1126/sciadv.abb5053

Poulin-Dubois, D., Azar, N., Elkaim, B., & Burnside, K. (2020). Testing the stability of theory of mind: A longitudinal approach. *PLoS ONE*, *15*(11), e0241721. https://doi.org/10.1371/journal.pone.0241721

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

Prasad, A., Niculescu-Mizil, A., & Ravikumar, P. (2017). On Separability of Loss Functions, and Revisiting Discriminative Vs Generative Models. *NIPS*.

Price, E. H. (2006). A critical review of congenital phantom limb cases and a developmental theory for the basis of body image. *Consciousness and Cognition*, *15*(2), 310–322. https://doi.org/10.1016/j.concog.2005.07.003

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

Prinz, W. (1997). Perception and Action Planning. *European Journal of Cognitive Psychology*, *9*(2), 129–154. https://doi.org/10.1080/713752551

Rabinowitz, N., Perbet, F., Song, F., Zhang, C., Eslami, S. M. A., & Botvinick, M. (2018). Machine Theory of Mind. *Proceedings of the 35th International Conference on Machine Learning*, 4218–4227. https://proceedings.mlr.press/v80/rabinowitz18a.html

Raileanu, R., Denton, E., Szlam, A., & Fergus, R. (2018). Modeling Others using Oneself in Multi-Agent Reinforcement Learning. *ArXiv:1802.09640 [Cs]*. http://arxiv.org/abs/1802.09640

Rakoczy, H. (2017). Theory of mind. In B. Hopkins, E. Geangu, & S. Linkenauger (Eds.), *The Cambridge Encyclopedia of Child Development* (2nd ed., pp. 505–512). Cambridge University Press. https://doi.org/10.1017/9781316216491.081

Ramirez, M., & Geffner, H. (2011). *Goal recognition over POMDPs: Inferring the intention of a POMDP agent*. 2009–2014. https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-335

Razmus, M., Daniluk, B., & Markiewicz, P. (2017). Phantom limb phenomenon as an example of body image distortion. *Current Problems of Psychiatry*, *18*(2), 153–159. https://doi.org/10.1515/cpp-2017-0013

Resnik, L., Borgia, M., & Clark, M. (2020). Function and Quality of Life of Unilateral Major Upper Limb Amputees: Effect of Prosthesis Use and Type. *Archives of*

Physical Medicine and Rehabilitation, *101*(8), 1396–1406. https://doi.org/10.1016/j.apmr.2020.04.003

Richardson, H., Koster-Hale, J., Caselli, N., Magid, R., Benedict, R., Olson, H., Pyers, J., & Saxe, R. (2020). Reduced neural selectivity for mental states in deaf children with delayed exposure to sign language. *Nature Communications*, *11*(1), 3246. https://doi.org/10.1038/s41467-020-17004-y

Richardson, H., & Saxe, R. (2020). Development of predictive responses in theory of mind brain regions. *Developmental Science*, *23*(1), e12863. https://doi.org/10.1111/desc.12863

Riva, F., Triscoli, C., Lamm, C., Carnaghi, A., & Silani, G. (2016). Emotional Egocentricity Bias Across the Life-Span. *Frontiers in Aging Neuroscience*, *8*, 74. https://doi.org/10.3389/fnagi.2016.00074

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*(9), 661–670. https://doi.org/10.1038/35090060

Ruder, S. (2017). An Overview of Multi-Task Learning in Deep Neural Networks. *ArXiv:1706.05098 [Cs, Stat]*. http://arxiv.org/abs/1706.05098

Ruffman, T., & Perner, J. (2005). Do infants really understand false belief?: Response to Leslie. *Trends in Cognitive Sciences*, *9*(10), 462–463. https://doi.org/10.1016/j.tics.2005.08.001

Rybarczyk, B., Nyenhuis, D. L., Nicholas, J. J., Cash, S. M., & Kaiser, J. (1995). Body image, perceived social stigma, and the prediction of psychosocial adjustment to leg amputation. *Rehabilitation Psychology*, *40*(2), 95–110. https://doi.org/10.1037/0090-5550.40.2.95

Saadah, E. S. M., & Melzack, R. (1994). Phantom limb experiences in congenital limb-deficient adults. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *30*(3), 479–485. https://doi.org/10.1016/S0010-9452(13)80343-7

Sak-Wernicka, J. (2016). Exploring Theory of Mind Use in Blind Adults During Natural Communication. *Journal of Psycholinguistic Research*, *45*(4), 857–869. https://doi.org/10.1007/s10936-015-9379-x

Sandini, G., Sciutti, A., & Vernon, D. (2021). Cognitive Robotics. In M. H. Ang, O. Khatib, & B. Siciliano (Eds.), *Encyclopedia of Robotics* (pp. 1–7). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41610-1_198-1

Saruco, E., Guillot, A., Saimpont, A., Di Rienzo, F., Durand, A., Mercier, C., Malouin, F., & Jackson, P. (2019). Motor imagery ability of patients with lower-limb amputation: Exploring the course of rehabilitation effects. *European Journal of Physical and Rehabilitation Medicine*, *55*(5), 634–645. https://doi.org/10.23736/S1973-9087.17.04776-1

Saxe, R., & Baron-Cohen, S. (2006). Editorial: The neuroscience of theory of mind. *Social Neuroscience*, *1*(3–4), 1–9. https://doi.org/10.1080/17470910601117463

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding Other Minds: Linking Developmental Psychology and Functional Neuroimaging. *Annual Review of Psychology*, *55*, 87–124. https://doi.org/10.1146/annurev.psych.55.090902.142044

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in ‘theory of mind’. *NeuroImage*, *19*(4), 1835–1842. https://doi.org/10.1016/s1053-8119(03)00230-1

Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia*, *42*(11), 1435–1446. https://doi.org/10.1016/j.neuropsychologia.2004.04.015

Scassellati, B. (2002). Theory of Mind for a Humanoid Robot. *Autonomous Robots*, *12*, 13–24. https://doi.org/10.1023/A:1013298507114

Schiffer, B., Pawliczek, C., Müller, B. W., Wiltfang, J., Brüne, M., Forsting, M., Gizewski, E. R., Leygraf, N., & Hodgins, S. (2017). Neural Mechanisms Underlying Affective Theory of Mind in Violent Antisocial Personality Disorder and/or Schizophrenia. *Schizophrenia Bulletin*, *43*(6), 1229–1239. https://doi.org/10.1093/schbul/sbx012

Schlinger, H. D. (2009). Theory of Mind: An Overview and Behavioral Perspective. *The Psychological Record*, *59*(3), 435–448. https://doi.org/10.1007/BF03395673

Schneider, D., Slaughter, V. P., Bayliss, A. P., & Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition*, *129*(2), 410–417. https://doi.org/10.1016/j.cognition.2013.08.004

Schuwerk, T., Vuori, M., & Sodian, B. (2015). Implicit and explicit Theory of Mind reasoning in autism spectrum disorders: The impact of experience. *Autism*, *19*(4), 459–468. https://doi.org/10.1177/1362361314526004

Scott, K., & Schulz, L. (2017). Lookit (Part 1): A New Online Platform for Developmental Research. *Open Mind*, *1*(1), 4–14. https://doi.org/10.1162/OPMI_a_00002

Scott, M. R., Richman, J. C., & Baillargeon, R. (2015). Infants understand deceptive intentions to implant false beliefs about identity: New evidence for early mentalistic reasoning. *Cognitive Psychology*, *82*, 32–56. https://doi.org/10.1016/j.cogpsych.2015.08.003

Scott, R. M. (2017). Surprise! 20-month-old infants understand the emotional consequences of false beliefs. *Cognition*, *159*, 33–47. https://doi.org/10.1016/j.cognition.2016.11.005

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

Scott, R. M., Baillargeon, R., Song, H., & Leslie, A. M. (2010). Attributing false beliefs about non-obvious properties at 18 months. *Cognitive Psychology*, *61*(4), 366–395. https://doi.org/10.1016/j.cogpsych.2010.09.001

Seiryte, A., & Rusconi, E. (2015). The Empathy Quotient (EQ) predicts perceived strength of bodily illusions and illusion-related sensations of pain. *Personality and Individual Differences*, *77*. https://doi.org/10.1016/j.paid.2014.12.048

Semmelmann, K., Hönekopp, A., & Weigelt, S. (2017). Looking Tasks Online: Utilizing Webcams to Collect Video Data from Home. *Frontiers in Psychology*, *8*, 1582. https://doi.org/10.3389/fpsyg.2017.01582

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*(2), 451–465. https://doi.org/10.3758/s13428-017-0913-7

Senju, A. (2012). Spontaneous theory of mind and its absence in autism spectrum disorders. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *18*(2), 108–113. https://doi.org/10.1177/1073858410397208

Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-months-olds really attribute mental states to others? A critical test.

Psychological Science, *22*(7), 878–880. https://doi.org/10.1177/0956797611411584

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science (New York, N.Y.)*, *325*(5942), 883–885. https://doi.org/10.1126/science.1176170

Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia*, *45*(13), 3054–3067. https://doi.org/10.1016/j.neuropsychologia.2007.05.021

Shimoda, S., Jamone, L., Ognibene, D., Nagai, T., Sciutti, A., Costa-Garcia, A., Oseki, Y., & Taniguchi, T. (2022). What is the role of the next generation of cognitive robotics? *Advanced Robotics*, *36*(1–2), 3–16. https://doi.org/10.1080/01691864.2021.2011780

Sisbot, E. A., Marin-Urias, L. F., Alami, R., & Simeon, T. (2007). A Human Aware Mobile Robot Motion Planner. *IEEE Transactions on Robotics*, *23*(5), 874–883. https://doi.org/10.1109/TRO.2007.904911

Skerry, A. E., Carey, S. E., & Spelke, E. S. (2013). First-person action experience reveals sensitivity to action efficiency in prereaching infants. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(46), 18728–18733. https://doi.org/10.1073/pnas.1312322110

Sodian, B., & Kristen, S. (2016). Theory of Mind. In *Handbook of epistemic cognition* (pp. 68–85).

Sodian, B., Kristen-Antonow, S., & Kloo, D. (2020). How Does Children’s Theory of Mind Become Explicit? A Review of Longitudinal Findings. *Child Development Perspectives*, *14*(3), 171–177. https://doi.org/10.1111/cdep.12381

Sodian, B., & Thoermer, C. (2004). Infants' Understanding of Looking, Pointing, and Reaching as Cues to Goal-Directed Action. *Journal of Cognition and Development*, *5*(3), 289–316. https://doi.org/10.1207/s15327647jcd0503_1

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*, 295–315. https://doi.org/10.1016/j.cognition.2008.08.008

Southgate, V. (2020). Are infants altercentric? The other and the self in early social cognition. *Psychological Review*, *127*(4), 505–523. https://doi.org/10.1037/rev0000182

Southgate, V., Johnson, M. H., & Csibra, G. (2008). Infants attribute goals even to biomechanically impossible actions. *Cognition*, *107*(3), 1059–1069. https://doi.org/10.1016/j.cognition.2007.10.002

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

Southgate, V., Johnson, M. H., Osborne, T., & Csibra, G. (2009). Predictive motor activation during action observation in human infants. *Biology Letters*, *5*(6), 769–772. https://doi.org/10.1098/rsbl.2009.0474

Southgate, V., Senju, A., & Csibra, G. (2007). Action Anticipation Through Attribution of False Belief by 2-Year-Olds. *Psychological Science*, *18*(7), 587–592. https://doi.org/10.1111/j.1467-9280.2007.01944.x

Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*, *130*(1), 1–10. https://doi.org/10.1016/j.cognition.2013.08.008

Spek, A. A., Scholte, E. M., & Van Berckelaer-Onnes, I. A. (2010). Theory of mind in adults with HFA and Asperger syndrome. *Journal of Autism and Developmental Disorders*, *40*(3), 280–289. https://doi.org/10.1007/s10803-009-0860-y

Spelke, E. S. (2002). Developmental neuroimaging: A developmental psychologist looks ahead. *Blackwell Publishers Ltd*, 5.

Spradlin, J. E., & Brady, N. (2008). A Behavior Analytic Interpretation of Theory of Mind. *Revista Internacional de Psicologia y Terapia Psicologica = International Journal of Psychology and Psychological Therapy*, *8*(3), 335–350.

Stankevicius, A., Wallwork, S. B., Summers, S. J., Hordacre, B., & Stanton, T. R. (2021). Prevalence and incidence of phantom limb pain, phantom limb sensations and telescoping in amputees: A systematic rapid review. *European Journal of Pain*, *25*(1), 23–38. https://doi.org/10.1002/ejp.1657

Stein, B. E., & Rowland, B. A. (2011). Organization and plasticity in multisensory integration: Early and late experience affects its governing principles. *Progress in Brain Research*, *191*, 145–163. https://doi.org/10.1016/B978-0-444-53752-2.00007-2

Steinbeis, N. (2016). The role of self-other distinction in understanding others’ mental and emotional states: Neurocognitive mechanisms in children and adults. *Philosophical Transactions of the Royal Society of London. Series B,* *Biological Sciences*, *371*(1686), 20150074. https://doi.org/10.1098/rstb.2015.0074

Stephens-Fripp, B., Jean Walker, M., Goddard, E., & Alici, G. (2020). A survey on what Australians with upper limb difference want in a prosthesis: Justification for using soft robotics and additive manufacturing for customized prosthetic hands. *Disability and Rehabilitation. Assistive Technology*, *15*(3), 342–349. https://doi.org/10.1080/17483107.2019.1580777

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, *10*(5), 640–656. https://doi.org/10.1162/089892998562942

Striem-Amit, E., Vannuscorps, G., & Caramazza, A. (2018). Plasticity based on compensatory effector use in the association but not primary sensorimotor cortex of people born without hands. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(30), 7801–7806. https://doi.org/10.1073/pnas.1803926115

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

Surian, L., & Franchin, L. (2020). On the domain specificity of the mechanisms underpinning spontaneous anticipatory looks in false-belief tasks. *Developmental Science*, *23*(6), e12955. https://doi.org/10.1111/desc.12955

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

Symons, D. K. (2004). Mental state discourse, theory of mind, and the internalization of self–other understanding. *Developmental Review*, *24*(2), 159–188. https://doi.org/10.1016/j.dr.2004.03.001

Taniguchi, T. (2016). Symbol Emergence in Robotics for Long-Term Human-Robot Collaboration**This research was partially supported by a Grant-in-Aid for Young Scientists (B) 2012-2014 (24700233) and a Grant-in-Aid for Young Scientists (A) 2015-2019 (15H05319) funded by the Ministry of Education, Culture, Sports, Science, and Technology, Japan, and by CREST, JST. *IFAC-PapersOnLine*, *49*(19), 144–149. https://doi.org/10.1016/j.ifacol.2016.10.476

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

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

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

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

Tran, M., Cabral, L., Patel, R., & Cusack, R. (2017). Online recruitment and testing of infants with Mechanical Turk. *Journal of Experimental Child Psychology*, *156*, 168–178. https://doi.org/10.1016/j.jecp.2016.12.003

Träuble, B., Marinović, V., & Pauen, S. (2010). Early Theory of Mind Competencies: Do Infants Understand Others' Beliefs? *Infancy: The Official Journal of the International Society on Infant Studies*, *15*(4), 434–444. https://doi.org/10.1111/j.1532-7078.2009.00025.x

Tsakiris, M. (2008). Looking for myself: Current multisensory input alters self-face recognition. *PloS One*, *3*(12), e4040. https://doi.org/10.1371/journal.pone.0004040

Tsakiris, M. (2017). The multisensory basis of the self: From body to identity to others. *Quarterly Journal of Experimental Psychology (2006)*, *70*(4), 597–609. https://doi.org/10.1080/17470218.2016.1181768

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, *121*(4), 649–675. https://doi.org/10.1037/a0037665

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

van Overwalle, F., & Vandekerckhove, M. (2013). Implicit and explicit social mentalizing: Dual processes driven by a shared neural network. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00560

Vanderelst, D., & Winfield, A. (2018). An architecture for ethical robots inspired by the simulation theory of cognition. *Cognitive Systems Research*, *48*, 56–66. https://doi.org/10.1016/j.cogsys.2017.04.002

Vannuscorps, G., Andres, M., & Caramazza, A. (2020). Efficient recognition of facial expressions does not require motor simulation. *ELife*, *9*, e54687. https://doi.org/10.7554/eLife.54687

Vannuscorps, G., Andres, M., Carneiro, S. P., Rombaux, E., & Caramazza, A. (2021). Typically Efficient Lipreading without Motor Simulation. *Journal of Cognitive Neuroscience*, *33*(4), 611–621. https://doi.org/10.1162/jocn_a_01666

Vannuscorps, G., & Caramazza, A. (2015). Motor simulation does not underlie action perception: Evidence from upper limb dysmelia. *Journal of Vision*, *15*, 559. https://doi.org/10.1167/15.12.559

Vannuscorps, G., & Caramazza, A. (2016). Typical action perception and interpretation without motor simulation. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(1), 86–91. https://doi.org/10.1073/pnas.1516978112

Vannuscorps, G., & Caramazza, A. (2019). Conceptual processing of action verbs with and without motor representations. *Cognitive Neuropsychology*, *36*(7–8), 301–312. https://doi.org/10.1080/02643294.2020.1732319

Vannuscorps, G., F Wurm, M., Striem-Amit, E., & Caramazza, A. (2019). Large-Scale Organization of the Hand Action Observation Network in Individuals Born Without Hands. *Cerebral Cortex (New York, N.Y.: 1991)*, *29*(8), 3434–3444. https://doi.org/10.1093/cercor/bhy212

Varni, J. W., Setoguchi, Y., Rappaport, L. R., & Talbot, D. (1991). Effects of stress, social support, and self-esteem on depression in children with limb deficiencies. *Archives of Physical Medicine and Rehabilitation*, *72*(13), 1053–1058.

Vinanzi, S., Patacchiola, M., Chella, A., & Cangelosi, A. (2019). Would a robot trust you? Developmental robotics model of trust and theory of mind. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1771), 20180032. https://doi.org/10.1098/rstb.2018.0032

Völlm, B. A., Taylor, A. N. W., Richardson, P., Corcoran, R., Stirling, J., McKie, S., Deakin, J. F. W., & Elliott, R. (2006). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *NeuroImage*, *29*(1), 90–98. https://doi.org/10.1016/j.neuroimage.2005.07.022

Vuillermin, C., Canizares, M. F., Bauer, A. S., Miller, P. E., Goldfarb, C. A., & CoULD Study Group. (2021). Congenital Upper Limb Differences Registry (CoULD): Registry Inclusion Effect. *The Journal of Hand Surgery*, *46*(6), 515.e1-515.e11. https://doi.org/10.1016/j.jhsa.2020.11.006

Wall, L. B., Wright, M., Samora, J., Bae, D. S., Steinman, S., & Goldfarb, C. A. (2021). Social Deprivation and Congenital Upper Extremity Differences—An Assessment Using PROMIS. *The Journal of Hand Surgery*, *46*(2), 114–118. https://doi.org/10.1016/j.jhsa.2020.08.017

Wang, S., Andrews, G., Pendergast, D., Neumann, D., Chen, Y., & Shum, D. (2021). A Cross-Cultural Study of Theory of Mind Using Strange Stories in School-Aged Children from Australia and Mainland China. *Journal of Cognition and Development*, 1–24. https://doi.org/10.1080/15248372.2021.1974445

Ward, J., Schnakenberg, P., & Banissy, M. J. (2018). The relationship between mirror-touch synaesthesia and empathy: New evidence and a new screening tool. *Cognitive Neuropsychology*, *35*(5–6), 314–332. https://doi.org/10.1080/02643294.2018.1457017

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

Wellman, H. M., Phillips, A. T., Dunphy-Lelii, S., & LaLonde, N. (2004). Infant social attention predicts preschool social cognition. *Developmental Science*, *7*(3), 283–288. https://doi.org/10.1111/j.1467-7687.2004.00347.x

Wiese, E., Metta, G., & Wykowska, A. (2017). Robots As Intentional Agents: Using Neuroscientific Methods to Make Robots Appear More Social. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/article/10.3389/fpsyg.2017.01663

Wilkins, K. L., McGrath, P. J., Finley, A. G., & Katz, J. (1998). Phantom limb sensations and phantom limb pain in child and adolescent amputees. *Pain*, *78*(1), 7–12. https://doi.org/10.1016/S0304-3959(98)00109-2

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

Winfield, A. F. T. (2018). Experiments in Artificial Theory of Mind: From Safety to Story-Telling. *Frontiers in Robotics and AI*, *5*. https://www.frontiersin.org/article/10.3389/frobt.2018.00075

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

Xie, J., Cheung, H., Shen, M., & Wang, R. (2018). Mental Rotation in False Belief Understanding. *Cognitive Science*, *42*(4), 1179–1206. https://doi.org/10.1111/cogs.12594

Yang, G.-Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., Jacobstein, N., Kumar, V., McNutt, M., Merrifield, R., Nelson, B. J., Scassellati, B., Taddeo, M., Taylor, R., Veloso, M., Wang, Z. L., & Wood, R. (2018). The grand challenges of Science Robotics. *Science Robotics*, *3*(14), eaar7650. https://doi.org/10.1126/scirobotics.aar7650

Yang, X., & Krajbich, I. (2020). *Webcam-based online eye-tracking for behavioral research*. PsyArXiv. https://doi.org/10.31234/osf.io/qhme6

Yeung, E., Askitis, D., Manea, V., & Southgate, V. (2022). *Emerging Self-Representation Presents a Challenge for Perspective Tracking in Infancy*. https://doi.org/10.31234/osf.io/tv2kb

Yott, J., & Poulin-Dubois, D. (2012). Breaking the rules: Do infants have a true understanding of false belief? *The British Journal of Developmental Psychology*, *30*(Pt 1), 156–171. https://doi.org/10.1111/j.2044-835X.2011.02060.x

Zanatto, D., Patacchiola, M., Cangelosi, A., & Goslin, J. (2020). Generalisation of Anthropomorphic Stereotype. *International Journal of Social Robotics*, *12*(1), 163–172. https://doi.org/10.1007/s12369-019-00549-4

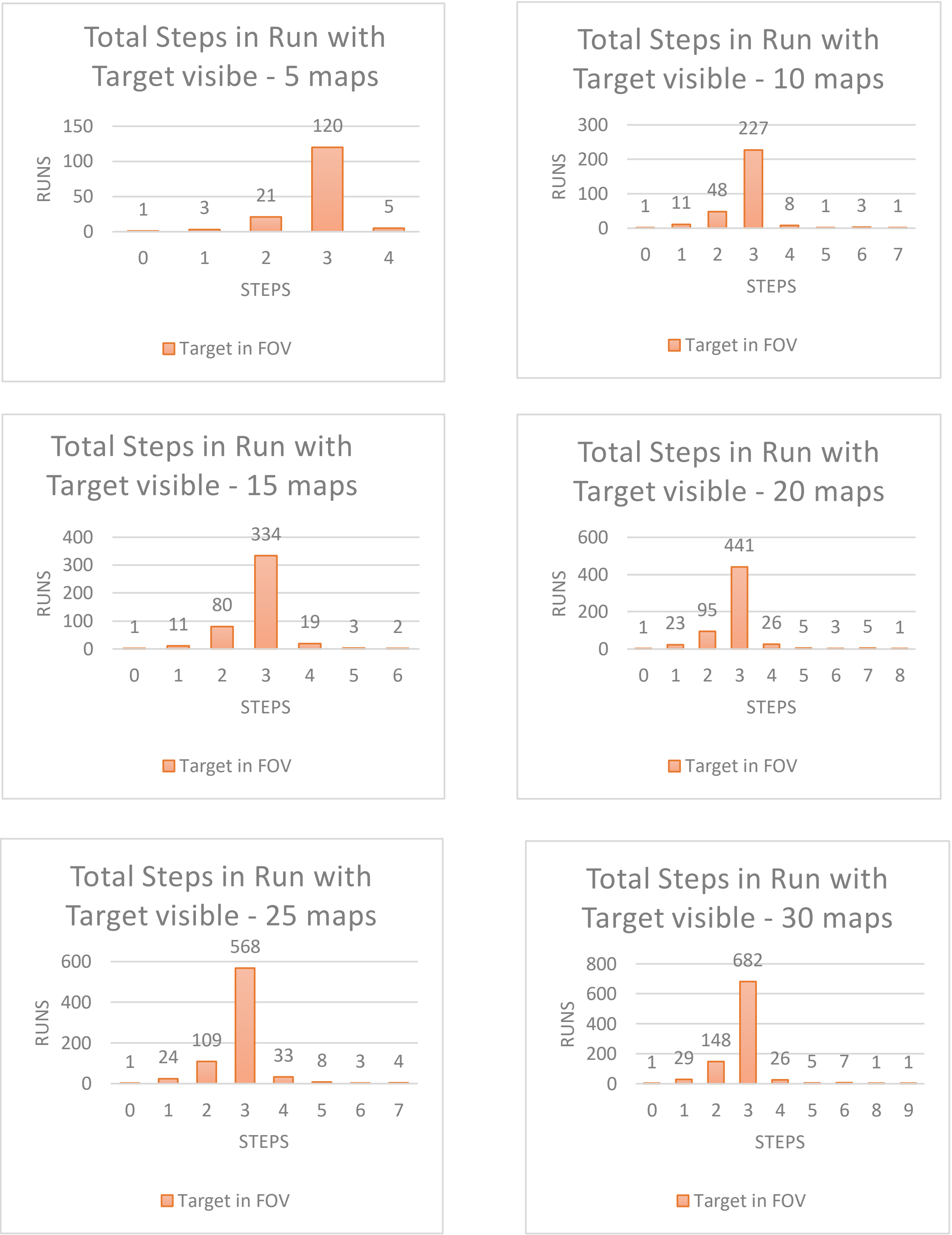
Zanatto, D., Patacchiola, M., Goslin, J., & Cangelosi, A. (2019). Investigating cooperation with robotic peers. *PLOS ONE*, *14*(11), e0225028. https://doi.org/10.1371/journal.pone.0225028

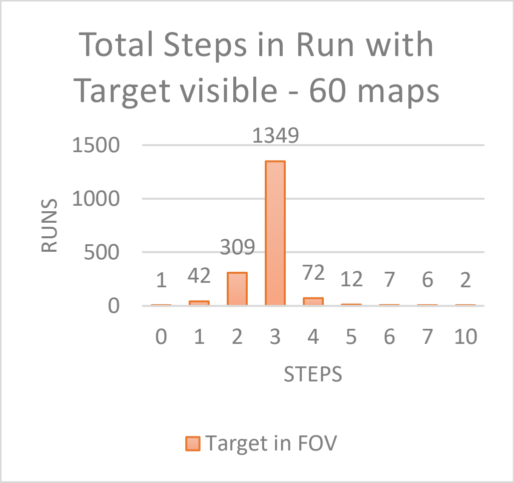
Zeng, Y., Zhao, Y., Zhang, T., Zhao, D., Zhao, F., & Lu, E. (2020). A Brain-Inspired Model of Theory of Mind. *Frontiers in Neurorobotics*, *14*, 60. https://doi.org/10.3389/fnbot.2020.00060

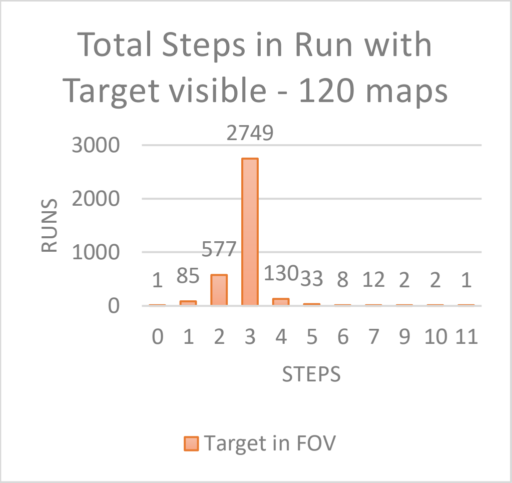
Ziegler-Graham, K., MacKenzie, E. J., Ephraim, P. L., Travison, T. G., & Brookmeyer, R. (2008). Estimating the prevalence of limb loss in the United States: 2005 to 2050. *Archives of Physical Medicine and Rehabilitation*, *89*(3), 422–429. https://doi.org/10.1016/j.apmr.2007.11.005

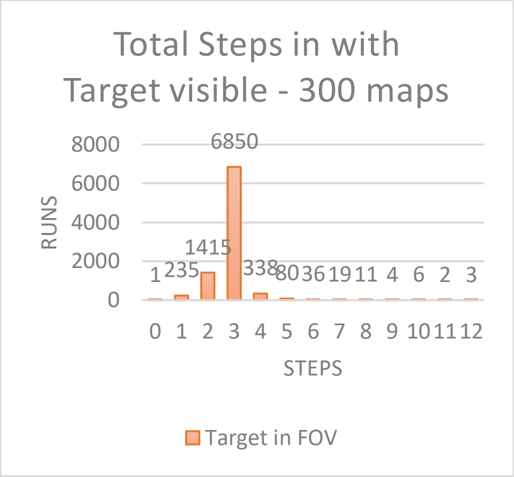
Supplementary Materials

Supplementary material 1. Number of steps performed by the actor to reach target once *visible*, by number of maps. Generally, the actor takes 3 steps to reach the target once it becomes *visible*.




Total Steps in Run with Target visible - 60 maps
RUNS
1500
1000
500
0
1349
309
1
42
72
12
7
6
2
0
1
2
3
4
5
6
7
10
STEPS
Target in FOV


Total Steps in Run with Target visible - 120 maps
RUNS
3000
2000
1000
0
2749
577
1
85
130
33
8
12
2
2
1
0
1
2
3
4
5
6
7
9
10
11
STEPS
Target in FOV


Total Steps in with Target visible - 300 maps
RUNS
8000
6000
4000
2000
0
6850
1415
1
235
338
80
36
19
11
4
6
2
3
0
1
2
3
4
5
6
7
8
9
10
11
12
STEPS
Target in FOV

Supplementary material 2. Best target prediction accuracies for the *Beliefs* vs *NoBeliefs* architectures by number of training maps in the conditions with the target *not visible* by the actor and No distractor objects *visible*, controlling for objects not being in the actor's past trajectory. As a result, equal probabilities of being targets (~25%) were assigned to all objects. Accuracies were calculated as averages across 18 initial network weights; the associated variances were reported.

	Target *not visible* - No Objects *visible*					
	BEL		NoBEL		BEL-NoBEL (%)	*p-value*
Train Maps (N)	Avg Acc (%)	Var	Avg Acc (%)	Var		
5	22.00	11.76	25.00	4.00	-3.00	***0.003***
10	24.33	1.41	26.11	2.81	-1.78	***0.001***
15	26.00	3.29	26.94	3.00	-0.94	0.119
20	26.11	2.46	27.56	4.26	-1.44	***0.024***
25	26.06	1.94	26.83	2.85	-0.78	0.141
30	25.72	1.51	26.94	2.41	-1.22	***0.013***
60	26.00	2.00	27.56	2.26	-1.56	***0.003***
120	27.39	1.90	28.17	3.91	-0.78	0.160
300	26.50	0.97	26.06	1.00	0.44	0.188
				avg.	-1.23	
				max.	0.44	

BEL: *Beliefs* architecture; NoBEL: *NoBeliefs* architecture