Modulating Mental State Decoding and Reasoning in Autism: The Importance of Context

Ruihan Wu

Institute of Cognitive Neuroscience

Division of Psychology and Language Sciences

Faculty of Brain Sciences

University College London (UCL)

Thesis submitted to UCL for the Degree of Doctor of Philosophy

March 2024

I, Ruihan Wu confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: [redacted]

Date 6th March 2024

Abstract

Autism affects social cognition in many ways, including mentalizing and perception. This thesis is particularly interested in studying false-belief reasoning and smile perception in autism by using a variety of measurement methods including eye-tracking, fNIRS, and behavioural assessments. Moreover, there are limited studies that look at factors that may improve autistic people's performance in mentalizing and social cue perception. Contextual factors hold promise in altering mentalizing performance, as has been found in non-autistic adults, but has not been explored in autistic people. There are two major themes in my PhD. The first theme is to investigate whether implicit mentalizing plays a role in autistic cognition. Through adapting an existing anticipatory-looking paradigm for measuring implicit mentalizing in autistic adults; I found autistic adults perceive social cues in the same way as non-autistic adults, but this information is not then used to update mental representations. I also found mentalizing, compensation, and mental health are associated with each other using another modified implicit mentalizing paradigm in a genetically predisposed population and a non-autistic sample. And, mothers of autistic children reported poorer mental health than mothers of non-autistic children. The second theme is to explore how contextual information (i.e., evaluative context and intergroup bias) would modulate mentalizing by using both the anticipatory-looking paradigm and a genuine-posed smile discrimination task, as well as the corresponding neural mechanism using fNIRS. I found autistic adults are equally affected by contextual information, but tend to possess difficulties in mental state decoding and reasoning and are less likely to identify with their in-group than their non-autistic counterparts. These findings extend the current understanding of mentalizing abilities in autism. The thesis will be discussed together with the current theories in mentalizing, intergroup bias, double empathy problem, and social mimicry.

Impact Statement

Identifying autism is challenging. Some autistic females or individuals with higher IQs do not receive a timely diagnosis. The delayed and missed diagnoses have likely resulted from the current behavioural observation and parental reports autism diagnostic framework. Autistic people have been thought to exhibit difficulties in mentalizing, yet, the evidence is mixed in the literature. Further, mentalizing can be enhanced or suppressed, however, we know very little about what factors can modulate mentalizing in autistic and non-autistic people and the corresponding cognitive and neural mechanisms. This work firstly investigates whether implicit mentalizing plays a role in autistic cognition by accurately assessing mentalizing abilities in autism and secondly explores potential factors that may modulate mentalizing. The findings have potential impacts both inside and outside of academia.

Inside academia

Scholarship. This thesis contributes to the limited body of literature on modulating mentalizing in autism. To the best of my knowledge, this is the first work exploring the modulation effect of contextual factors (i.e., contextual evaluability and intergroup bias) on mentalizing in diverse populations. The findings in all the experimental chapters have been reported in conferences and have been published(2), submitted(1) to or are in preparation(1) for submission to scholarly journals. The findings establish that certain social cognition and social judgement can be modulated by contextual factors, which might mitigate social difficulties in autism.

In addition, they suggest a reconsideration of past findings that might have misrepresented the social judgements of autistic people through introducing an outgroup disadvantage which contributes to a better understanding of autism. The findings also provide innovative insights into the intricate interplay between behaviour and neural activities in modulating mentalizing, advocating for investigating mentalizing within contexts characterized by diverse evaluability and intergroup dynamics.

Methodology. Through resolving criticisms of past mentalizing paradigms identified in the literature, three mentalizing tasks were designed which are potentially more robust according to the findings. This research also implemented a multimodal approach to disentangle the complexity of measuring and modulating mentalizing and to reveal the underpinning neural mechanisms. This integrated behavioural, cognitive and neural measures and applied behavioural assessments, self-report inventories, eye tracking, facial movements, video recordings, and brain activity recordings (fNIRS) in social cognitive studies. This multi-dimensional approach provides more information than the traditional single-dimensional approach in understanding social cognition in autism.

Outside academia

Well-being, diagnosis, and life quality. The findings emphasize the need to support families with autistic members in terms of *mental health* and psychological resilience and could be used in further research, *policy*-making or *service* delivery in relation to autistic people and families. The findings imply that the contextual information might impact the results of autism *diagnosis* and *interventions*, and the mentalizing paradigms designed have the potential to be adapted for use in autism *diagnostic assessments*. This work promotes the design of tailored *educational and working supports and policies* for autistic social differences that emphasize similarities and transparency between diverse people. These have the potential to improve the social experience and *life quality* of autistic people and make society more inclusive.

UCL Research Paper Declaration Form (1/3)

This form references the doctoral candidate's own published work(s)

1. **For a research manuscript prepared for publication but that has not yet been published**

   a) **What is the current title of the manuscript?**

   Do autistic adults spontaneously reason about belief? A detailed exploration of alternative explanations

   b) **Has the manuscript been uploaded to a preprint server?** (e.g. medRxiv; if 'Yes', please give a link or doi)

   https://doi.org/10.21203/rs.3.rs-267044/v1

   c) **Where is the work intended to be published?** (e.g. journal names)

   Royal Society Open Science

   d) **List the manuscript's authors in the intended authorship order**

   Ruihan Wu, Jing Tian Lim, Zahra Ahmed, Ensar Acem, Ishita Chowdhury, Sarah J White

   e) **Stage of publication** (e.g. in submission)

   In review

2. **For multi-authored work, please give a statement of contribution covering all authors**

RW, SW and JL conceived and designed the study; JL filmed the stimuli; JLT and ZA implemented the experiment; ZA collected the data; ZA, EA and IC conducted preliminary data analysis, RW performed the final data analysis, RW and SW interpreted the results and drafted the manuscript. SW supervised the project and acquired the funding.

3. **In which chapter(s) of your thesis can this material be found?**

   Chapter 2

4. **e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

   Candidate

   Ruihan Wu

   Date: 3rd March 2024

   Supervisor/ Senior Author (where appropriate)

   Sarah J. White

   Date: 3rd March 2024

UCL Research Paper Declaration Form (2/3)

This form references the doctoral candidate's own published work(s)

1. **For a research manuscript that has already been published** (if not yet published, please skip to section 2)

   a) **What is the title of the manuscript?**

   Evaluative contexts facilitate implicit mentalizing: relation to the broader autism phenotype and mental health

   b) **Please include a link to or doi for the work**

   https://doi.org/10.1038/s41598-024-55075-9

   c) **Where was the work published?**

   Scientific Reports

   d) **Who published the work?** (e.g. OUP)

   Nature Portfolio

   e) **When was the work published?**

   26th February 2024

   f) **List the manuscript's authors in the order they appear on the publication**

Ruihan Wu, Karen Leow, Nicole Yu, Ciara Rafter, Katia Rosenbaum, Antonia F. de C. Hamilton & Sarah J. White

g) **Was the work peer reviewed?**

Yes

h) **Have you retained the copyright?**

Yes

i) **Was an earlier form of the manuscript uploaded to a preprint server?** (e.g. medRxiv). If 'Yes', please give a link or doi)

Yes

https://doi.org/10.31234/osf.io/m2n9u

If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

□

*I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.*

2. **For multi-authored work, please give a statement of contribution covering all authors** (if single-author, please skip to section 4)

R.W. the conception and design of the project; the acquisition, analysis, and interpretation of data; and have drafted the manuscript and substantively revised it. K.L., N.Y., C.R. and

K.R. the acquisition of data. A.H. the design of the project; supervision; the analysis of data. S.W. the conception and design of the project; project supervision; funding acquisition; the analysis and interpretation of data; and have substantively revised the manuscript.

3. **In which chapter(s) of your thesis can this material be found?**

   Chapter 3

4. **e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

   Candidate

   Ruihan Wu

   Date: 3rd March 2024

   Supervisor/ Senior Author (where appropriate)

   Sarah White

   Antonia Hamilton

   Date: 3rd March 2024

UCL Research Paper Declaration Form (3/3)

This form references the doctoral candidate's own published work(s)

1. **For a research manuscript that has already been published**

j) **What is the title of the manuscript?**

Can group membership modulate the social abilities of autistic people? An intergroup bias in smile perception

k) **Please include a link to or doi for the work**

https://doi.org/10.1016/j.cortex.2023.12.018

l) **Where was the work published?**

Cortex

m) **Who published the work?** (e.g. OUP)

Elsevier

n) **When was the work published?**

13th February 2024

o) **List the manuscript's authors in the order they appear on the publication**

Ruihan Wu, Antonia F. de C. Hamilton, Sarah J. White

p) **Was the work peer reviewed?**

Yes

q) **Have you retained the copyright?**

Yes

r) **Was an earlier form of the manuscript uploaded to a preprint server?** (e.g. medRxiv). If 'Yes', please give a link or doi)

Yes

https://doi.org/10.31234/osf.io/8x5zf

If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

□

*I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.*

2. For multi-authored work, please give a statement of contribution covering all authors

Ruihan Wu: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. Antonia F. de C. Hamilton: Conceptualization, Data curation, Investigation, Methodology, Supervision, Visualization, Writing – review & editing. Sarah J. White: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Visualization, Writing – review & editing.

3. **In which chapter(s) of your thesis can this material be found?**

   Chapter 4

4. **e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

Candidate

Ruihan Wu

Date: 3rd March 2024

Supervisor/ Senior Author (where appropriate)

Sarah White

Antonia Hamilton

Date: 3rd March 2024

Acknowledgements

I would like to sincerely thank and acknowledge those people who, in one way or another, helped and supported me during my PhD journey, and contributed to the work presented in this thesis.

I am indebted to my incredible primary supervisor, Dr Sarah White, who has always been full of inspiration, wisdom and compassion. Her unconditional support and encouragement as well as her everlasting patience throughout my PhD and her invaluable feedback and comments have always been extremely beneficial. I have greatly enjoyed working with her. Not only being a great mentor, Sarah also cared for my well-being and always promoted a healthy work-life balance. As an example, inspired by the fact that she is a great dancer, I fell in love with dancing!

I am deeply grateful to my secondary supervisor, Prof Antonia Hamilton, for guiding me to the world of social neuroscience. She had always been generous with her time to help and give me feedback. I learned a wealth of skills, techniques and insightful ideas from her, which I believe will bring me tremendous benefits in my future research. I feel extremely honoured to have the opportunity to work with these two exceptional scientists.

I sincerely thank the UCL's Research Excellence Scholarship (UCL-RES; also known as the UCL Overseas Research Scholarships) and the CSC PhD Scholarship for funding my PhD.

My sincere thanks also go to lab members in the Developmental Diversity group and Social Neuroscience group for a fantastic, supportive and inspirational environment. In particular, I want to thank Dr David Ruttenberg, Dr Ishita Chowdhury, Nevin Ozden, Dr Kat

Ellis, Maryna Nosyk, Beth Jones, Gillian Hughes, Anushay Mazhar, Isla Jones, Dr Sara De Felice, Dr Tian Ye, Dr Roser Canigueral, Paula Wicher, Dwaynica Greaves, Gal Rozic, Emre Yavuz, Laina Townsend, Francesco Di Cio', Magdalena Matyjek, Karen Leow, Nicole Yu, Ciara Rafter, Katia Rosenbaum, Jing Tian Lim, Zahra Ahmed, Ensar Acem, Tina Markandaier, Lina Erilli, and so many past and present lab colleagues. They have all been so unbelievably kind and helpful.

Thank you to all who assisted in the technical aspects of this thesis, in particular Dr Paola Pinti, Prof Ilias Taschtsidis, Uzair Hakim and Natalie Gunasekara. Their vital assistance and support in implementing and analysing the fNIRS study in this thesis made the chapter in its presented form. Thank you also to Dr Harry Farmer, Dr Lucy A Livingston, Prof Punit Shah and Prof Francesca Happé for sharing your expertise and valuable experimental materials.

Very special thanks to my dear thesis buddies Isla Jones and Dr David Ruttenberg, thank you so much for working with me and making the thesis writing much more enjoyable and efficient. I am also grateful to the ICN community and staff in professional services for creating such a supportive work and research environment.

Thanks must go to all the autistic and non-autistic participants and their families for dedicating their time, effort and insight to understanding autism and neurodivergent social cognition. It was a great joy to work with. This research could not have been completed without them.

I can find no words in the world to express my gratitude to Dr Rong Yao, Dr Xiao Fu, Qiurui Feng, Dr Sarah White, Prof Sam Gilbert, Prof Antonia Hamilton, Nova Matthias, Jenny L C Geh, Dr Pei-Chun Tsai, the UCL Doctoral School and many practitioners in the

NHS. In 2022 which was supposed to be the final year of my PhD, my right knee was severely injured. There were 6 months I could barely walk, and still need knee pads until now. My life was in serious danger because of the long-term open wound. Without the help of these kind people before and after my surgery and during my long and painful rehabilitation, I would have lost my life or gotten irreversible injury. Thank them so much for saving my life and helping me return to my life and my beloved research.

Thanks to my amazing friends whom I know from research and hobby (dancing) for making the PhD journey such an enjoyable and memorable part of my life. There is no space to mention you all, but you know how much I love you.

Thanks also to my family for supporting me since always, for giving me the life opportunities that got me here today. It is a privilege to come from a loving and smart home, and I am grateful for that. Thanks to Andrea, who started and ended this journey with me, and for being together even when we were not. I couldn’t ask for a more fun, loving, smart and supportive partner.

Last but not least, I want to thank my mom and dad. They are my life mentors and my friends. I thank them for their unconditional love and support. I was fortunate to grow up in a loving and encouraging family that gave me the courage to overcome fear and illness and complete this long and lonely journey one step at a time. I would like to dedicate this thesis to them. I hope you are proud. I love you, both!

最后，我想感谢我的妈妈爸爸。他们是我的人生导师和我的朋友。感谢他们对我无条件倾尽所有的爱与支持。我很幸运成长在一个充满爱与鼓励的家庭，使我有勇气战胜恐惧和病痛，一步一个脚印完成这个漫长而孤寂的旅程。我想把这篇论文献给他们。我希望你们感到骄傲。我爱你们。

Ruihan Wu

3rd March 2024

London

Table of Contents

Table of Figures

Table of Tables

Abbreviations

AU = Action Unit

BAP = Broader Autism Phenotype

CBSI = correlation-based signal improvement

COVID-19 = Coronavirus disease 2019

dlPFC = dorsolateral prefrontal cortex

DLS = differential looking score

EEG = electroencephalogram

EVC = extrastriate visual cortex

fMRI = functional magnetic resonance imaging

fNIRS = functional Near-infrared spectroscopy

HbO_2 = oxygenated hemoglobin

HbR = deoxygenated hemoglobin

HRF = Hemodynamic Response Function

IFG = inferior frontal gyrus

MFG = middle frontal gyrus

MNS = mirror neuron system

mPFC = medial prefrontal cortex

PCG = postcentral gyrus

pSTS = posterior superior temporal sulcus

SIMS = Simulation of Smiles

STC = superior temporal cortex

TS = temporal sulcus

ToM = Theory of Mind

TPJ = Temporoparietal junction

Chapter 1. General Introduction

To be able to navigate the complex social world effectively, we decipher social cues and formulate inferences about others' thoughts on a daily basis. This ability to understand others' mental states is known as mentalizing (or Theory of Mind; Leslie, 1987; Premack & Woodruff, 1978). People who exhibit difficulties in comprehending others' thoughts, such as autistic people, can experience challenges in their endeavours to communicate and engage with others. Yet, the evidence for mentalizing difficulties in autism is mixed in the literature. Further, people do not equally mentalize about every single person they encounter every day. Mentalizing is not immutable, it can be enhanced or suppressed by a range of contextual and individual factors. In a bustling thoroughfare, for instance, one might be less inclined to ponder the mental processes of a stranger on the opposite sidewalk, whose likelihood of engaging in further interactions with oneself is minimal. However, we know very little about what factors can modulate mentalizing, what processes in mentalizing are capable of being modulated, and whether modulating these factors could facilitate mentalizing in autistic people and thus improve their life quality.

My PhD thesis aims to accurately assess mentalizing abilities in autism (Chapters 2 and 3) and investigate potential factors that may modulate mentalizing (Chapters 3, 4, and 5). In this introductory chapter, I first point out the difficulties in identifying autism under the current diagnostic framework. I then introduce definitions and the theoretical framework of mentalizing. I also review the evidence in autism and highlight the key methodological and theoretical challenges in existing paradigms from the literature. In the second part, I identify factors that modulate mentalizing and emotion processing in non-autistic people, and which may have similar modulation effects in autistic people. In the third part, I present neuroimaging studies to identify the corresponding neural correlates of these modulation effects. In the last part, I describe how my thesis attempts to deal with some of the challenges and gaps mentioned in the first three parts and outline the rationale for the four studies covered in the following chapters.

1.1 Mentalizing and Autism

1.1.1 Identifying Autism is Challenging

Identifying autism is not always easy. Some autistic individuals do not receive a timely diagnosis, and presumably others are not being identified at all. Indeed, adult diagnosis is increasingly common (Lai & Baron-Cohen, 2015; Lehnhardt et al., 2013; Lewis, 2016; Stagg & Belcher, 2019), individuals with higher IQs are diagnosed later than those with lower IQs (Baio, 2014; Gillberg, 1998; Hofvander et al., 2009), and females on average receive their diagnosis considerably later than males and are more likely to have been previously misdiagnosed (Begeer et al., 2013; Lai & Baron-Cohen, 2015; Leedham et al., 2020; Rutherford et al., 2016; Willey, 2014). These have likely resulted from the current autism diagnostic framework, which still relies on clinicians' interpretation of behaviours through observation and/or parental reports (Livingston & Happé, 2017).

Moreover, Frith (2004) argued that the reduction in the severity of autistic symptoms across development and the great heterogeneity in behavioural symptoms are not necessarily genuine remediation or delayed maturation. Instead, some individuals with autism may develop coping strategies to demonstrate fewer behavioural symptoms, despite the persistent existence of cognitive deficits (Frith, 2004, 2013; Hull et al., 2017; Livingston & Happé, 2017). Livingston and Happé (2017) proposed the compensation framework that autistic individuals can compensate for their cognitive deficits to improve their behavioural presentation, with no genuine remission in cognitive and/or neural levels. Therefore, there should be a mismatch between behavioural performance and underpinning cognition and/or neural basis in autism (Livingston & Happé, 2017). Similarly, the camouflaging framework suggests autistic people use conscious or unconscious coping strategies and techniques to minimise the visibility of their autistic characteristics and to appear socially competent (Hull et al., 2019; Hull et al., 2017; Lai et al., 2011; Lai et al., 2017) in the non-autistic society. Additionally, compensation (or camouflaging) is related to stress, mood and self-esteem levels; long-term, unsuccessful and strenuous usage may compromise mental health, thus the outcome of compensation can be toxic (Hull et al., 2019; Hull et al., 2017; Lai et al., 2017; Livingston & Happé, 2017).

The possibility of compensation makes identifying autism even more difficult because those individuals who receive a diagnosis are likely to be least able to compensate or to have particular characteristics that cannot easily be camouflaged (Dworzynski et al., 2012; Hull et al., 2017; Livingston & Happé, 2017; Mandy & Tchanturia, 2015). This does not necessarily mean that the neurodevelopmental difficulties of those who can circumvent diagnosis have genuinely remitted at a cognitive level; rather, these difficulties persist and such behavioural compensation comes at a great cost to mental health (Bargiela et al., 2016; Hull et al., 2017; Livingston & Happé, 2017; Portway & Johnson, 2005). It is therefore critical to develop more sensitive assessments that are not susceptible to compensation and that target underlying cognitive ability (Frith, 2012).

Another idea that is also related to the compensation issue in autism diagnosis is the Broader Autism Phenotype (BAP). The BAP indicates a collection of sub-clinical expressions of autistic traits (Green et al., 2019; Ingersoll, Hopwood, et al., 2011; Piven et al., 1997; Sucksmith et al., 2011; Wainer et al., 2011). BAP populations have similar social cognition challenges as autistic people (Gliga et al., 2014; Green et al., 2019; Rea et al., 2019), but can compensate for those difficulties at a behavioural level (Livingston et al., 2019), which may potentially cause missed or late diagnosis (Mandy & Tchanturia, 2015). However, how social cognition difficulties might be compensated in BAP populations has yet to be fully understood (Green et al., 2019; Livingston & Happé, 2017). As mentioned earlier, missed or late diagnosis occurs more often in females than in males (Hull et al., 2020; Lai et al., 2017; McQuaid et al., 2022; Wood-Downie et al., 2021). It is therefore essential to explore BAP females' socio-cognitive functioning (Green et al., 2019; Ingersoll, Hopwood, et al., 2011), which could, in turn, improve understanding of the endophenotypes of autism (An et al., 2021; Billeci et al., 2016; Palmen et al., 2005). As the BAP is especially prevalent in the relatives of autistic people (Green et al., 2019), one way to identify BAP females is as the mothers of autistic children. BAP mothers are also more vulnerable to mental health problems, such as depression and anxiety, compared with non-BAP mothers (Carpita et al., 2020; DeMyer, 1979; Ekas et al., 2010). If BAP mothers engage in greater compensation, the heightened mental health problems in BAP relatives may result from the cost of compensation (Livingston & Happé, 2017). I will look at BAP mothers' social cognition and compensation in Chapter 3.

1.1.2 Mentalizing Difficulties in Autism

The social difficulties characterising autism (American Psychiatric Association, 2013) have been suggested to result from mentalizing (or Theory of Mind) difficulties (Baron-Cohen et al., 1985; Brüne & Brüne-Cohrs, 2006; Leslie, 1987; Wellman et al., 2001). Mentalizing is the ability to attribute mental states (e.g. belief, intention, desire) to the self and others to explain and predict behaviours (Baron-Cohen et al., 1985; Leslie, 1987; Premack & Woodruff, 1978). It is thought to consist of two systems: explicit mentalizing allows for a deliberate consideration of mental states, which is cognitively demanding and operates in a slow, flexible and conscious way; while implicit mentalizing allows for the efficient processing of mental states in a fast, rigid and unconscious way (Apperly & Butterfill, 2009). The implicit pathway was suggested to exist since infancy, presumably present across the lifespan in parallel with the later-developing (at around 4-years-old) explicit mentalizing (Apperly & Butterfill, 2009; Happé et al., 2017). Accordingly, some autistic individuals without verbal impairments may acquire the capacity to explicitly mentalize through compensatory learning (Frith, 2004), but struggle to implicitly attribute mental states (Senju et al., 2009). For example, when we watch a movie, we may spontaneously engage in implicit mentalizing to understand the protagonist's mental states behind their behaviours; whereas we may engage in explicit mentalizing when we are explicitly asked to tell what the protagonist thought. This ability allows people to understand everyday social contexts, thus the integrity of mentalizing ability is crucial for the effectiveness of social communication and interaction (Frith & Frith, 2006). Therefore, it is important to better understand mentalizing to aid in identifying autism, help design more appropriate supports, and improve the lives of autistic people and their families.

1.1.3 Implicit vs. Explicit Mentalising Tasks

Although some autistic adults perform less well than their non-autistic counterparts on mentalizing tasks (Happé, 1994; White et al., 2009), many autistic children and adults with greater verbal ability can pass mentalizing tasks (Baron-Cohen et al., 1999; Bowler, 1992; Happé, 1995; Steele et al., 2003). This may relate to the type of mentalizing system that each task design taps into. It has been suggested that some autistic people without language difficulties may acquire the capacity to explicitly mentalize through compensatory learning, but still struggle to spontaneously attribute mental states (Frith, 2004). In explicit mentalizing tasks, participants are encouraged to deliberately reason about mental states because these tasks involve direct questioning and require verbal responses. Thus, apart from explicit mentalizing, these tasks could also rely on language (Happé, 1995) and other cognitive abilities, such as executive functions (Abell, 2000) and memory (Ullman & Pullman, 2015). Implicit mentalizing paradigms were developed to bypass this issue to reveal the ability to spontaneously and quickly reason about others' mental states (Clements & Perner, 1994) by using more objective measurements, like eye movements (Southgate et al., 2007) and reaction time (Kovács et al., 2010).

1.1.4 False-belief Reasoning: Implicit Anticipatory-looking Paradigm

To target potential underlying implicit mentalizing difficulties in autism, implicit assessments hold promises as they are less susceptible to compensation. Among various implicit tasks, Southgate et al. (2007)'s non-verbal anticipatory-looking paradigm has been considered that can detect more subtle false-belief reasoning than traditional explicit, and even than some other implicit, paradigms (Hayashi et al., 2020; Southgate et al., 2007). In this paradigm, participants first watched two familiarization trials to set up the contingency, in which a puppet hid an object in the left (see *Figure 1.1a*) and the right box (see *Figure 1.1b*) respectively, and then an agent retrieved the object from the box by reaching through the corresponding window in an occluding screen between herself and the boxes after the windows illuminated. Next participants watched one of the two false-belief conditions. In both conditions, the puppet first placed the object in the left box, second moved it to the right box, and third removed it from the scene. In the false-belief 1 condition, the agent turned her back to the scene before the third step and turned back to face the scene before the windows illuminated (see *Figure 1.1c*), whereas in the false-belief 2 condition, she turned before the second step and back before the illumination (see *Figure 1.1d*). Thus, the agent in the false-belief 1 condition held a false belief that the object was in the right box, while she believed the object was in the left box in the false-belief 2 condition. Eye movements were recorded to assess which window participants expected the agent to reach through. Southgate and colleagues found that non-autistic infants made eye movements toward the window/box that were consistent with the agent's false belief about the object location (belief-congruent), indicating an ability to represent others' false-beliefs. Although a considerable number of infant studies have not replicated Southgate et al. (2007)'s findings and the authors now argue that this paradigm should not be used with infants (Kampis et al., 2020), substantial evidence supports the idea that it can reliably detect mentalizing in adults (reviewed in Schneider et al., 2017).

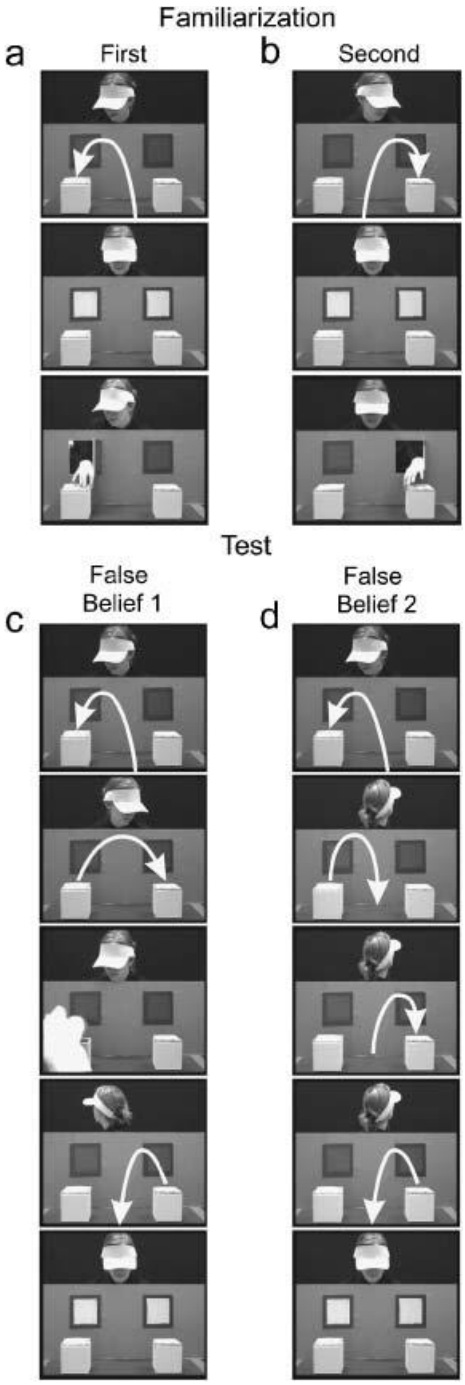


Figure 1.1. **Selected frames from the events used in Southgate et al. (2007)'s paradigm.**

Senju et al. (2009) provided the first evidence for a dissociation between implicit and explicit mentalizing task performance in autism. They compared spontaneous mentalizing between 19 autistic adults and 17 non-autistic adults using Southgate et al. (2007)'s paradigm. Differential looking score (DLS) was calculated, by dividing the difference of the total looking duration between the belief-congruent area (where the agent believed the object was) and the belief-incongruent area by the sum of the two, to measure looking bias. They found that autistic adults' looking behaviour was not biased by the agent's false belief (see *Figure 1.2*), indicating that they were not spontaneously mentalizing, despite performing comparably to their non-autistic counterparts on explicit mentalizing tasks. Presumably, in the latter instance, autistic adults may 'hack' the solution through compensatory strategies, such as linguistic abilities or executive functions (e.g. Abell, 2000; Eisenmajer & Prior, 1991; Frith, 2004; Happé, 1995; Hull et al., 2017; Livingston & Happé, 2017; Ullman & Pullman, 2015). However, although studies with similar paradigms replicated the finding that autistic children and adults have difficulties with implicit mentalizing but not explicit mentalizing (e.g. Schneider et al., 2013; Schuwerk et al., 2016; Senju et al., 2010), this promising finding has been challenged in terms of the reliability of the paradigm, but also in the interpretation of the data (e.g. Burnside et al., 2018; Heyes, 2014; Kulke, Wübker, et al., 2019).

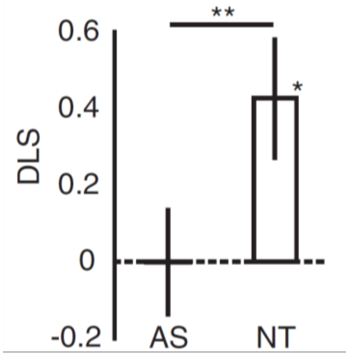


***Figure 1.2.* Mean differential looking score of autistic and non-autistic participants in Senju et al. (2009).**

1.1.4.1 Replication Problem

As already mentioned, substantial evidence supports the idea that the anticipatory-looking paradigm can reliably detect implicit mentalizing in adults (Schneider, Bayliss, et al., 2012; Schneider et al., 2017; Schuwerk et al., 2018). However, a considerable number of infant studies have not replicated Southgate et al. (2007)'s finding that 2-year-old non-autistic children can spontaneously appreciate others' false beliefs and have argued that this paradigm should not be used with infants (Dörrenberg et al., 2018; Kampis et al., 2021; Schuwerk et al., 2018).

Moreover, Kulke and colleagues conducted a series of replication studies to detect implicit mentalizing in non-autistic children and adults, closely following Southgate et al. (2007)'s paradigm, which involved two subtly different but conceptually similar false-belief trial types. Kulke, Reiß, et al. (2018) replicated the anticipatory looking bias for the false-belief 1 condition in all age groups, but the false-belief 2 only in young adults; Kulke, von Duhn, et al. (2018) replicated the results for the false-belief 1, but not false-belief 2, condition; and none of the other studies successfully detected implicit mentalizing in children or adults at all (Kulke, Johannsen, et al., 2019; Kulke, Wübker, et al., 2019). The two false-belief conditions are similar in nature, they only differ when the agent turns her back to the scene which leads to different false beliefs held by the agent (see section 1.1.2.2). Accordingly, they suggested that there might not be spontaneous/implicit mentalizing, or that it exists but is hard to detect by anticipatory-looking paradigms.

1.1.4.2 Three Challenges of Southgate et al. (2007)'s Paradigm

Three specific challenges have been made about the reliability of Southgate et al. (2007)'s paradigm, and several studies have endeavoured to overcome them. Because of these challenges, poor performance on the commonly reported outcome measures (i.e., the anticipatory looking bias towards the belief-congruent area) of this task alone cannot be used to conclusively deduce that autistic individuals have difficulties in spontaneous mentalizing. Several studies have endeavoured to improve the reliability of Southgate et al. (2007)'s paradigm, however, the empirical results from these studies are mixed.

Single-trial design. Each participant was only presented with one test trial, either the false-belief 1 or the false-belief 2. This single-trial design is problematic, because trial-by-trial variation is particularly large between participants, which escalates error variance. The single trial design also exacerbates the dropout rate. For example, 44% of participants in Southgate et al. (2007)'s infants, although only 3% in Senju et al. (2009)'s adults, were excluded due to failure to meet inclusion criteria, missed key events or prediction period, or technical issues, which attenuates reliability (Dang et al., 2020; Kulke, Wübker, et al., 2019). A multi-trial design can improve the signal-to-noise ratio and increase power, allowing for a better estimation of individual performance. With such a design, Schneider, Bayliss, et al. (2012) found that implicit mentalizing can be sustained over the course of a multi-trial procedure lasting about an hour.

No true-belief condition. Both Southgate et al. (2007) and Senju et al. (2009) only presented a false-belief condition; no matched true-belief control condition, in which the actor's beliefs should be consistent with reality, was included as a baseline. This lack of control condition opens the door to alternative explanations. One prominent example is by Heyes (2014), who proposed that non-autistic individuals pass this task due to submentalizing abilities rather than mentalizing. Specifically, the submentalizing hypothesis claimed that non-autistic individuals exhibit correct anticipatory looking in false-belief trials because they get distracted by the agent's head turning, and therefore do not pay attention to, or remember, the subsequent object displacement. Heyes (2014) therefore argued that non-autistic individuals predict the agent's action based on their own false belief of the object's location, rather than the agent's false belief. Additionally, she claimed that autistic individuals are less distracted by the agent and hence know the object is not in either box but are simply less likely to predict people's actions. Accordingly, including true-belief conditions that closely match false-belief conditions and providing a detailed analysis of eye movements throughout the paradigm can examine the submentalizing hypothesis. Based on Heyes (2014), there should be differences between autistic and non-autistic people in visual attention to the key events. Specifically, at the onset of the head-turn period in both false-belief and true-belief conditions, non-autistic individuals should attend and therefore fixate more on the agent but less on the puppet moving the object than autistic individuals. Also, autistic individuals should be less likely to predict the agent's action in both false-belief and true-belief trials.

However, the results from studies implementing true-belief conditions are mixed. It has been found that non-autistic infants and adults were able to attribute both true beliefs and false beliefs with low cognitive demands (Surian & Geraci, 2012; Wang & Leslie, 2016); but, with the same paradigm, Kulke, Reiß, et al. (2018) did not find positive correlations between the two in any age groups. These might indicate that the true-belief conditions were not well matched with the false-belief conditions, or that different strategies or cognitive abilities were used between the two conditions. Gliga et al. (2014) used a familiarization trial in Southgate et al. (2007)'s paradigm as a true-belief condition and concluded that siblings of autistic children were able to attribute others' true beliefs, but not false beliefs. As they found there was no group-specific pattern of attention on the key events in their false-belief condition, it seems like their results can be better explained by the mentalizing difficulty in autism than the submentalizing hypothesis.

Non-evaluative context. The anticipatory-looking paradigm might not be sufficiently engaging to elicit implicit mentalizing, which is intrinsically a social ability (Kulke & Hinrichs, 2021; Kulke, Johannsen, et al., 2019; Schuwerk et al., 2018; Woo et al., 2023). Indeed, half of the children in Southgate et al. (2007), 35-50% of adults in Kulke, Johannsen, et al. (2019), and 70% of data in Schneider et al. (2013) were excluded due to failure to predict actions. Thus, Kulke, Johannsen, et al. (2019) called for creating more engaging implicit paradigms to encourage mentalizing. To make Southgate et al. (2007)'s paradigm more engaging for children, Kulke and Rakoczy (2019) added verbal narrations of the events to the original non-verbal videos, and Kulke and Hinrichs (2021) moved the entire task to a more realistic social scenario; however, none of them replicated the original findings.

Although the replication was unsuccessful, Kulke and Hinrichs (2021) argued that when observers know there would not be any social consequence to not anticipating the actor's action, reasoning about her mental state is less likely to be prioritized. The importance of social context is consistent with Woo et al. (2023)'s suggestion that socially evaluative contexts can facilitate mentalizing, defined as contexts where agents' actions have interactive potential, including both prosocial and antisocial. They further proposed that the mixed results in replications using Southgate et al. (2007)'s paradigm may be because those studies have only detected false-belief reasoning within non-evaluative contexts, which provide observers less reason to care about agents' mental states, as their actions are irrelevant. Thus, it is necessary to develop a more evaluative implicit mentalizing paradigm to assess whether social contexts can facilitate mentalizing.

1.1.5 Smile Discrimination: an Alternative Way to Index Mentalizing

In addition to false-belief reasoning, discriminating between genuine and posed emotional expressions can be an alternative way to index mentalizing. Emotional expressions can be isolated from genuine feelings to a certain extent for a variety of purposes (e.g., Ekman, 2003; Hess et al., 1997; Lazarus, 1991; Niedenthal et al., 2010; Rosenberg & Ekman, 2020). Thus, understanding others' emotional expressions is essential in social interaction and communication, which gives clues about their affective states and intentions.

People spontaneously evaluate the authenticity behind others' emotional expressions (Cosme et al., 2021). It has been suggested that mentalizing plays an important role in distinguishing between genuine and posed emotional expressions (Cosme et al., 2021; Lavan et al., 2017; McGettigan et al., 2015; Szameitat et al., 2010). For example, three studies consistently found that the mentalizing network (i.e., *anterior medial prefrontal cortex, amPFC*) (Frith & Frith, 2006) was engaged more strongly for posed than genuine laughter during passive listening (Lavan et al., 2017; McGettigan et al., 2015; Szameitat et al., 2010). They concluded that people involuntarily reason about others' mental states when an expression is perceived as posed, even though this conclusion may suffer from reverse inference issues, and the *amPFC* has also been identified to be involved in other higher-order cognitive processes. They further suggested that it is the social-emotional ambiguity of posed expressions that engages mentalizing to a greater degree.

From an ethical perspective, I specifically chose smiles in my thesis because they are positive facial expressions. Smiles would not cause any potential risks or adverse effects, like psychological stress or distress. In contrast, watching smile videos may release the stress results from the worldwide lockdown, as the corresponding study reported in Chapter 4 was initiated and carried out during the COVID-19 pandemic.

1.1.5.1 Genuine and Posed Smiles

Smiles are important social cues but are not always a reliable indicator of affective states (e.g., Ekman, 2003; Lazarus, 1991). Genuine (or Duchenne) smiles are considered to be spontaneous and associated with enjoyment emotions, while posed (or non-Duchenne) smiles are not necessarily related to positivity but act as purposeful communication tools in social situations and can be a potential signal that the real emotional state is obscured (Ekman, 2003; Krumhuber et al., 2007). Biland et al. (2008) suggested that posed smiles might indicate deception, which may need mentalizing ability to reason the hidden intentions. As decoding genuine and posed smiles can be associated with the ability to reason about another's mental state (Boraston et al., 2008), the accuracy of distinguishing between the two types of smiles can be used as an indicator of mentalizing ability.

There are two ways to differentiate genuine and posed smiles: one according to subjective perception, the other one based on muscle activation. It has been found that smilers expressing genuine smiles are perceived as more cooperative, likeable and trustworthy, as well as less disingenuous and misleading than those displaying posed smiles (e.g., Biland et al., 2008; Ekman, 2003; Frank & Ekman, 1993; Johnston et al., 2010; Krumhuber et al., 2007; Mehu et al., 2007; Schug et al., 2010). Similarly, genuine smiles are intrinsically more rewarding than posed smiles, so people prefer the former over the latter and are willing to offer a higher monetary value to receive genuine than posed smiles (Shore & Heerey, 2011). However, these subjective feelings can be very subtle and vary from person to person to a great degree (Hess et al., 1997), so it can be difficult to discern such nuanced displays only based on subjective perceptions.

The other most robust yet subtle feature that differentiates genuine from posed smiles is muscle activation because there is a physical reality to the differences between the two types of smiles. Genuine smiles tend to involve a spontaneous contraction of both the zygomatic major (i.e., AU12; in the cheek) and the orbicularis oculi (i.e., AU6; around the eyes) muscles, whereas posed smiles only involve a deliberate activation of the AU12 (Duchenne & de Boulogne, 1990; Ekman et al., 1990; Ekman & Friesen, 1982).

1.1.5.2 Smile Discrimination in Autism

Accurate differentiation between, and response to, genuine and posed smiles is challenging but also an essential ability to effectively cope with the complexity of social interaction and communication (e.g., Blampied et al., 2010; Boraston et al., 2008; Ekman, 2003; Lazarus, 1991; Song et al., 2016; Young et al., 2015). Difficulties recognising and responding to others' emotional states are intrinsic to the definition of autism, both historically and currently (ICD-11; World Health Organization, 2018) and have long been suggested to be a central feature of autism (Hobson, 1986). However, a substantial body of behavioural research into facial emotion recognition in autism has produced mixed findings, with some studies showing difficulties (e.g., Uljarevic & Hamilton, 2013) but others showing seemingly typical responses (e.g., Cook et al., 2013; Ketelaars et al., 2016). Moreover, autistic people are more accurate than those with attention deficit hyperactivity disorder (ADHD) in recognizing basic emotions, such as happiness, sadness, anger, fear and disgust (Downs & Smith, 2004; Sinzig et al., 2008). One suggested explanation of these observations is that autistic adults are able to recognize basic facial emotions (Castelli, 2005), but struggle to identify more complex and subtle facial emotions, such as embarrassment and guilt (for reviews, see Harms et al., 2010; Liu & Humpolíček, 2013), but also differentiating genuine from posed smiles (Boraston et al., 2008). More detailed studies of the perception of these subtly different expressions are therefore important for understanding autistic social communication. However, it remains unclear whether such difficulties are characteristic of autism per se or are, for example, related to alexithymia, a commonly co-occurring condition affecting emotion recognition (Cook et al., 2013; Dyck et al., 2001; Hill et al., 2004; Salminen et al., 1999; Shah et al., 2016).

To the best of my knowledge, only two studies have investigated this ability in autism, but all had small sample sizes which may compromise their statistical power. Boraston et al. (2008) compared the ability of 18 autistic and 18 non-autistic adults to distinguish between genuine and posed smiles from static images. They found autistic adults were less accurate in discriminating the two smile types than non-autistic adults, but they were just as good at discriminating between neutral and smiling faces. Additionally, in the autism group, the ability to differentiate the two types of smile was negatively associated with the degree of social communication difficulties measured by the ADOS (Lord et al., 2000): the more severe the social difficulties, the more affected the smile discrimination ability. Boraston et al. (2008) suggested that failure to decode these subtle social cues is likely to be associated with reasoning about another’s mental state and could lead to social difficulties in autism.

Blampied et al. (2010) compared the smile discrimination ability between 8 autistic and 11 non-autistic boys matched on chronological age and sex using face images displaying a neutral expression, a genuine smile or a posed smile. They first verified each child understood the difference between looking and feeling happy by giving them an example and asking them to provide an example. Then, each child watched a set of 18 pictures and answered whether each target was ‘looking happy’ and a further set of 18 pictures and answered whether the target was ‘feeling happy on the inside’ by touching a YES or NO button on a touch screen. They observed a group difference between autistic and non-autistic groups in smile discrimination, similar to Boraston et al. (2008)’s finding, but this group effect was marginal (*p* = .09). Specifically, compared with non-autistic boys, autistic boys were less sensitive to the difference between genuine and posed smiles. However, social communication ability as measured by the Social Communication Questionnaire (Rutter et al., 2003) was not related to the sensitivity to make subtle distinctions between the two smile types.

Using a similar smile discrimination task as Boraston et al. (2008), Heerey (2014) and Manera et al. (2011) recruited 45 and 120 non-autistic adults respectively to investigate the relationship between smile discrimination ability and autistic traits as measured by the autism-spectrum quotient (Baron-Cohen, Wheelwright, Skinner, et al., 2001). Consistent with Blampied et al. (2010)'s finding, both did not observe any relationship between individual differences in recognizing smile authenticity and autistic traits. Because of the limited and mixed results in the literature, it remains unclear whether smile discrimination ability contributes to the social communication difficulties in autism.

1.2 Modulating Mentalizing and Neural Mechanisms

Given the observation of mentalizing difficulties in autism (i.e., false-belief reasoning and smile discrimination), it is important to study whether it is possible to increase or decrease mentalizing generally, and further whether it is possible to modulate mentalizing, specifically to increase mentalizing, in autistic people. As discussed in section 1.1.4.2, it appears that context is important for mentalizing to occur, and therefore context may be capable of modulating mentalizing. In addition to the evaluability of context, other possible contextual modulators include intra-personal factors, such as group membership. Therefore, I specifically explore whether evaluative contexts facilitate implicit mentalizing in Chapter 3, and how intergroup bias modulates social cognition in autism and the corresponding neural correlates in Chapters 4 and 5.

1.2.1 What is Intergroup Bias

Exploring factors that may modulate mentalizing might not only help explain some of the variations in difficulties autistic people experience but also highlight possible ways to make social interactions easier to navigate. One potential factor is intergroup bias, which refers to the systematic finding that people tend to favour those who are more similar to themselves (i.e., ingroup members) over those who are less similar to themselves (i.e., outgroup members). This ingroup favouritism is not a conscious choice, people spontaneously preferred to process ingroup over outgroup information (reviewed in Scheepers & Derks, 2016). Intergroup bias can be generated not only when the group boundary is definite in the real world, such as gender and race (e.g., Montagu, 1997; Rudman & Goodwin, 2004), but also when it is completely arbitrary and people are randomly assigned to one of two mutually exclusive groups (i.e., minimal group; e.g., Allen & Wilder, 1975; Doosje et al., 1995; Howard & Rothbart, 1980; Tajfel, 1970).

Ingroup favouritism can regulate people's perception, evaluation, and behaviours towards ingroup over outgroup (e.g., Balliet et al., 2014; Jordan et al., 2014), with higher ingroup identification resulting in stronger biases (Doosje et al., 1995; Ellemers et al., 2002). For example, people tend to cooperate and share resources more with ingroup members but punish outgroup members more harshly (Balliet et al., 2014; Jordan et al., 2014), and selectively discount negative behaviours from ingroup, but not outgroup, members (Park & Young, 2020). Grounded from an evolutionary perspective, ingroup favouritism serves an adaptive role with multiple benefits that could facilitate building and maintaining intragroup relationships to cope with the complexity of the physical and social worlds (Park & Young, 2020). However, on the other hand, intergroup bias has also been associated with a range of negative outcomes, such as prejudice, discrimination, and even dehumanization of outgroup members (e.g., Borinca et al., 2023; Brewer, 1999; MacInnis & Hodson, 2012).

1.2.2 Intergroup Bias in Social Cognition

Intergroup biases have been shown to affect social cognition, in particular mentalizing. Harris and Fiske (2006) suggested that people attribute fewer mental states to outgroup members than ingroup members. They found that the *medial prefrontal cortex (mPFC)*, a key brain region associated with the mentalizing network, was less activated when viewing images of outgroup members than their ingroup members. Consistently, behavioural evidence also supports this idea. Harris and Fiske (2011) asked adults to imagine and describe a day in the life of their ingroup and outgroup members. They found that adults used fewer mental-state words (e.g., believe, feel, think) for their outgroup members than ingroup members. Similarly, McLoughlin and Over (2017) investigated whether 5- and 6-year-old children spontaneously attribute more mental states to their ingroup than outgroup members. Children were asked to describe the actions of interacting geometric shapes and were led to believe that those shapes constituted their ingroup and outgroup (based on gender and geographic location).

Because the corresponding study reported in Chapter 4 was carried out during the COVID-19 pandemic, it was impossible to use the in-person false-belief reasoning task discussed in section 1.1.4 based on the constrained situation. Therefore, I decided to use the smile discrimination task as an alternative way to index mentalizing, which can be easily adapted to use online. Consistent with Harris and Fiske (2006)'s idea, people tend to be more accurate in decoding basic facial emotions displayed by ingroup members than by outgroup members under definite group boundaries (e.g., Elfenbein & Ambady, 2002) as well as in minimal group settings (e.g., Bernstein et al., 2007; Young & Hugenberg, 2010).

An intergroup bias has been detected in identifying genuine smiles from posed smiles within a minimal group setting. Young (2017) measured participants' ability to tell whether a smile was genuine or posed using videos, as well as their tendency to identify themselves as an ingroup and outgroup member, by randomly assigning them to one of two made-up personality categories. Surprisingly, an outgroup advantage was observed: people were not only more accurate but also faster in differentiating genuine from posed smiles for outgroup than ingroup smilers. Specifically, people were more likely to mistake posed smiles for genuine ones from ingroup members. Young (2017) suggested that ingroup favouritism may explain this effect. More positive feelings towards ingroup members may have biased people to interpret ingroup smiles as genuine even when smiles were posed, and attracted them to look at ingroup faces longer. On the other hand, the wariness of outgroup members may have led to a more vigilant approach to outgroup smiles. However, although people conveyed higher identification with their ingroup than outgroup members, this was not related to their accuracy or speed in determining the smile authenticity.

Xie et al. (2019) looked at intergroup bias in recognizing micro-expressions, fleeting facial expressions lasting up to 500 ms. It has been suggested that micro-expressions can be a reliable indicator of lies (Matsumoto & Hwang, 2018), similar to posed facial expressions (Biland et al., 2008; Ekman, 2003), which might engage mentalizing because of their social-emotional ambiguity (Lavan et al., 2017; McGettigan et al., 2015). In Xie et al. (2019)'s study, 30 Asian participants were asked to identify the facial expressions of 12 White and 12 Asian actors. Each expression was presented for 100 or 333 ms with two 1000 ms presentations of the same actor's neutral expression before and after respectively. After each trial, participants were asked to identify the micro-expression just displayed. Consistent with Young (2017)'s finding, they showed an outgroup advantage in identifying micro-expressions; the recognition accuracy of outgroup members was higher than that of ingroup members. However, they did not separate emotion categories (i.e., sadness, happiness, fear, surprise, anger, and disgust) in their analysis to achieve sufficient statistical power, which opens to the possibility that dropping this factor might cancel out some opposite effects, as different emotion categories might be influenced by intergroup bias differently.

Increasing neuroscience research has begun to unpack the neural correlates underpinning intergroup social influences on a wide range of perceptions, attitudes and behaviours, and found evidence that the human brain perceives and responds differently to ingroup and outgroup information (reviewed in Molenberghs & Louis, 2018; Moradi et al., 2020). It has been suggested that intergroup bias is potentially underpinned by the variation of attentional saliency (e.g., Moradi et al., 2020; Mullen et al., 1992; Schupp et al., 2003). Accordingly, two brain functional systems, the mentalizing network and the executive control network, have been suggested to be involved in social cognition in intergroup settings. The mentalizing network is responsible for identifying and evaluating others' mental states, which is putatively located in the *temporoparietal junction (TPJ), posterior superior temporal sulcus (pSTS)* and *mPFC* (Frith & Frith, 2006; Frith & Frith, 2003; Schurz et al., 2014). The executive control network is necessary for reorienting attention to salient stimuli, which contains but not limited to the *dorsolateral prefrontal cortex (dlPFC)* and *middle frontal gyrus (MFG)* (e.g., Decety & Lamm, 2007; Eberhardt, 2005; Moradi et al., 2020; Mullen et al., 1992; Schupp et al., 2003; Seeley et al., 2007; Smith et al., 2019).

Evidence from neuroscience has shown that the mentalizing network is recruited when intergroup bias alters social cognition. For example, Adams Jr et al. (2010) employed the fMRI technique to look at the neural correlates of racial intergroup bias in mentalizing by using the Reading the Mind in the Eyes Task (Baron-Cohen, Wheelwright, Hill, et al., 2001). Adams Jr et al. (2010) found both Asian and Caucasian participants were more accurate in detecting mental states from their ingroup. They also found greater activation in the *TPJ* area when reasoning the mental state of ingroup members than that of outgroup members. In another example of racial intergroup bias from Katsumi and Dolcos (2018), participants were presented with non-verbal ingroup or outgroup guest-host social encounters, including approach and avoidance conditions. Then, participants were asked to rate the host’s competence and their own interest in interacting with the hosts. Katsumi and Dolcos (2018) found that ingroup members were rated more positively than outgroup members. They also found that the *pSTS* and *mPFC* showed greater activation when observing ingroup than outgroup approach behaviour.

Neuroimaging studies have also observed that the executive control network is involved when group membership modulates social cognition. In the study mentioned above, Katsumi and Dolcos (2018) also compared dynamic social interactions to non-social control scenes where the hosts were replaced with non-interactive cardboard cut-outs. They observed that the *dlPFC*, *extrastriate visual cortex* and *pSTS* engaged higher activation in the social than non-social conditions for ingroup members, indicating the role of these regions in processing non-verbal social behavioural cues. As the involvement of the *dlPFC* has been primarily related to voluntary regulation of spontaneous racial intergroup bias (Bartholow & Henry, 2010), the attenuated activity level in *dlPFC* during the ingroup non-social condition is likely to indicate a reduced executive control and regulatory processes devoted. This context-based mechanism seems to possess an adaptive function. Specifically, when social interaction is absent or when the outgroup is present, the context provides observers less reason to monitor and regulate the information, thus more cognitive resources may be allowed to focus on other more relevant top-down control-related processes and evaluations. This is consistent with the idea explored in Chapter 3 that social evaluative context facilitates social cognition (Woo et al., 2023).

1.2.3 Mimicry, Social Cognition, and Intergroup Bias

People spontaneously and often unconsciously mimic a variety of others' behaviours in social interactions (reviewed in Chartrand & Van Baaren, 2009), which has been suggested to be related to social cognition (e.g., Lee et al., 2023; Niedenthal et al., 2010; Wang & Hamilton, 2012). One of the major theories that has been proposed to explain the role of social mimicry in social cognition is the simulation theory of social cognition (Gallese, 2007, 2009). Simulation theories claim that automatic facial mimicry facilitates understanding of others' mental states, aligns emotions of both sides and improves interaction (e.g., Niedenthal et al., 2010). Substantial research evidence has reported that simulation contributes to accurately and efficiently process features of emotions conveyed by facial expressions. Particularly, restricting movements compromises people's ability to process facial expressions. For example, Stel and Van Knippenberg (2008) asked participants to avoid facial movements which led them to judge the valence of emotional expressions slower. With a similar method, Borgomaneri et al. (2020) more recently replicated this finding and reported that blocking facial mimicry compromises recognition of facial and body expressions. Hennenlotter et al. (2009) applied botulinum toxin to participants' frown muscles to prevent movement and found that reducing facial imitation of angry expressions attenuated the amygdala activity and its functional connectivity with other brain areas.

Niedenthal et al. (2010) claimed that facial mimicry may be especially beneficial for understanding subtle or ambiguous emotion expressions, such as the authenticity of smiles, and therefore proposed the simulation of smiles (SIMS) model. Evidence in favour has reported that facial mimicry is sensitive to smile features like intensity. Korb et al. (2014) found that the degree of participants' smile muscles (i.e., AW6 & AU12) contraction predicted smile authenticity judgments, suggesting that facial mimicry influences smile perception. Similarly, Rychlowska et al. (2014) showed that blocking facial mimicry made participants judge genuine and posed smiles to be equally authentic, suggesting that disrupting mimicry impairs smile discrimination ability.

Simulation theories have been primarily supported by the discovery of the mirror neuron system (MNS). The MNS is a group of specialized neurons, putatively located in the *inferior frontal gyrus (IFG), STS, and inferior parietal cortex (IPC)* (Iacoboni et al., 1999; Rizzolatti & Craighero, 2004), that fire both when the same action is acted and observed (Rizzolatti, 2005; Rizzolatti & Craighero, 2004; Rizzolatti & Sinigaglia, 2016). A growing body of neuroscience literature has revealed mimicry and possibly social cognition rely on the MNS (e.g., Bastiaansen et al., 2009; Heyes, 2011; Krautheim et al., 2019; McLellan et al., 2012; Olsson & Ochsner, 2008; Shamay-Tsoory, 2011; Spunt & Lieberman, 2012). The MNS directly links perception especially visual perception and motor behaviour, specifically observing an action can automatically activate the same neurons when that action is performed (Brass & Heyes, 2005; Rizzolatti, 2005). Thus, according to simulation theories, the MNS plays a fundamental role in mimicry and understanding of action and emotion (Rizzolatti & Craighero, 2004; Rizzolatti & Sinigaglia, 2016; Thompson et al., 2022).

The activation of the MNS is sensitive to social cognition. McLellan et al. (2012) investigated the neural correlates underpinning discriminating genuine and posed facial expressions of happiness (i.e., smile) and sadness using fMRI. Participants were able to identify genuine from posed facial expressions, indicating that they were sensitive to the underlying mental states. The fMRI results showed greater neural activity in response to genuine compared to posed facial expressions in the *IFG*, which may reflect spontaneous mimicry facilitating smile discrimination. However, only 7 females were recruited, the small single-gender sample might compromise the statistical power and the generalizability of their results. Lee et al. (2023) examined the neural mechanism of smile authenticity identification using fMRI with a larger sample (*n* = 44). Consistent with McLellan et al. (2012), they found that accurately identifying genuine from posed smiles activated the *IFG*.

Human mimicry is sensitive to the mimicked targets, specific goals of the current interaction, and social contexts (reviewed in Chartrand & Van Baaren, 2009), for example, intergroup bias (e.g., Krautheim et al., 2019; Peng et al., 2021). Thus, one way that intergroup bias may modulate social cognition is through influencing mimicry. Indeed, empirical evidence has demonstrated that people are more likely to mimic ingroup than outgroup members (e.g., Bourgeois & Hess, 2008; Mondillon et al., 2007; Peng et al., 2020; Peng et al., 2021), which in turn can improve ingroup affiliation (reviewed in Hale & Hamilton, 2016). For example, Bourgeois and Hess (2008) found that people increase facial mimicry toward their ingroup members who are important for their own social standing and benefits compared with outgroup members, also expressions facilitating affiliation (e.g., sad) were mimicked more than expressions endangering affiliation (e.g., anger). Similarly, Peng et al. (2021) replicated that people were more likely to mimic happiness of racial ingroup rather than outgroup members, and this intergroup bias on emotional mimicry was mediated by the perceived degree of interpersonal closeness or self-other overlaps.

The activation of the MNS can be modulated by social contextual factors, such as group membership. Krautheim et al. (2019) explored the effect of intergroup bias on the MNS mechanisms for the perception and production of facial emotional expressions (i.e., happy, angry, and neutral) within a minimal group setting using fMRI. Participants were asked to watch ingroup and outgroup facial expressions and reproduce these expressions themselves. They found enhanced neural activity in the *IFG, MFG* and *postcentral gyrus (the* *location of the primary somatosensory cortex)* for perceiving ingroup compared to outgroup members' facial expressions.

However, it cannot be ignored that a number of researchers have critically assessed the MNS and argued that it may tell us little about social cognition (Borg, 2007; Csibra, 2008; Hickok, 2009; Jacob, 2008). An obvious objection to the mediating role of mimicry between intergroup bias and social cognition is from visual theories of social cognition (e.g., Allison et al., 2000; Kanwisher, 2000). Unlike simulation theories, visual theories propose that people understand actions through visually analysing each element that it consists of and the interactions between elements, which does not require simulation and the involvement of the MNS. For example, when we see someone reaching for a cup of water, the elements would be the person's hand, the cup of water, and the movement of the person towards the cup. By analysing how these three elements are associated and inferencing how they may interact, we can understand the person's action and reason for their intention. According to visual theories, action understanding would be primarily associated with the activity of the superior temporal sulcus (STS) and extrastriate visual areas that selectively respond to body, motion, objects, and interactions between them (Allison et al., 2000; Kanwisher et al., 1997). Visual theories also propose an alternative explanation of the MNS mechanism that the MNS is activated after an action has been understood through visual analysis, thus, instead of contributing action understanding, the MNS reflects action understanding per se (Csibra, 2008; Hickok, 2013).

Moreover, to understand an action, we need to understand the goal and intention behind it (Hickok, 2013). Thus, to analyse visual cues from others, people may develop theories to understand others' minds through causal models and Bayesian learning which does not necessarily involve mimicry (Gopnik & Wellman, 1994, 2012). Indeed, people can understand mental states that they cannot simulate, in other words, simulation alone is insufficient to explain social cognition (Csibra, 2008; Csibra & Gergely, 2007; Gallese et al., 2011; Hickok, 2009; Hickok & Hauser, 2010). Rizzolatti and Craighero (2004) also suggested that there can be other mechanisms, apart from simulation, for supporting action understanding.

1.2.4 Intergroup Bias in Autism

Despite the growing evidence that autistic people are less sensitive to social stimuli (e.g., Dawson et al., 2004; Dubey et al., 2018; Fletcher-Watson et al., 2009; Klin et al., 2002; Steele et al., 2003), their behaviour under intergroup settings has rarely been examined (Qian et al., 2022).

A few recent studies have suggested that intergroup bias is attenuated and even absent in autistic people in studies using definite intergroup boundaries (e.g., nationality) and in non-autistic adults with higher autistic traits in studies using minimal group settings. Qian et al. (2022) investigated intergroup bias in "third-party punishment" behaviours in autistic adults. Participants observed arbitrary ingroup/outgroup proposers making decisions to distribute money to outgroup/ingroup receivers. Then, they were asked to penalise proposers when they violated social norms (i.e., distributed the money unfairly) by removing their money. Qian et al. (2022) found that non-autistic adults penalised outgroup proposers more harshly than ingroup proposers, but this ingroup favouritism was attenuated in autistic adults. With a similar paradigm, Vaucheret Paz et al. (2020) found the effect of intergroup bias was completely absent in autistic children compared with three other neurodivergent groups, children with ADHD, learning disabilities and intellectual disability. With a different paradigm, Uono et al. (2021) observed an attenuated racial intergroup bias in perceiving self-directed gazes (i.e., gazes that look at self) in autistic compared to non-autistic adults. They found, compared with outgroup gazes, ingroup gazes were more likely to be perceived as self-directed gazes in non-autistic adults, but this intergroup bias was absent in autistic adults, even though autistic and non-autistic adults did equally well in distinguishing self-directed gaze from averted gaze.

The cross-race effect, the tendency to recognize own racial (i.e., ingroup) faces more easily than other racial faces, has been considered as a type of intergroup bias manifestation. The intergroup bias in face recognition between own and other races has also been studied in autism, but the findings are inconsistent. For example, some studies examined face discrimination for own- and other-race faces and found that autistic adults (n = 19; Hadad et al., 2019) and autistic children with lower intellectual abilities (n = 77; Kang et al., 2020) showed substantially smaller processing advantage and significantly attenuated specialization for own-race faces than their non-autistic counterparts. However, some other studies with similar methods reported that autistic adults and children showed the same cross-race effect in face recognition as the non-autistic populations (n = 24-29; Wilson et al., 2011; Yi et al., 2016; Yi et al., 2015). Using eye-tracking techniques, Kang et al. (2020) also showed that autistic children generally pay less attention to faces, especially eyes, than non-autistic children which may affect their ability to process face race information. However, some of the other findings showed the same gaze pattern between autistic and non-autistic people alongside a typical ability in recognizing facial identity in autism (Yi et al., 2016; Yi et al., 2015), Thus, intergroup bias should exist in autism and is possible to be detected under the right conditions.

Accordingly, the autism-group identification should in theory cause intergroup bias. Indeed, it has been suggested that autistic people could more easily decode social cues and reason about the mental states of other autistic people than about non-autistic people, and the opposite would be true for non-autistic people (e.g., Fletcher-Watson & Happé, 2019; Komeda et al., 2019). Furthermore, non-autistic people are less successful in understanding autistic targets than their peers' behaviours (Sheppard et al., 2016). This idea is consistent with Milton (2012)'s 'double empathy problem' which suggests that social interaction and communication difficulties are bidirectional. On the one hand, autistic people struggle to navigate in a non-autistic society; on the other hand, it should be equally difficult for non-autistic people to fit into an autistic society. The idea of ingroup favouritism between autistic and non-autistic groups has been partially supported by Sasson et al. (2017) and Alkhaldi et al. (2019) who both reported that non-autistic people rated autistic people as less favourable than other non-autistic people, without knowing who was autistic. This in turn could discourage autistic people from interacting with others (Mitchell et al., 2019). In reality, more than 97% of the general population is non-autistic (e.g., Brugha et al., 2009; Chown, 2014; Li et al., 2022), so the sense of being disfavoured by the non-autistic majority is likely to be harmful (Milton, 2012; Mitchell et al., 2021). The aforementioned evidence raises an important possibility that if we emphasize similarities and inclusion between neurodiverse groups, they might favour and empathize more with each other, and eventually perhaps they could interact with and understand each other better (Mitchell et al., 2021).

Importantly, none of the aforementioned studies measured group identification, so the potential effect of the subjective attitude of autistic people on their group membership is not clear. If they did not feel so closely affiliated with ingroup members, it might not be surprising that an intergroup bias was reduced or absent. Consequently, intergroup bias seems to be a compelling factor that may potentially modulate the accuracy of autistic people when mentalizing, in particular discriminating genuine and posed expressions. However, to the best of my knowledge, this has not been studied in autistic people.

1.3 Overview of Experimental Chapters

In the current chapter, I first stressed the difficulties in identifying autism under the current diagnostic framework and the importance of understanding implicit mentalizing for aiding in autism identification and improving the lives of autistic people and their families. Accordingly, I further reviewed the criticisms in the literature about the challenges in detecting mentalizing using one of the most promising anticipatory-looking paradigms from Southgate et al. (2007), and pointed out the urgent need to develop more sensitive paradigms through addressing these issues. Then, I introduced how contextual information, especially intergroup bias, may modulate mentalizing and emotion processing, and the corresponding neural mechanism. I shed light on the importance of better understanding these modulation effects that may not only help explain the difficulties autistic people experience but also highlight possible ways to make social interactions easier to navigate. There are two major themes in the current thesis. The first theme is to investigate whether implicit mentalizing plays a role in autistic cognition by using and adapting Southgate et al. (2007)'s anticipatory-looking paradigm (Chapters 2 and 3). The second theme is to explore how contextual information (i.e., evaluative context and intergroup bias) would modulate mentalizing (Chapters 3, 4, and 5) by using both the anticipatory-looking paradigm (evaluative context) and a genuine-posed smile discrimination task (intergroup bias). To achieve these aims, I attempt to implement a multimodal approach to disentangle the complexity of measuring and modulating mentalizing and the underpinning neural mechanisms in the following chapters. In particular, I integrate behavioural performance, subjective awareness, individual personality traits, eye movements, facial movements, video recordings, and brain activity recordings.

Specifically, in Chapter 2, I compare implicit and explicit mentalizing abilities between autistic and non-autistic populations. To achieve this, I modify Southgate et al. (2007)'s paradigm by implementing a multi-trial experiment with matched true-belief conditions. I also scrutinize the alternative explanations for implicit mentalizing difficulties in autism proposed in the literature.

Chapter 3 primarily investigates the modulatory effect of contextual information, the evaluability of context, on mentalizing. I developed a more evaluative paradigm that provides more reason for eliciting mentalizing to compare with the same task in a less evaluative context. To do so, a question is added to prompt observers to anticipate agents' actions. I also attempt to identify the differences between mothers of autistic and non-autistic children in implicit and explicit mentalizing abilities, autistic traits, compensatory tendencies and mental health outcomes; and to explore how the aforementioned factors might relate to and predict implicit mentalizing performance.

Chapter 4 investigates another contextual modulation effect on mentalizing – the effect of intergroup bias in discriminating between genuine and posed smiles in autism using a minimal group paradigm. To extend my findings in Chapter 4, I attempt to capture the underlying neural representations of intergroup bias on social cognition using fNIRS in Chapter 5. I implement the same procedure as Young (2017) as well as a non-mentalizing control condition. I also examine whether facial imitation is modulated by intergroup bias, as well as the corresponding neural correlates.

It is worth noting that throughout the thesis 'we' will be used when the experimental team made collaborative contributions to the work, with the default assumption that the candidate Ruihan Wu was the leading researcher who designed and conducted the experiment. Moreover, the studies reported in Chapters 4 and 5 were carried out during the

COVID-19 pandemic, with severe delays and great challenges for the in-person neural imaging study. This not only postponed the whole research programme (e.g., applying for ethical amendments, extra time for sanitizing the testing equipment), but also resulted in redesigning the two studies and making a series of adaptations based on the constrained situation.

Chapter 2. Do Autistic Adults Spontaneously Reason about Belief? A Detailed Exploration of Alternative Explanations

Abstract

Southgate et al. (2007)’s anticipatory looking paradigm has presented exciting yet inconclusive evidence surrounding spontaneous mentalizing in autism. The present study, therefore, aimed to develop this paradigm to address alternative explanations for a lack of predictive eye-movements on false-belief tasks by autistic adults. This was achieved through implementing a multi-trial design with matched true-belief conditions, and both high and low inhibitory demand false-belief conditions. We also sought to inspect if any group differences were related to group-specific patterns of attention to key events. Autistic adults were compared with non-autistic counterparts on this adapted implicit mentalizing task and a well-established explicit task. The two groups performed equally well in the explicit task; however, autistic adults did not show anticipatory-looking behaviour in false-belief conditions of the implicit task. Critically, both groups showed the same attentional distribution in the implicit task prior to action prediction, indicating that autistic adults process information from social cues in the same way as non-autistic adults, but this information is not then used to update mental representations. Our findings further document that many autistic individuals struggle to spontaneously mentalize others’ beliefs. We also discuss alternative theoretical explanations for this pattern of performance, leading to a better understanding of mentalizing mechanisms.

2.1 Introduction

As discussed in Chapter 1, identifying autism is challenging. The current autism diagnostic framework still relies on clinicians' interpretation of behaviours (Livingston & Happé, 2017). This is likely to cause delayed diagnosis or misdiagnosis because some autistic people might be able to circumvent diagnosis through compensation at the behavioural level but still have difficulties at the neural and cognitive levels (Livingston & Happé, 2017). Therefore, it is necessary to develop more sensitive assessments that target underlying cognitive ability but are not susceptible to compensation (Frith, 2012). Southgate et al. (2007)'s anticipatory-looking paradigm measuring implicit mentalizing holds promise as one such assessment (see the task details in Chapter 1, Section 1.1.3). Evidence supports that this paradigm can reliably measure mentalizing in non-autistic adults (see review in Schneider et al., 2017) and can detect mentalizing difficulties in autistic adults (e.g., Schneider et al., 2013; Senju et al., 2009) and children (e.g. Schuwerk et al., 2016; Senju et al., 2010).

However, three main criticisms have been made about Southgate et al. (2007)'s paradigm (see detailed discussion in Chapter 1, Section 1.1.4). Accordingly, poor performance on the commonly reported outcome measures of this task alone cannot be used to conclusively deduce that autistic individuals have difficulties in spontaneous mentalizing. The current chapter is going to focus on the first two criticisms. First, this paradigm possesses a single-trial design, which can attenuate its reliability via escalating error variance and dropout rate (Dang et al., 2020; Kulke, Wübker, et al., 2019). A multi-trial design would improve the signal-to-noise ratio and increase power, allowing for a better estimation of individual performance. Therefore, the current study first set out to increase the number of trials to improve task reliability.

Second, Heyes (2014) proposed the submentalizing hypothesis, which suggests that non-autistic individuals pass the anticipatory-looking task due to domain-general cognitive mechanisms, rather than mentalizing. She claimed that non-autistic people predict the agent's action based on their own, but not the agent's, false belief of the object's location; while autistic people are less distracted but also less likely to predict people's actions. Thus, the current study also aimed to address this by including true-belief conditions as baseline control conditions that closely match false-belief conditions and by providing a detailed analysis of eye movements throughout the paradigm. Based on Heyes (2014), we should see differences between autistic and non-autistic people in visual attention to the key events. Specifically, non-autistic people should be more distracted by the agent and therefore look at the puppet less than autistic people. Additionally, autistic people should have less tendency to predict the agent's action in both false-belief and true-belief control conditions than non-autistic people.

Several studies have endeavoured to improve the reliability of Southgate et al. (2007)'s paradigm. However, the results from studies implementing true-belief conditions are mixed. For example, non-autistic infants and adults were observed to be able to reason both true beliefs and false beliefs with low cognitive demands (Surian & Geraci, 2012; Wang & Leslie, 2016). Nonetheless, Kulke, Reiß, et al. (2018) used the same paradigm but did not find a relationship between true beliefs and false beliefs attribution in any age groups. Using the same task as Senju et al. (2009), Gliga et al. (2014) considered one of the familiarization trials as a true-belief condition and concluded that siblings of autistic children were able to attribute the agent's true belief, but not false belief. However, this familiarization trial was shorter and simpler than the false-belief condition and did not involve a head turn. Also, as the agent's true belief was consistent with reality and the child's own belief, plus the agent's reaching action was presented in every familiarization trial, it is possible that they predicted the agent's action according to their own belief (Russell et al., 1991; Van der Meer et al., 2011; Wang & Leslie, 2016) or learned the behavioural contingency from the repeated action (Schuwerk et al., 2015; Sodian et al., 2015). Therefore, it is needed to verify whether autistic individuals struggle specifically with mentalizing with a well-matched true-belief condition.

To date, the empirical results from studies adopting a multi-trial design are mixed. Schneider, Bayliss, et al. (2012) were the first to investigate how spontaneous mentalizing operates over time in non-autistic adults, and claimed that spontaneous mentalizing can be sustained over the course of a multi-trial procedure. Using the same paradigm in autistic adults, Schneider et al. (2013) replicated the observations of Senju et al. (2009): autistic individuals did not spontaneously mentalize across the trials, despite performing well in explicit mentalizing tasks, indicating that multi-trial designs are viable and are not susceptible to compensatory learning in autistic individuals.

However, it is important to note that Schneider's studies did not analyse whether participants showed a clear looking bias towards the belief-congruent box. Instead, they compared the looking bias towards the belief-congruent box on false-belief trials and the belief-incongruent box on true-belief trials. Thus, it is unclear whether the belief manipulation was successful within each condition; indeed, the autistic participants surprisingly appeared similarly likely to look at either box in the true-belief condition (Schneider et al., 2013, pp. 414-415, Figures 2 & 3) and the non-autistic participants appeared more likely to look at the belief incongruent location in the false-belief condition (Schneider, Bayliss, et al., 2012, pp. 435-436, Figures 2 & 3). Additionally, Schneider's studies consisted of 20 test trials, each more than one minute in duration, plus at least 20 familiarization trials, amplifying the total duration of the task. The present study, therefore, chose to make the paradigm more streamlined by removing any unnecessary actions and potential social confounds (i.e. an extra object displacement in the original false-belief trials and the agent's wave and smile), and keeping familiarization trials to a minimum.

Another factor worth highlighting in Schneider et al. (2013) is that, during the anticipatory period, both groups allocated their first fixation to the agent's face in more than 70% of trials. This meant that more than 70% of their data was excluded from the analysis. One possible reason could be the absence of an occluder between the agent and the scene, a disparity with Southgate et al. (2007)'s paradigm. It is possible that by removing the occluder, participants looked to the agent in anticipation of her action rather than making anticipatory saccades to the belief-congruent area as the first place where the action was expected. A second reason may be that the agent left the room, rather than turning to the back as in Southgate et al. (2007). This meant that the participant could not be sure of the agent's knowledge of the object's location whilst off-scene. A further contention is whether the reappearance of the agent is a salient event, which could result in retroactive memory interference on object displacement during the agent's absence (Heyes, 2014). Moreover, it is worth noting that in Schneider et al. (2013), the object was displaced twice in the true-belief scenario but only once in the false-belief scenario before the agent came back. Therefore, the higher memory load required in the true-belief condition may have caused the lack of looking difference between the belief-congruent area in true-belief trials and the belief-incongruent area in false-belief trials in both autism and non-autism groups. In order to avoid these potential caveats and by doing so increase the number of trials included in the analysis, we chose to retain the occluder, to keep the agent visible at all times, and to displace the object only once.

Furthermore, both of Schneider's studies used a false-belief condition with high-inhibitory demands, as the object was displaced to the other box, rather than removed from the scene (i.e. low-demand) as in Southgate et al. (2007). Wang and Leslie (2016) directly contrasted high- and low-demand false-belief conditions. They found that both non-autistic 3-year-olds and adults showed clear anticipatory-looking behaviours towards the belief-congruent area in the low-demand condition, but no looking bias in the high-demand condition. As a result, they suggested that the high-demand scenario requires greater cognitive resources to inhibit one's own belief about the object's location. This same suggestion of a reality bias (or true-/own-belief bias), has also been attributed to autistic individuals as an explanation for their poor false-belief task performance compared to non-autistic individuals (Russell et al., 1991; Van der Meer et al., 2011). If this inhibition difficulty is indeed even stronger in autistic individuals, they may therefore show a bias to look towards the object's current location in the high-demand false-belief condition and true-belief conditions, rather than the lack of bias shown by non-autistic individuals. Accordingly, we chose to study the effectiveness of our belief manipulation and compare high- and low-demand false-belief conditions in both autistic and non-autistic individuals.

Schuwerk et al. (2015) also reported a multi-trial study across just two trials, suggesting that experience might improve autistic individuals' performance when the outcome action (i.e. the agent opening the belief-congruent window and retrieving the object) is shown; only the second trial tested this learning effect. Autistic and non-autistic adults differed in looking bias in the first false-belief trial, consistent with Senju et al. (2009), but not in the second. However, the looking bias of autistic adults did not differ from chance in either trial and no improvement was seen in autistic children (Schuwerk et al., 2016). Moreover, any improvement in performance by the autism group could be due to compensatory learning of a behavioural contingency during the first trial, without representing the agent's mental state (Sodian et al., 2015). Hence, we chose not to show the outcome action in experimental trials. If the improved performance observed by Schuwerk et al. (2015) was due to an increase in mentalizing, then the performance of autistic individuals in our task should also increase over time to the level of non-autistic individuals. On the other hand, in line with Schneider et al. (2013), we expect to see no change in performance over time.

To summarise, the existing adaptations to the original Southgate et al. (2007) paradigm have presented exciting, yet inconclusive, evidence surrounding spontaneous mentalizing in autism. The present study, therefore, sought to advance the paradigm by implementing a multi-trial experiment with shorter trials, matched true-belief conditions, and both high- and low-inhibitory demand false-belief conditions to scrutinize the claims that have been made in the literature. In the low-demand false-belief condition, according to the literature, we predicted that autistic individuals should show no looking bias, whilst non-autistic individuals should be able to anticipate the agent's false-belief-based action. In the high-demand false-belief condition, both groups would show no looking bias. In the true-belief conditions, non-autistic adults would look significantly longer at the belief-congruent than the belief-incongruent area, whilst the prediction for autistic adults was bidirectional based on different theories. Following Gliga et al. (2014), we also expected both groups to show belief-congruent performance in the familiarization trials. Additionally, attentional bias differences would be shown between groups during the object displacement.

2.2 Methods

2.2.1 Participants

Participants were recruited through a local participant database, local autism support groups, and advertisements placed around the local community. This study was approved by the UCL Research Ethics Committee, and all methods were performed in accordance with the approved guidelines and regulations. Written informed consent was obtained from all participants. Assuming a medium effect size (*F* = 0.25) as seen in Senju et al. (2009) and Schneider et al. (2013)'s studies and power of .80, a sample size calculation indicates that we needed 17 participants per group to detect the critical interaction between group and belief. To ensure sufficient statistical power, a total of 67 participants were recruited, 40 before and 27 after the COVID-19 pandemic. Five participants (three autistic and two non-autistic) were excluded from the analysis due to poor data quality (see *Data pre-processing* below), leaving 32 autistic adults, aged 18-64 years; and 30 non-autistic adults, aged 18-50 years. The resulting two groups were comparable for age, sex, Verbal IQ, Performance IQ, and Full-scale IQ, as measured by the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II; Wechsler, 2011), and all participants had a Full-scale IQ greater than 80 (see Tables 2.1 & 2.2).

None of the non-autistic participants reported a diagnosis, or family history, of psychiatric or neurodevelopmental disorders. Each participant in the autism group had previously received a diagnosis of autism spectrum disorder, Asperger syndrome, high-functioning autism or atypical autism from a qualified clinician. The Autism Diagnostic Observation Schedule Second Edition (ADOS-2; Lord et al., 2012) was used to verify participants' autism diagnosis. ADOS scores were available for nine autistic participants (see Table 2.1); and seven of those met the criteria for autism or autism spectrum classification. The two participants who scored below the threshold, and the eight who had no ADOS score, were retained within the sample, because first scores below the cut-off are not uncommon in highly intelligent autistic adults (de Bildt et al., 2016), and second the autism group reported significantly higher autistic traits than the non-autism group (see Tables 2.1 & 2.2).

Table 2.1. ***Descriptive statistics of each group, Mean (Standard Deviation).***

	Autism (*n* = 32, 15 females)	Non-autism (*n* = 30, 14 females)
Age	32.00 (13.84)	30.70 (10.41)
Verbal IQ (WASI-II[a])	116.29 (17.17)	115.77 (18.06)
Performance IQ (WASI-II[a])	117.65 (21.09)	117.93 (18.83)
Full-scale IQ (WASI-II[a])	121.37 (21.14)	120.33 (18.12)
ADOS-2[b]	8.44 (5.08)	
Autistic traits (AQ[c])	33.55 (7.65)	17.47 (7.36)

Note. [a]WASI-II = Wechsler Abbreviated Scale of Intelligence, Second Edition; [b]ADOS = Autism Diagnostic Observation Schedule Second Edition (data was unavailable for eight autistic participants); [c]AQ = Autism-Spectrum Quotient.

Table 2.2. ***Group-wise comparison between the autism and non-autism groups.***

	Autism vs. Non-autism
Age	*t*(60) = 0.42, *p* = .679, *d* = 0.106
Sex	$\chi^2(1) = 0.99$, *p* = .594, odds ratio = 1.008
Verbal IQ (WASI-II[a])	*t*(59) = 0.12, *p* = .908, *d* = 0.030
Performance IQ (WASI-II[a])	*t*(59) = -0.06, *p* = .955, *d* = -0.014
Full-scale IQ (WASI-II[a])	*t*(59) = 0.21, *p* = .838, *d* = 0.053
Autistic traits (AQ[b])	***t*(59) = 8.37, *p* < .001, *d* = 2.143**

Note. [a]WASI-II = Wechsler Abbreviated Scale of Intelligence, Second Edition; [b]AQ = Autism-Spectrum Quotient.

2.2.2 Procedure

Participants started the session by completing the WASI-II, then Section A of the implicit mentalizing task, followed by the explicit mentalizing task and questionnaires measuring demographics and autistic traits, and finished with Section B of the implicit mentalizing task. Participants were then fully debriefed. The overall duration of the experiment was 1.5 hours. Testing was conducted at the Institute of Cognitive Neuroscience, University College London.

2.2.3 Mentalizing Tasks

2.2.3.1 Implicit Mentalizing Task

The implicit mentalizing task used a multi-trial anticipatory-looking paradigm with matched true-belief and false-belief conditions, which was adapted from the anticipatory-looking paradigm in Southgate et al. (2007). In an attempt to maintain participants' attention, the task was split into two sections (A and B) with a 20-minute interval in between (see *Figure 2.1*). Participants were instructed to passively view some videos and informed they would be asked questions about their content at the end, to encourage them to pay attention and watch carefully. The questions asked about basic features of the videos (e.g. the colour of the puppet) and participants' judgements (e.g. the most frequent final location of the object), but participants were not informed of the style of question in advance to avoid directing their attention to particular features of the videos.

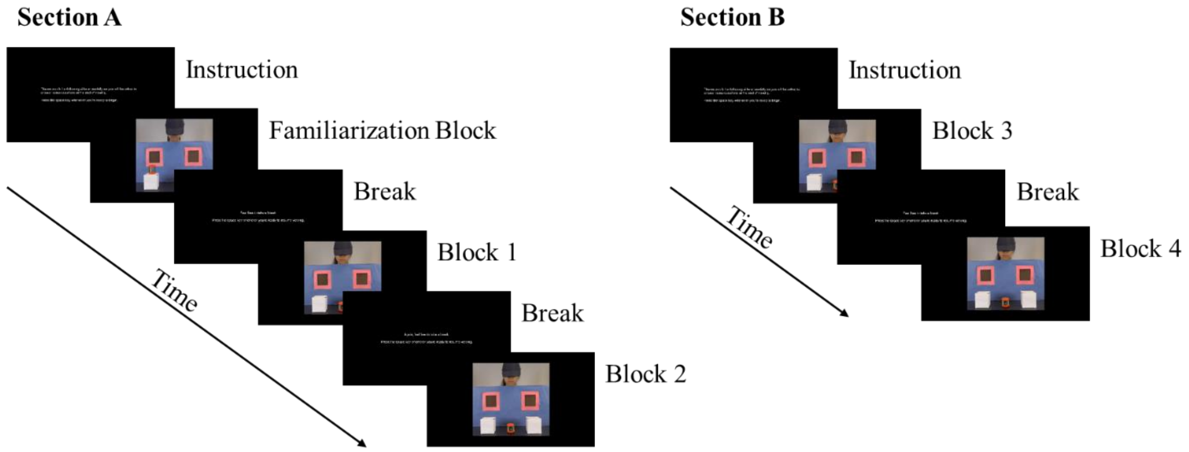


***Figure 2.1.* Implicit mentalizing task procedure.**

Section A contained one familiarization block, and both Sections A and B contained two experimental blocks (see *Figure 2.1*). The familiarization block included four short and four long familiarization trials, which enabled participants to learn the contingency that the agent would retrieve the object after an alert signal (the windows illuminated and a chime sounded simultaneously for 800ms). The short familiarization trials started with the object on top of one of two boxes (see *Figure 2.2b*). The scene was frozen for 2,800ms from the onset of the alert signal. The agent then reached through the corresponding window and retrieved the object. During the long familiarization trial, a puppet hid the object in one of the boxes while the agent was watching (see *Figure 2.2d*). After the puppet left the scene, the alert signal occurred and the scene froze, and then the agent reached through the window, opened the box and retrieved the object. To make the contingency between the alert signal and the agent reaching through the window more salient, we also filmed two short and two long familiarization trials using transparent boxes to give participants a more direct perception of the object location (see *Figures 2.2a* & *2.2c*). The end location of the object was counterbalanced in the short and long, transparent and opaque trials, producing eight possible videos, which were all displayed in the familiarization block in random order.

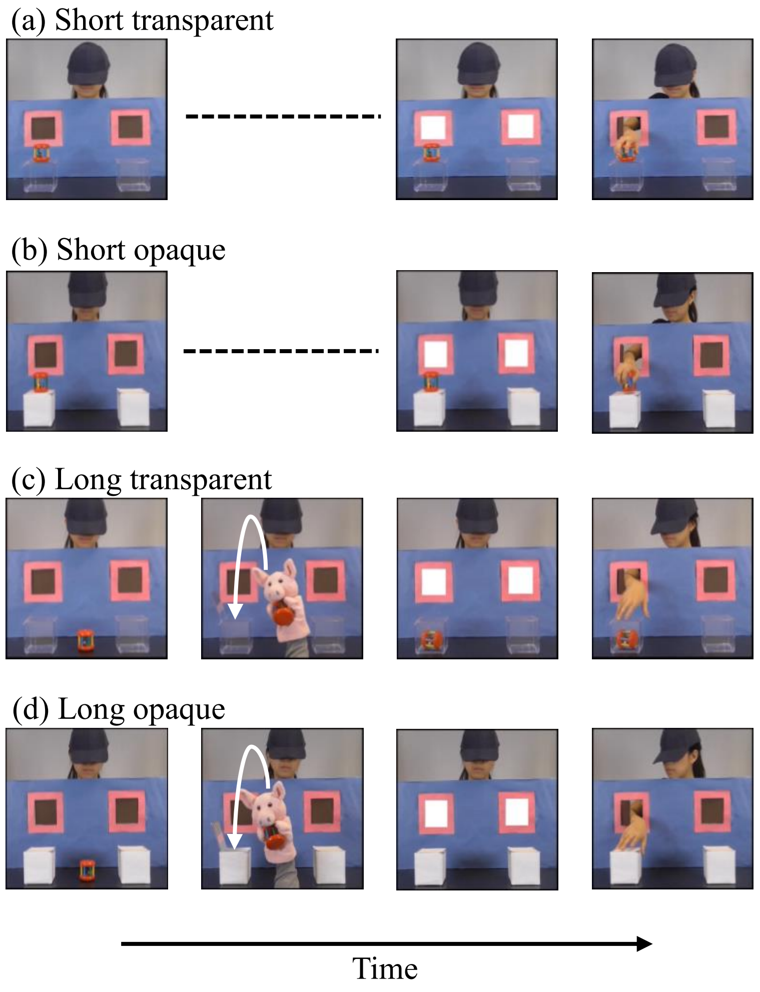


Figure 2.2. **Short (5,000ms) and long (15,000ms), opaque and transparent familiarization trials. The scenarios of (a) and (c) were identical to (b) and (d), respectively. Long familiarization scenarios include an additional object transfer event.**

Each experimental block started with one short and one long familiarization trial, randomly selected from the eight videos without replacement, to remind participants of the contingency. This was followed by four true-belief and four false-belief trials, consisting of two true- and two false-belief conditions: true-belief short-turn, true-belief long-turn, false-belief high-demand and false-belief low-demand. True-belief short-turn and true-belief long-turn conditions were matched to false-belief high-demand and false-belief low-demand conditions, respectively. In the true-belief conditions, the agent's belief about the object's location was congruent with its actual location, while these two locations were incongruent in the false-belief conditions so the agent held a false belief about the object's location. These conditions only used the opaque boxes and the agent did not retrieve the object; instead the scene remained frozen for the full 4,800ms from the onset of the alert signal (see *Figure 2.3*).

In the true-belief short-turn condition (see *Figure 2.3a*), the puppet hid the object in one of the boxes. A doorbell then rang and the agent turned away from the scene, followed almost immediately by the sound of a door closing, whereupon the agent turned back to the scene and witnessed the puppet move the object to the other box. Once the puppet had disappeared, the alert signal occurred and the scene froze. In the true-belief long-turn condition (see *Figure 2.3b*), the only difference was that the puppet returned the object to the original box whilst the agent was turned away from the scene and the agent did not turn back until the puppet had disappeared, at which point the alert signal occurred. The false-belief high-demand condition (see *Figure 2.3c*) only differed from the true-belief short-turn condition in that the agent remained turned to the back whilst the puppet moved the object to the other box. The false-belief low-demand condition (see *Figure 2.3d*) was similar to the true-belief long-turn condition except the puppet removed the object from the scene whilst the agent was turned away. Each sound was paired with the same corresponding event in all of the experimental videos, and the agent's head movements always followed the puppet's movement when she was facing the front to indicate that she was paying attention to the situation.

The box that first contained the object and the direction in which the agent turned were both counterbalanced, producing four possible videos for each condition. In each experimental block, two videos were randomly selected from each condition, giving a total of eight videos presented in random order. Participants watched each experimental video once in each section. Mathematica (Wolfram Research, Inc. Version 11.1) was used to code the random presentation sequences of the videos, which were then imported into the presentation software.

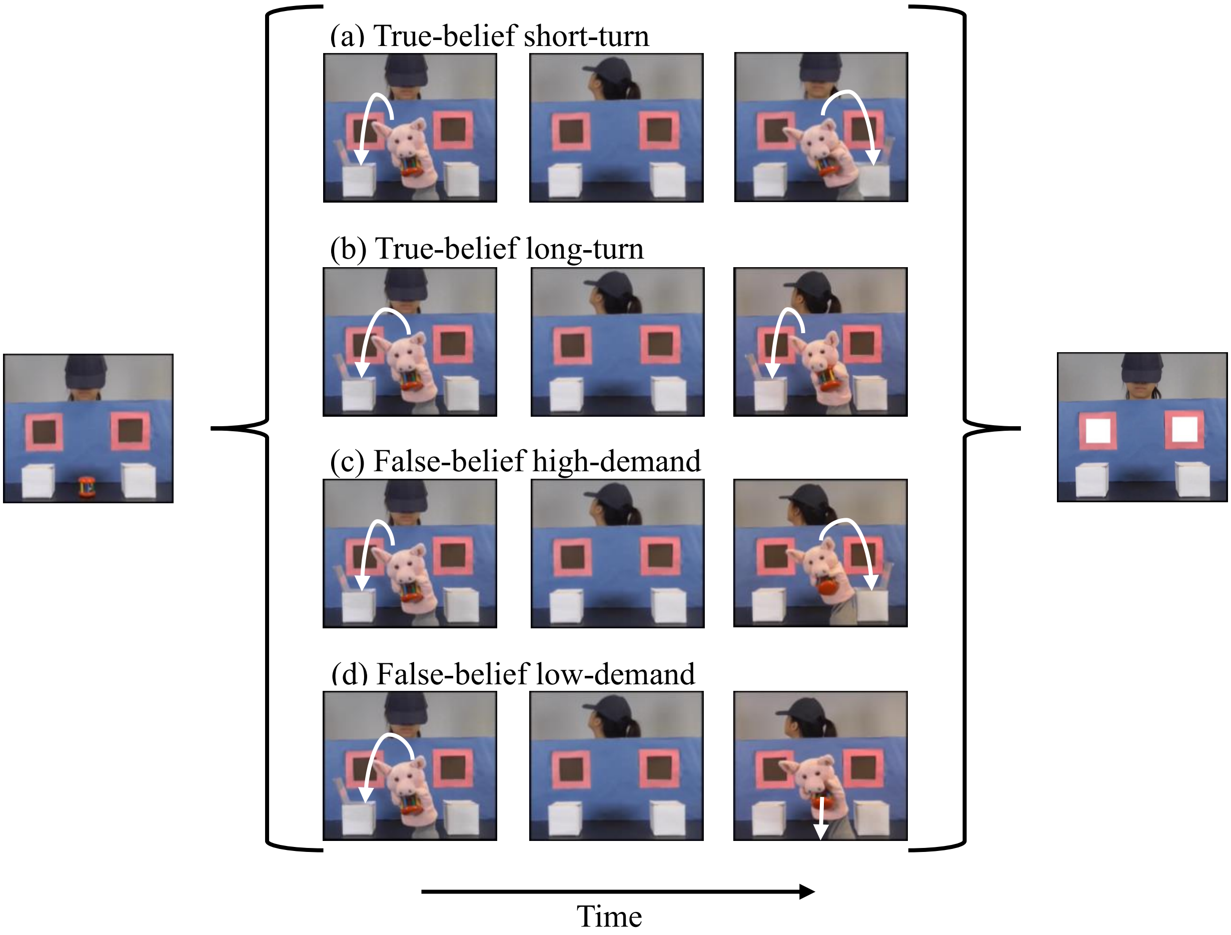

(a) True-belief short-turn
(b) True-belief long-turn
(c) False-belief high-demand
(d) False-belief low-demand
Time

Figure 2.3. **Sequence of events in the true-belief and false-belief condition videos (a) true-belief short-turn (37,000ms), (b) true-belief long-turn (33,000ms), (c) false-belief high-demand (37,000ms), (d) false-belief low-demand (33,000ms).**

Apparatus. A remote screen-based Tobii Pro X3-120 eye-tracker system, with a sampling rate at 120Hz, was used to record gaze data (Tobii, Sweden). Visual and auditory stimuli were presented via a Dell Precision 5520 laptop (15.6-inch) with Tobii Pro Studio 3.4.8 software, integrated with the eye-tracker. Participants sat approximately 70cm from the eye-tracker and were instructed to sit still throughout the eye-tracking assessment. A 9-point calibration was performed before each section began.

Areas of interest (AOIs). Nine AOIs within five timeframes were identified across each trial (see Table 2.3 & *Figure 2.4*). The total fixation duration was encoded and extracted through Tobii Pro Studio, measuring the sum of the duration of all fixations within each AOI. According to the scenarios, timeframes *1* and *3* captured object displacement, timeframes *2* and *4* captured the agent's head-turn, and timeframe *5* (*af*) captured action anticipation after the onset of the alert signal. Therefore, to investigate group differences in attention distribution, the total fixation durations of *Head_1* and *Head_3* were combined as *Head_bf, Puppet_1* and *Puppet_ 3* were combined as *Puppet_bf*, *HeadTurn_2* and *HeadTurn_4* were combined as *HeadTurn*, and that of *Belief-congruent* and *Belief-incongruent* were combined as *Anticipation_af*. For the long familiarization trials using opaque boxes, the total fixation durations of *Belief-congruent* and *Belief-incongruent* for two different timeframes were extracted (see Table 2.3). For timeframe *5*, 4,800ms AOIs were used to evaluate if participants were able to predict the agent's action through mentalizing her beliefs, while 2,500ms AOIs were used in familiarization trials to examine if they paid attention to learn the contingency of the task.

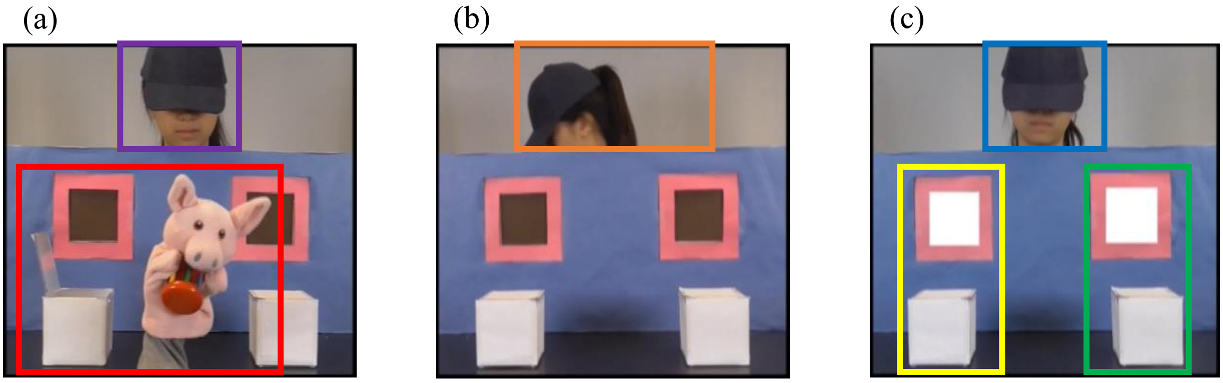


***Figure 2.4.* Examples of the areas of interest: (a) *Head_ 1* and *Head_ 3* (purple), *Puppet_ 1* and *Puppet_ 3* (red); (b) *HeadTurn_2* and *HeadTurn_4* (orange); (c) *Head_af* (blue), *Belief-congruent* (yellow) and *Belief-incongruent* (green).**

Table 2.3. ***Definition of each AOI.***

AOI_timeframe	Location	Event
Head_1	Agent's head area	The agent watches the puppet hiding the object in one of the boxes
Puppet_ 1	Puppet's moving area	
HeadTurn_2	Agent's head area	The agent turns away from the scene
Head_ 3	Agent's head area	The puppet displaces the object
Puppet_ 3	Puppet's moving area	
HeadTurn_4	Agent's head area	The agent turns back to the scene
Head_af	Agent's head area	From the onset of the alert signal to the end of the trial, total duration 4,800ms; for the long familiarization trials using opaque boxes, data were also encoded from the onset of the alert signal up until the agent reaches through the window, total duration 2,500ms
Belief-congruent	Window & box area consistent with agent's belief	
Belief-incongruent	Window & box area inconsistent with agent's belief	

Data pre-processing. Differential looking scores (DLS), which measure participants' looking preference between two visual targets, were calculated by dividing the difference between the total looking time to the *Belief-congruent* and *Belief-incongruent* AOIs, by the sum of the two. DLS ranged from 1 to -1, closer to 1 if participants showed a looking bias towards the *Belief-congruent* AOI, closer to -1 if they were biased towards the *Belief-incongruent* AOI, and closer to 0 if they looked equally to both AOIs, equivalent to chance performance.

Three exclusion criteria were applied to ensure participants were paying attention to the key events in the videos (e.g. watching the hand retrieve the object). First, the data from the entire task were excluded for any participant whose average DLS in the familiarization block (based on the full 4,800ms post-flash) was missing or below chance, to confirm that they had paid attention to the key event (a combination to the prediction and the action itself). Second, the data from each experimental block were excluded if the average DLS of the two familiarization trials at the beginning of that block was missing or below chance. Third, participants were excluded if they missed more than 25% of data. After data cleaning, five participants (three autistic and two non-autistic) were excluded, and two additional participants each had their data removed from one experimental block.

2.2.3.2 Explicit Mentalizing Task

The Mental State set from the Strange Stories Task (Happé, 1994; White et al., 2009), an advanced mentalizing test, was used to assess participants' ability to infer mental states in social situations explicitly. In addition to accuracy (maximum score of 16), comprehension time was recorded (i.e. time elapsed from the start of reading a story to the start of answering the question).

2.2.4 Autistic Traits

Autistic traits were measured by the Autism-Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, et al., 2001), with higher scores indicating more autistic traits, ranging between 0-50.

2.3 Results

2.3.1 Implicit Mentalizing Task

2.3.1.1 Differential Looking Sores (DLS)

One-sample *t*-tests, comparing the average DLS of the long familiarization trials with opaque boxes (based on the first 2,500ms post-flash, before the agent reached through the window) to chance performance, were conducted in each group separately to assess action prediction. Both groups performed significantly above chance: autism: *M* = 0.27, *SD* = 0.36, *t*(31) = 4.17, *p* < .001, *d* = 0.737; non-autism: *M* = 0.26, *SD* = 0.44, *t*(29) = 3.23, *p* = .003, *d* = 0.590, but no difference was found between the groups, *t*(60) = 0.07, *p* = .943, *d* = 0.018, indicating that both groups were able to correctly predict the agent's actions.

The same tests were conducted for each experimental condition in the non-autism group to check the task validity. The DLSs for the true-belief short-turn (*M* = 0.22, *SD* = 0.42), *t*(29) = 2.79, *p* = .009, *d* = 0.510, true-belief long-turn (*M* = 0.32, *SD* = 0.36), *t*(29) = 4.91, *p* < .001, *d* = 0.897, and false-belief low-demand conditions (*M* = 0.19, *SD* = 0.34), *t*(29) = 3.00, *p* = .005, *d* = 0.548, were significantly above chance, but that of the false-belief high-demand (*M* = -0.10, *SD* = 0.30) did not significantly differ from zero, *t*(29) = -1.90, *p* = .068, *d* = -0.346. That is, non-autistic participants showed a preference for the *Belief-*

congruent location in the true-belief short-turn, true-belief long-turn and false-belief low-demand conditions, but did not show any looking bias in the false-belief high-demand condition. This indicated that all the conditions, but not the false-belief high-demand, were valid for detecting implicit mentalizing ability.

The results in the autism group revealed that the DLS was significantly above zero in true-belief but not false-belief conditions: true-belief short-turn (M = 0.18, SD = 0.39), t(31) = 2.62, p = .014, d = 0.462, true-belief long-turn (M = 0.14, SD = 0.33), t(31) = 2.49, p = .018, d = 0.440, false-belief low-demand (M = 0.07, SD = 0.37), t(31) = 1.05, p = .304, d = 0.185, false-belief high-demand (M = -0.06, SD = 0.28), t(31) = -1.14, p = .264, d = -0.201. Since the false-belief high-demand condition was not valid, we decided to focus the following analysis on the false-belief low-demand condition and its matched true-belief long-turn condition.

A 2x2x2 mixed-design analysis of variance was conducted using the DLS as the outcome variable, with Time (Section A vs. Section B) and Belief (true-belief vs. false-belief) as within-subject factors, and Group (non-autism vs. autism) as a between-subject factor. There was a marginal main effect of Group, F(1, 60) = 3.88, p = .054, *partial* η^2 = .061, with the non-autism group displaying higher DLSs than the autism group (see *Figures 2.5* & *2.6*). The results also revealed a main effect of Belief, F(1, 60) = 7.67, p = .007, *partial* η^2 = .113, the means indicate that higher DLSs occurred in the true-belief than false-belief condition (see *Figures 2.5* & *2.6*). There was also a Time*Belief interaction, F(1, 60) = 4.82, p = .032, *partial* η^2 = .074; see *Figure 2.5*). Post-hoc tests (with α-level adjusted to p = .0125 for multicomparison correction) indicated that the true-belief condition was performed significantly better than the false-belief condition in Section A, t(61) = 3.56, p < .001, d = 0.453, but not Section B, t(61) = 0.53, p = .597, d = 0.068; and neither condition was different between Section A and B: true-belief: *t*(61) = 1.43, *p* = .157, *d* = 0.182; false-belief: *t*(61) = -1.46, *p* = .149, *d* = -0.186.

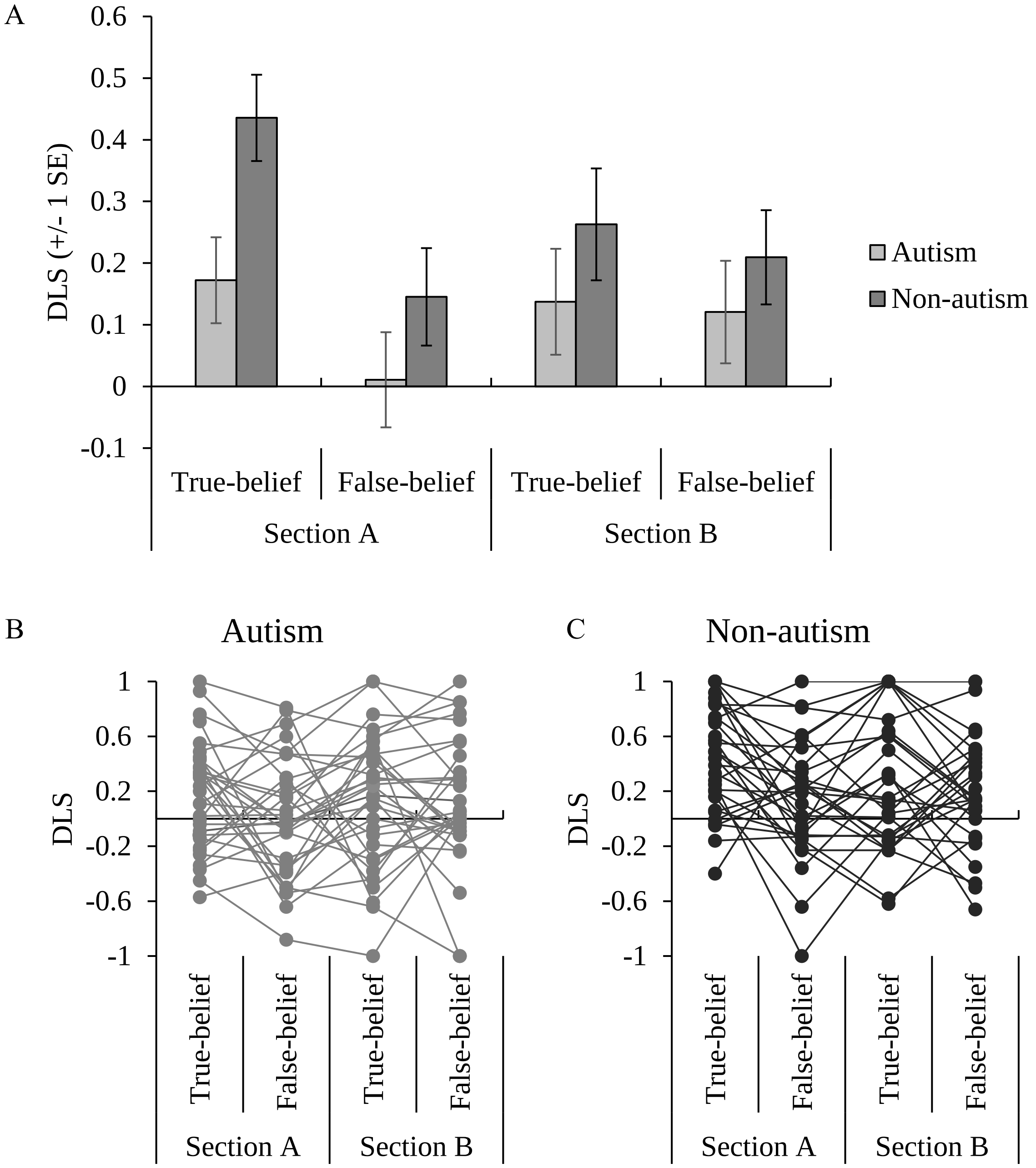


Figure 2.5. **A: Mean true-belief and false-belief DLS of the implicit mentalizing task in Sections A and B for both the autism and non-autism groups; B & C: Each line represents a participant.**

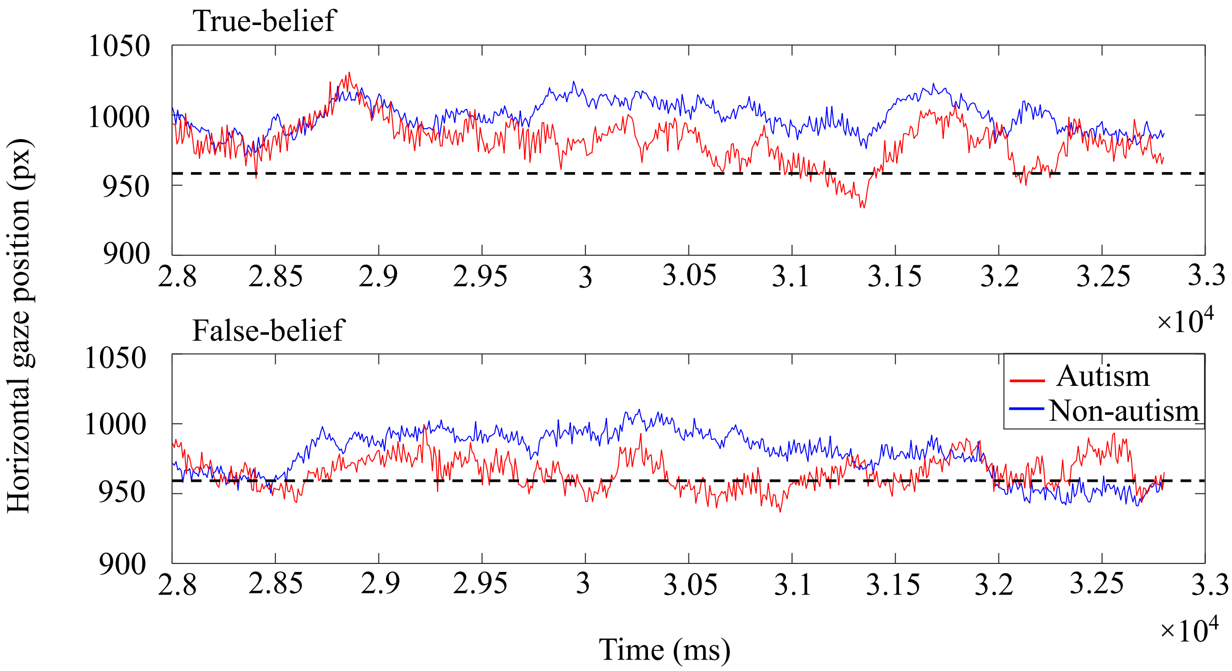


Figure 2.6. **Grand averaged horizontal gaze position in response to stimulus presentations across true-belief and false-belief conditions. This timeframe begins at the onset of the alert signal until the end of the video, from 2.8 to 3.28x10^{4}ms. The origin (0 px) of the frame is defined as the edge of the screen on the belief incongruent side. The dashed line indicates the midline of the screen. Values above the dashed line are therefore biased towards the *Belief-congruent* side of the screen.**

2.3.1.2 Fixation Pattern

We explored group differences in fixation patterns on the total fixation durations at critical time points. Given the critical frames were identical in all true-belief and false-belief trials and we found no interaction between group and belief, the data were collapsed across these two conditions. No group difference was found in *Head_bf*, *t*(60) = -0.54, *p* = .590, *d* = -0.138, *Puppet_bf*, *t*(60) = 0.91, *p* = .368, *d* = 0.231, or *HeadTurn*, *t*(60) = 1.19, *p* = .237, *d* = 0.304 (see *Figure 2.7*). More specifically, participants looked significantly longer at

Puppet_3 (*M* = 7.37, *SD* = 1.59) than *Head_3* (*M* = 0.88, *SD* = 0.74) whilst the agent was turned away, *t*(61) = 27.53, *p* < .001, *d* = 3.50, and this result held within each group (autism: *p* < .001; non-autism: *p* < .001). Indeed, as shown in *Figure 2.8*, the gaze patterns of both groups were almost identical in timeframes *1*, *3* and *4*. During timeframe 2, non-autistic adults seemed to mostly look at the agent (i.e. above the dashed line, indicating the top of the occluder), while autistic adults gazed at both the agent and the scene. Autistic participants seem on average to spend most of the timeframe *5* gazing within the belief congruent and incongruent areas (i.e. below the dashed line), whereas non-autistic participants gazed at both the agent and the anticipatory areas. However, there was no group difference in *Head_af*, *t*(60) = 0.45, *p* = .655, *d* = 0.114, or *Anticipation_af*, *t*(60) = 1.49, *p* = .141, *d* = 0.379. Altogether, these analyses indicated that the looking pattern of the two groups did not significantly differ during the key events.

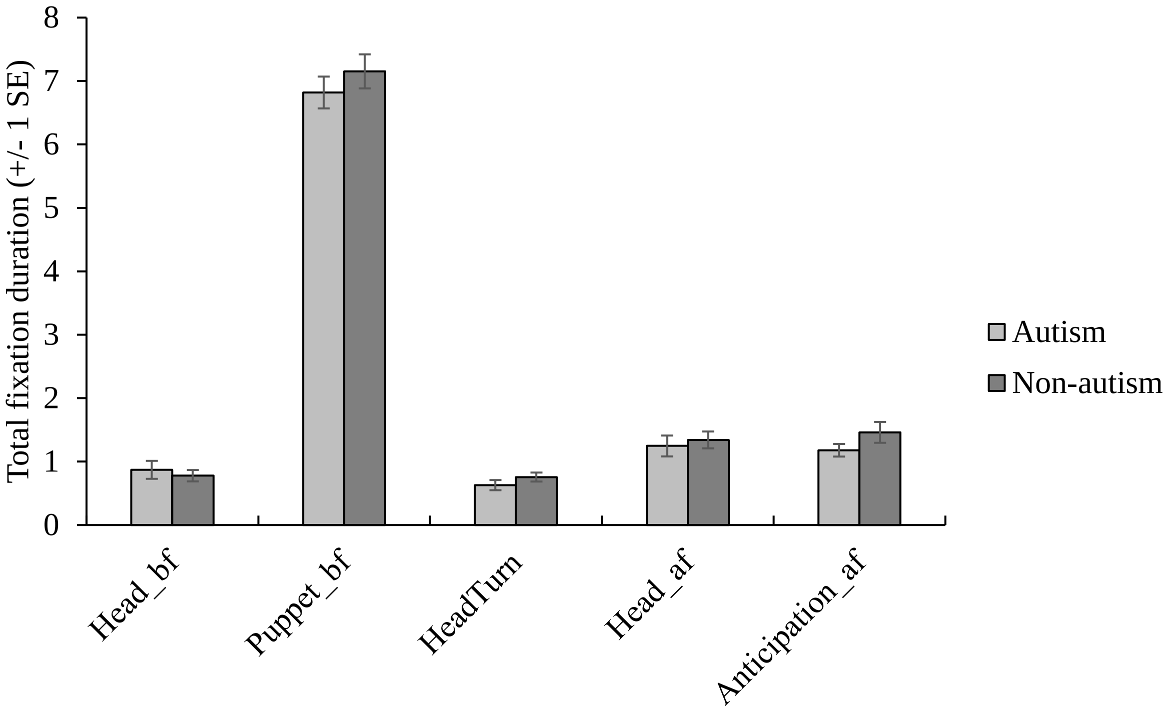


Figure 2.7. **Mean total fixation duration indicated no group difference during all key events.**

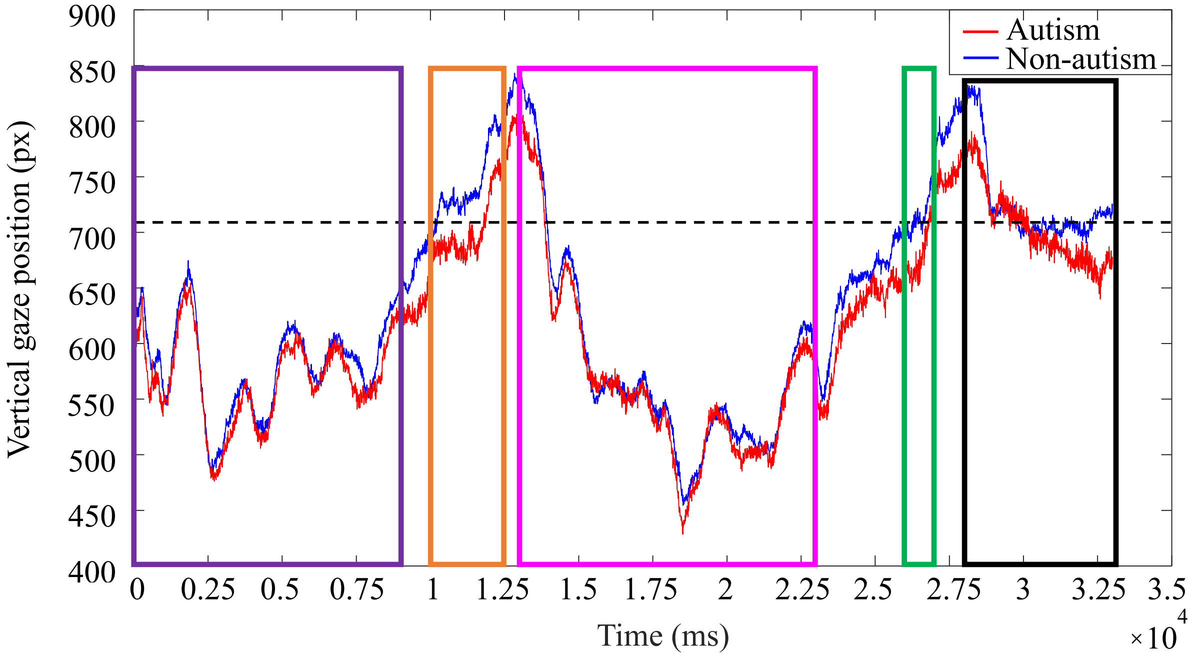


Figure 2.8. **Grand averaged vertical gaze position in response to stimulus presentations across true-belief and false-belief conditions. Timeframe *1*: *Head_ 1* and *Puppet_ 1* (purple), Timeframe *2*: *HeadTurn_2* (orange), Timeframe *3*: *Head_ 3* and *Puppet_ 3* (pink), Timeframe *4*: *HeadTurn_4* (green), Timeframe *5*: *Head_af* and *Anticipation_af* (black). Video onset occurred at 0ms. The origin (0,0) of the frame is in the bottom left corner of the screen. The dashed line indicates the edge of the purple board in the scene.**

2.3.2 Explicit Mentalizing Task

Independent samples *t*-tests revealed that performance on the Strange Stories Task was comparable between the autism and non-autism groups, both in terms of accuracy (autism: *M* = 13.31, *SD* = 1.93; non-autism: *M* = 13.63, *SD* = 2.28), *t*(60) = -0.60, *p* = .551, *d* = -0.152, and comprehension time (autism: *M* = 28.93, *SD* = 10.55; non-autism: *M* = 27.00, *SD* = 6.74), *t*(57) = 0.82, *p* = .417, *d* = 0.214.

2.4 Discussion

The present study aimed to probe spontaneous mentalizing in autism. To overcome the methodological difficulties seen in previous work, a multi-trial paradigm with well-matched true-belief control conditions was used. We conducted a detailed analysis of gaze patterns throughout individual trials, as well as changes in performance over the test session. Our results support the presence of spontaneous mentalizing difficulties in autistic adults, despite a typical allocation of attentional resources to complex social stimuli.

2.4.1 Belief Reasoning in Autism

In line with Senju et al. (2009), we found a dissociation between implicit and explicit mentalizing tasks in the autism group. Specifically, in the explicit mentalizing task, which is considered an advanced test of mentalizing (Happé, 1994), the performance of autistic adults was indistinguishable from non-autistic adults, indicating sophisticated mentalistic reasoning. On the other hand, in the false-belief condition of our implicit mentalizing task, while the non-autism group showed a bias to look at the belief congruent target location, the autism group split their time equally between the belief-congruent and -incongruent target locations, indicating that they failed to appreciate the agent's false belief. Of note, these difficulties could not be accounted for by difficulties predicting actions, submentalizing processes or attentional differences (discussed in more detail below). These findings are consistent with the idea that some autistic individuals with average-to-high IQs may acquire the capacity to explicitly 'mentalize' about complex mental states (Frith, 2004), but still struggle to implicitly attribute simple mental states (Senju et al., 2009).

In true-belief conditions, both autistic and non-autistic participants displayed a clear looking bias towards the belief-congruent area, which is more consistent with Gliga et al.

(2014) than Kulke, Reiß, et al. (2018). This deserves detailed consideration as it has distinct implications for the specificity of the mentalizing differences thought to lie at the very core of autism. This finding does not corroborate the prediction according to Heyes (2014) that autistic individuals may struggle to spontaneously predict others' actions per se, regardless of mentalizing requirements (also see in Van de Cruys et al., 2014), leading to a lack of preference for the belief congruent location. Consistently, we also found that both groups of participants showed anticipatory eye movements predicting the agent's hand reach to retrieve the object in the long familiarisation trials (using opaque boxes, based on the first 2,500ms after the onset of the alert signal before the agent reached through a window), implying that they were capable of action prediction in its most basic form. These two evidences can immediately dismiss the possibility according to the submentalizing hypothesis.

In both false-belief and true-belief conditions, all of our participants spent the majority of their time looking at the puppet rather than the agent's head whilst the object was displaced, consistent with Gliga et al. (2014)'s analysis. Thus, it is unlikely that the non-autistic adults were distracted by the first head-turn and therefore failed to notice the critical events (i.e. the object displacement) thereafter, as predicted by the submentalizing hypothesis (Heyes, 2014). Similarly, during the action anticipatory period, the two groups spent a similar amount of time looking at the agent and the object (i.e. windows and boxes), revealing that autistic individuals have neither paid more attention to non-social stimuli nor less attention to social stimuli than non-autistic individuals, contrary to social attentional theories of autism (e.g. Klin et al., 2002). Likewise, our autistic participants did not show a tendency to fixate on the hidden object location at the end of the videos in the false-belief high-demand condition where it remained on the scene. This indicates that a true-belief bias is not preventing them from passing the task. Therefore, these results cannot be explained by a manifestation of submentalizing, attentional differences or a true-belief bias.

2.4.2 Belief Reasoning Task Implementation

We found a main effect of Belief and an interaction between Time and Belief, neither of which interacted with Group, with participants showing more of a belief-congruent bias in the true- than false-belief condition in the first half of the task, although no difference in the second half. While both types of condition involve belief reasoning, it is likely that false-belief reasoning requires more sophisticated mentalizing abilities than true-belief reasoning due to the need to represent an alternative mental state that differs from one's own. Indeed, true-belief conditions have been consistently performed better than false-belief conditions in the literature (Surian & Geraci, 2012; Wang & Leslie, 2016), which indicates that it is easier than false-belief reasoning. Consistent with this, Nijhof et al. (2018) observed that the right temporoparietal junction was recruited in both true-belief and false-belief reasoning, but more so during false-belief than true-belief conditions. Similarly, Schneider et al. (2014) found the same pattern in the superior temporal sulcus. Accordingly, given the differences in performance in our task and differences in brain activation in the literature, false-belief reasoning seems to require a greater degree of mentalizing than true-belief reasoning, although this difference may diminish with exposure. This idea that mentalizing is involved in both true-belief and false-belief reasoning also helps to explain the marginal main effect of Group, whereby autistic participants showed a tendency to look less at the belief-congruent location across both Belief conditions, despite performing above chance on the true-belief condition but at chance on the false-belief condition.

Considering our false-belief high-demand condition, neither the looking bias in either group nor any differences between groups were observed. Although these results support that autistic adults did not affect by a true-belief bias, it might indicate that this condition failed to elicit mentalizing in non-autistic adults, consistent with findings from Wang and Leslie

(2016). However, this explanation seems unlikely as the two false-belief scenarios differed only in the final location of the object. A more plausible explanation is that mentalizing test performance, as indexed by anticipatory eye gaze, is also subject to the availability of executive resources, especially inhibitory control (Schneider, Lam, et al., 2012; Wang & Leslie, 2016). Both the true-belief bias hypothesis (Wang & Leslie, 2016) and the similarity-contingency model (Ames, 2004) propose that non-autistic adults may default to utilizing their own mental states as a basis on which to mentalize others. Taken together with the inhibition model of mentalizing proposed by Leslie and Polizzi (1998), it seems likely that the inability to inhibit the specified own belief in the false-belief high-demand condition may lead to greater uncertainty in non-autistic adults when predicting the agent's action (Wang & Leslie, 2016).

2.4.3 Advantages & Limitations of This Study

Overall, we found that our multi-trial design was sensitive to detecting group differences in mentalizing. We expected that increasing the number of trials would effectively decrease not only the dropout rate but also the error variance, addressing concerns from Dang et al. (2020) and Kulke, Wübker, et al. (2019). Through comparing Sections A and B, the group difference was held in both sections, consistent with Schneider et al. (2013); however, the difference between true- and false-belief conditions only existed in Section A. Accordingly, to maximise the sensitivity of the task, future research could remove the later blocks and the false-belief high-demand and true-belief short-turn conditions, and instead increase the number of false-belief low-demand and true-belief long-turn trials in a single section.

Although we have been able to address some important issues that until now have remained unanswered, an obvious limitation to this study was that all participants were adults and had average-to-high IQs, thus our findings cannot be generalized to autistic adults with language delay and/or intellectual disability, or autistic children. Still, this non-verbal paradigm holds promise as being adaptable to a much wider range of individuals than traditional mentalizing tests. Indeed, future studies should investigate whether spontaneous mentalizing varies between autistic individuals with different levels of general ability.

2.5 Conclusion

In closing, we extended Southgate et al. (2007)'s paradigm to critically examine spontaneous mentalizing in autism through a multi-trial, multi-condition eye-tracking study with a more nuanced analysis of eye movements over the timecourse of each trial. Replicating the findings of Senju et al. (2009), we found that although many autistic individuals perform well in explicit mentalizing tasks, they do not engage in spontaneous false-belief reasoning in implicit tasks, consistent with their everyday social difficulties. We have been able to rule out alternative theoretical explanations for this pattern of performance, leading to a better understanding of mentalizing in both non-autistic and autistic individuals. We have presented evidence that autistic adults are capable of processing information from social cues in the same way as non-autistic adults but that this information is not then used to update alternative mental representations. Future studies should directly test the point at which implicit mental state reasoning breaks down in autism.

Chapter 3. Evaluative Contexts Facilitate Implicit Mentalizing: Relation to the Broader Autism Phenotype and Mental Health

Abstract

One promising account for autism is implicit mentalizing difficulties. However, this account and even the existence of implicit mentalizing have been challenged because the replication results are mixed. Those unsuccessful replications may be due to the task contexts not being sufficiently evaluative. Therefore, the current study developed a more evaluative paradigm by implementing a prompt question. This was assessed in 60 non-autistic adults and compared with a non-prompt version. Additionally, parents of autistic children are thought to show a genetic liability to autistic traits and cognition and often report mental health problems, but the broader autism phenotype (BAP) is an under-researched area. Thus, we also aimed to compare 33 BAP and 26 non-BAP mothers on mentalizing abilities, autistic traits, compensation and mental health. Our results revealed that more evaluative contexts can facilitate implicit mentalizing in BAP and non-BAP populations, and thus improve task reliability and replicability. Surprisingly, BAP mothers showed better implicit mentalizing but worse mental health than non-BAP mothers, which indicates the heterogeneity in the broader autism phenotype and the need to promote BAP mothers’ psychological resilience. The findings underscore the importance of contexts for implicit mentalizing and the need to profile mentalizing and mental health in BAP parents.

3.1 Introduction

3.1.1 The Challenges and Solutions of Southgate et al. (2007)'s Paradigm

In Chapter 2, we addressed the first two criticisms about the reliability of Southgate et al. (2007)'s paradigm mentioned in Chapter 1 (i.e., *single-trial design* and *no true-belief control condition*), and critically examined spontaneous mentalizing in autism. These were achieved by developing a multi-trial, multi-condition eye-tracking study and conducting a nuanced analysis of gaze patterns during key events. We found that although many autistic participants perform well in explicit mentalizing tasks, they do not engage in spontaneous false-belief reasoning in implicit tasks, which was consistent with the findings of Senju et al. (2009). We also found no group-specific patterns of attentional distribution on key events of the implicit task. Taken together, we concluded that autistic adults are capable of processing information from social cues in the same way as non-autistic adults but that this information is not then spontaneously used to update alternative mental representations.

This chapter is going to primarily focus on the last challenge of Southgate et al. (2007)'s paradigm that the current thesis identified in the literature. As reviewed in Chapter 1 (see detailed discussion in Section 1.1.4.2), it has been criticized that this paradigm might not be sufficiently engaging to elicit implicit mentalizing (Kulke & Hinrichs, 2021; Kulke, Johannsen, et al., 2019; Schuwerk et al., 2018; Woo et al., 2023). Both Kulke and Hinrichs (2021) and Woo et al. (2023) have emphasised the importance of social context. Kulke and Hinrichs (2021) argued that mentalizing may not be prioritized if observers know there would not be any social consequence for not doing it. Consistently, Woo et al. (2023) suggested that social contexts where agents' actions have interactive potential can facilitate mentalizing. Accordingly, they proposed that the failed replications using Southgate et al. (2007)'s paradigm may have non-evaluative contexts, which provide observers less reason to care about agents' mental states, as their actions are irrelevant. Therefore, it is necessary to create more engaging implicit mentalizing paradigms to address this criticism and investigate whether the evaluability of context can modulate mentalizing.

3.1.2 Broader Autism Phenotype (BAP)

The Broader Autism Phenotype (BAP) was proposed to indicate a collection of sub-clinical expressions of autistic traits (Green et al., 2019; Ingersoll, Hopwood, et al., 2011; Piven et al., 1997; Sucksmith et al., 2011; Wainer et al., 2011). The BAP is qualitatively similar to autism, but neither leads to the full autism phenotype nor results in significant difficulties in socio-cognitive functioning (Green et al., 2019; Piven et al., 1997). Studies have observed that BAP is especially prevalent in the relatives of autistic people, for example, 20-40% of first-degree relatives but 2-9% of the general population (Green et al., 2019), indicating that autism is highly heritable (An et al., 2021; Freitag et al., 2010; Hill & Frith, 2003). Parents of autistic children (BAP) are about three times more likely to have autistic traits than parents of non-autistic children (non-BAP), especially in communication and social skills domains (Bishop et al., 2004; Bora et al., 2017; Sasson et al., 2013). Importantly, autistic traits in BAP parents are extremely heterogeneous: Rubenstein and Chawla (2018) found great variation in prevalence rates of the BAP across studies, ranging from 2.6% to 80% (e.g. Rubenstein & Chawla, 2018; Sucksmith et al., 2011; Wheelwright et al., 2010).

BAP populations have been found to have similar social cognition challenges as autistic people (Gliga et al., 2014; Green et al., 2019; Rea et al., 2019). Relatives of autistic people have moderate difficulties in mentalizing compared to non-autistic and autistic people (Gliga et al., 2014; Green et al., 2019), and people with higher self-reported autistic traits show more difficulties in mentalizing (Stewart et al., 2020). Moreover, Livingston et al.

(2019) found that BAP co-twins have mentalizing difficulties but can compensate for them at a behavioural level, which may potentially cause missed or late diagnosis (Mandy & Tchanturia, 2015). Interestingly, compensation at the behavioural level, which may potentially cause missed or late diagnosis, has been observed more in autistic females than males (Hull et al., 2020; Lai et al., 2017; McQuaid et al., 2022; Wood-Downie et al., 2021). It is possible that genetically predisposed individuals and those who are more likely to compensate have therefore been excluded from both the mentalizing and compensation literature because they do not meet the diagnostic criteria under the current clinical approaches (Hull et al., 2017). Thus, implicit mentalizing and how its difficulties might be compensated in BAP populations, and BAP females in particular, have yet to be fully understood (Green et al., 2019; Livingston & Happé, 2017). It is essential to explore BAP females' socio-cognitive functioning (Green et al., 2019; Ingersoll, Hopwood, et al., 2011), which could, in turn, improve understanding of the endophenotypes of autism (An et al., 2021; Billeci et al., 2016; Palmen et al., 2005).

One way to identify BAP females is as the mothers of autistic children. BAP mothers are also more vulnerable to mental health problems, such as depression and anxiety, compared with non-BAP mothers (Carpita et al., 2020; DeMyer, 1979; Ekas et al., 2010). First, parenting and caring for an autistic child can be stressful (Bishop et al., 2007; Bitsika et al., 2013; Ekas et al., 2010). Second, given that levels of autistic traits are associated with mental health outcomes, the elevated prevalence of autistic traits in BAP mothers may increase mental health problems (Bolton et al., 1998; Ingersoll & Hambrick, 2011; Ingersoll, Meyer, et al., 2011; Micali et al., 2004; Pruitt et al., 2018; Sucksmith et al., 2011). Third, if BAP mothers engage in greater compensation, the heightened mental health problems in BAP relatives may result from the cost of compensation (Livingston & Happé, 2017). Given mothers' primary role in parenting, it is vital to examine the relationship between BAP characteristics and mental health in mothers.

3.1.3 The Current Study

The primary aim of the current study was to develop a more evaluative paradigm that provides more reason for eliciting mentalizing. To make Southgate et al. (2007)'s paradigm more evaluative, a question was added, prompting observers to anticipate agents' actions. It might be argued that the prompt question might transform the task into an explicit task. Notably, only action anticipation, but not mentalizing, was prompted, to keep it implicit. Moreover, since eye-tracking has been considered as an applied implicit evaluation technique and widely used in autism research (Mazza et al., 2020; Senju et al., 2009), eye gaze as the outcome measure is implicit. Thus, the task did not make or require any explicit statement about mentalizing (Kulke & Rakoczy, 2019). So far, more versus less socially evaluative contexts have not been compared directly in any replications (Woo et al., 2023), thus we also include a comparable non-prompt version. According to Woo et al. (2023)'s proposal, the prompt implicit mentalizing task should enhance mentalizing compared with the non-prompt version. Additionally, according to our findings in Chapter 2, we set out to employ a multi-trial design and include matched true-belief conditions to improve task reliability and replicability. This prompt paradigm would be evaluated in a sample of non-autistic young adults, and compared with a comparable non-prompted version to examine its potential to facilitate implicit mentalizing. According to Woo et al. (2023), we hypothesized that the prompt task would be better at enhancing belief reasoning than the non-prompt version.

By using the prompt task, our second aim was to identify the differences between BAP and non-BAP mothers in implicit and explicit mentalizing abilities, autistic traits, compensatory tendencies and mental health outcomes. According to the existing literature, we predicted that BAP mothers would perform less well in mentalizing tasks, and reported more autistic traits, compensatory tendencies and mental health problems than non-BAP mothers. Last but not least, we aimed to explore how the aforementioned factors might relate to and predict implicit mentalizing performance in a non-clinical sample with sufficient statistical power.

3.2 Method

3.2.1 Participants

Two samples, a total of 128 participants, were recruited. In the *traits sample*, 68 participants from a local participant database were tested, aged 18-38 years (see demographics in Tables 3.1 & 3.2). Five participants were excluded because of poor data quality (see *Data pre-processing* below), and two who reported an autism diagnosis were excluded from data analyses. Given the majority of the sample is college students, we reasonably assumed that they had average-to-high IQs which therefore were not tested.

In the *mother sample*, 60 participants took part but one was excluded from the analysis due to poor data quality (see *Data pre-processing* below), leaving 33 mothers of autistic children (BAP mothers), aged 20-57 years; and 26 mothers of non-autistic children (non-BAP mothers), aged 28-60 years. They were recruited through autism support groups in London, and advertisements placed around the local community. All participants in the BAP group stated that at least one of their children has an autism diagnosis from a qualified clinician but not themselves. None of the non-BAP mothers reported or were known to have a diagnosis of psychiatric or neurodevelopmental conditions or related family history. To avoid confounding variables, the two groups were required to be matched on age, handedness, highest education, and IQ as measured by the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II; Wechsler, 2011) (see Tables 3.2 & 3.3).

Participants in both samples were required to be fluent in English and have normal or corrected-to-normal vision and hearing. Ethical approval for the study was received from the UCL Research Ethics Committee and all methods were performed in accordance with the approved guidelines and regulations. All participants gave written informed consent and were reimbursed for their time and effort.

Table 3.1. ***Descriptive statistics of the traits sample, Mean (Standard Deviation).***

	Traits (*n* = 61)
Age	21.97 (4.97)
Gender	Females (67.2%)
	Males (32.8%)
Handedness	Right (91.8%)
	Left (8.2%)
Education	High school (16.4%)
	UG[f] (47.5%)
	PG[g] (36.1%)
Anxiety (STAI Y-2[a])	39.27 (12.45)
Depression (BDI[b])	8.40 (9.36)
Autistic traits (AQ[c])	17.72 (7.66)
Autistic traits (BAPQ[d])	2.80 (0.70)
Camouflaging (CAT-Q[e])	3.40 (0.93)

Note. [a]STAI Y-2 = Spielberger State-Trait Anxiety Inventory Form Y-2; [b]BDI = Beck Depression Inventory; [c]AQ = Autism-Spectrum Quotient; [d]BAPQ = Broad Autism Phenotype Questionnaire; [e]CAT-Q = Camouflaging Autistic Traits Questionnaire; [f]UG = undergraduate; [g]PG = postgraduate.

Table 3.2. ***Descriptive statistics of the mothers sample, Mean (Standard Deviation).***

	BAP (*n* = 33)	Non-BAP (*n* = 26)
Age	42.55 (7.62)	41.73 (6.97)
Handedness	Right (87.9%)	Right (88.46%)
	Left (12.1%)	Left (11.54%)
Education	High school (21.2%)	High school (23.1%)
	UG[g] (48.5%)	UG (34.6%)
	PG[h] (30.3%)	PG (42.3%)
IQ (WASI-II[a]) with range	107.09 (12.82): 81-132	106.23 (13.21): 72-133
Anxiety (STAI Y-2[b])	44.41 (10.36)	39.66 (7.92)
Depression (BDI[c])	13.11 (9.09)	8.12 (6.23)
Autistic traits (AQ[d])	16.94 (8.28)	15.81 (6.36)
Autistic traits (BAPQ[e])	2.93 (0.92)	2.70 (0.59)
Camouflaging (CAT-Q[f])	3.08 (1.17)	2.75 (0.84)

Note. [a]WASI-II = Wechsler Abbreviated Scale of Intelligence, Second Edition; [b]STAI Y-2 = Spielberger State-Trait Anxiety Inventory Form Y-2; [c]BDI = Beck Depression Inventory; [d]AQ = Autism-Spectrum Quotient; [e]BAPQ = Broad Autism Phenotype Questionnaire; [f]CAT-Q = Camouflaging Autistic Traits Questionnaire; [g]UG = undergraduate; [h]PG = postgraduate.

3.2.2 Procedure

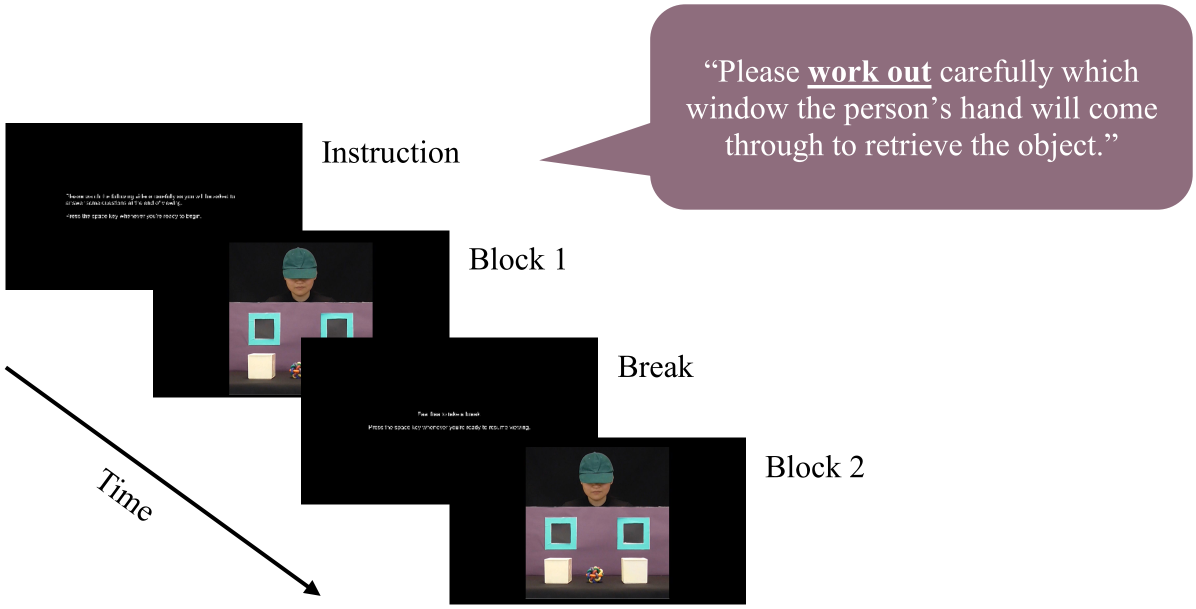
Participants started the session by completing a demographic questionnaire, then a non-prompt implicit mentalizing task and a prompt version of the task, followed by the WASI-II (not for the *traits sample*) and an explicit mentalizing task. The session finished with a series of questionnaires measuring individual differences in autistic traits, camouflaging behaviour, anxiety, and depression. Participants were then fully debriefed. The overall duration of the experiment was two hours. One participant's non-prompt task data in the *traits sample* were excluded as they did the prompt task before the non-prompt task. Testing was conducted either in participants' homes or in the Institute of Cognitive Neuroscience, University College London.

3.2.3 Implicit Mentalizing Tasks

The implicit mentalizing tasks were adapted from the anticipatory-looking paradigm in Senju et al. (2009), based on Southgate et al. (2007)'s classic false-belief task. One debriefing question was administered after the two tasks to investigate whether participants were aware of any differences between the non-prompt and prompt tasks.

Prompt task. Participants were prompted to reason about the actor's mental state by asking them to predict her behaviour (see *Figure 3.1*). In order to accurately predict the behaviour, they needed to be able to mentalize the actor's belief. Participants were instructed to work it out in their minds, not answer out loud. There were 2 types of false-belief conditions and 2 types of true-belief conditions (see *Figure 3.2*). The false-belief conditions included a Book condition in which the puppet removed the object from the scene while the agent was reading a book, and a Turn condition in which the puppet removed the object from the scene while the agent was distracted by the doorbell. Thus, observers should have different beliefs of the object's location in false-belief conditions than the agent. The corresponding matched true-belief conditions included a Book condition in which the puppet moved the object out of a box and then back to the same box while the agent was reading a book, and a Stretch condition in which the agent came back to watch the scene after a quick stretching. Accordingly, both observers and the agent should have the same belief of the object's location in true-belief conditions. The agent's head always followed the puppet's movement when she could see it, to indicate her attention.

The prompt task contained 2 experimental blocks (see *Figure 3.1*). Each block had 2 trials of each of the 4 conditions. All participants watched the same pseudorandomized sequence of the trials to reduce inter-individual variability. The box where the puppet put the object, the hand that held the puppet and the side the agent's head turned to were counterbalanced across the videos. An eye tracker was used to measure whether participants could predict which window the agent would open to retrieve the object by making anticipatory eye movements. If participants are mentalizing, they should look at the window/box which is consistent with the agent's belief about the location of the object (*belief-congruent*). This task was 15 minutes long with 1 break. Eye movements were recorded. Two questions were asked at the end of the task to encourage concentration. The questions asked about basic features of the videos (e.g. the colour of the puppet) and participants' judgements (e.g. the most frequent final location of the object), but participants were not informed of the style of question in advance to avoid directing their attention to particular features of the videos.



Please watch the following video carefully as you will be asked to answer some questions at the end of viewing.

Press the space key whenever you're ready to begin.

Feel free to take a break.

Press the space key whenever you're ready to resume viewing.

"Please **work out** carefully which window the person's hand will come through to retrieve the object."

Figure 3.1. **Prompt implicit mentalizing task procedure.**

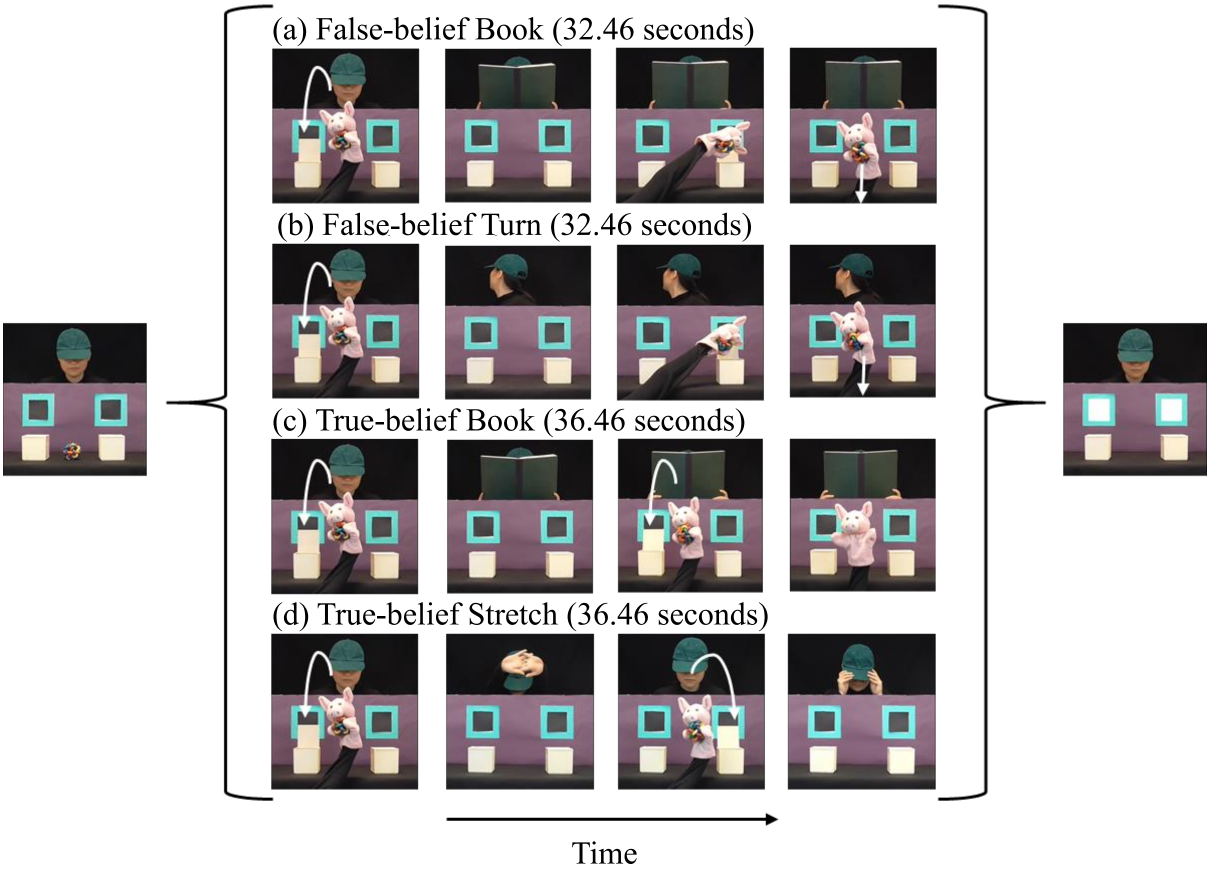

(a) False-belief Book (32.46 seconds)
(b) False-belief Turn (32.46 seconds)
(c) True-belief Book (36.46 seconds)
(d) True-belief Stretch (36.46 seconds)
Time

Figure 3.2. **Selected key frames from the videos. (a) False-belief Book condition: the puppet removed the object from the screen, while the agent was reading a book; (b) False-belief Turn condition: the puppet removed the object from the screen while the agent was distracted by the doorbell; (c) True-belief Book condition: the puppet moved the object out of a box and then back to the same box while the agent was reading a book; (d) True-belief stretch condition: the agent came back to watch the scene after a quick stretching.**

Non-prompt task. The same stimuli were presented in the non-prompt task but participants were instructed to passively view the videos and answer some questions accordingly at the end. There was 1 familiarization block as well as 2 experimental blocks. The familiarization block enabled participants to implicitly learn the contingency that the agent was going to retrieve the object after the windows illuminated, which included 4 short and 4 long familiarization trials (see *Figure 3.3*). Specifically, in the short trials, the object was on one of the boxes, and then the agent's hand came through the window to retrieve it after the windows illuminated; while in the long trials, the puppet put the object into one of the boxes, and then the agent's hand came through the window to open the box and retrieve it after the windows illuminated.

Each experimental block also started with 1 short and 1 long familiarization trials, followed by the 2 trials of each of the 4 conditions. The task was 20 minutes long with 2 breaks. At the end of this task, to encourage the participant to concentrate, two questions were asked about the details in the videos; to check that this task examined implicit processing, an 8-item funnelled debriefing procedure, adapted from Schneider et al. (2014), was administered.

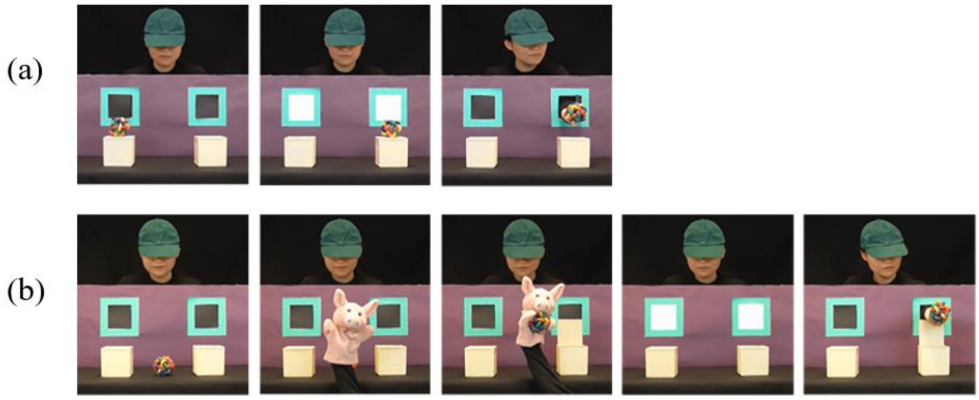


***Figure 3.3.* Familiarization trials. (a) Short trials: the object was on one of the boxes, and then the agent's hand came through the window to retrieve it after the windows illuminated. (b) Long trials: the puppet put the object into one of the boxes, and then the agent's hand came through the window to open the box and retrieve it after the windows illuminated.**

Apparatus. A remote screen-based Tobii Pro X3-120 eye-tracker system, with a sampling rate at 120Hz, was used to record eye movements (Tobii, Sweden). Visual and auditory stimuli were presented via a Dell Precision 5520 laptop (15.6-inch) with Tobii Pro Studio 3.4.8 software, integrated with the eye-tracker. Participants sat approximately 70cm from the eye-tracker and were instructed to sit still throughout the eye-tracking assessment. A 5-point calibration was performed before each implicit task.

Areas of interest (AOIs). Data were coded from the windows illumination onset to the end of each video, with a total duration of 5 seconds in each trial. Two AOIs were identified: *Belief-congruent* and *Belief-incongruent* (see *Figure 3.4* as an example). Gaze data were extracted from both AOIs.

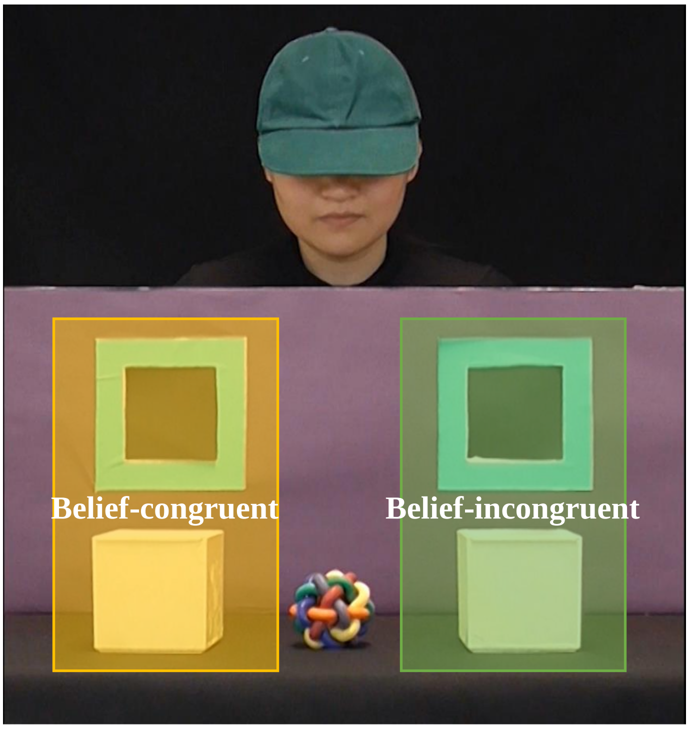


***Figure 3.4.* An example of the areas of interest: *Belief-congruent* (yellow) and *Belief-incongruent* (green).**

Fixation analysis. Data points with angular velocity below 30 degrees per second were classified as ***fixations*** (i.e. the visual gaze on a single location) while those above were ***saccades*** (i.e. the rapid eye movement between fixations). Two adjacent fixations with less than 75ms time interval or less than 0.50 degrees visual angle were merged as one fixation. Fixations with less than 60ms time duration were discarded. The total fixation duration was extracted, measuring the sum of the fixation durations within each AOI, by using Tobii Studio.

Dara pre-processing. Differential looking scores (DLS), which measure participants' looking preference between two visual targets, were calculated by dividing the difference between the total fixation duration to the *Belief-congruent* and *Belief-incongruent* AOIs by the sum of the two. DLS ranged from 1 to -1: closer to 1 if participants showed a looking bias towards the *Belief-congruent* AOI, closer to -1 if they were biased towards the *Belief-incongruent* AOI, and closer to 0 if they looked equally to both AOIs, equivalent to chance performance.

Three exclusion criteria were applied to ensure participants were paying attention to the task and the key events in the videos (e.g. watching the hand retrieving the object in the familiarisation trials). First, participants' data from a task were excluded if they missed more than 25% data from that task. Second, the data from the non-prompt task were excluded for any participant whose average DLS in the familiarization block was missing or below chance, to confirm that they had paid attention to the key event (a combination of the prediction and the action itself). Third, the data from each experimental block of the non-prompt task were excluded if the average DLS of the two familiarization trials at the beginning of that block was missing or below chance. Accordingly, five *traits* participants and one BAP mother were excluded from the whole analysis.

3.2.4 Explicit Mentalizing Task

The Strange Stories Task is an advanced mentalizing test assessing participants' ability to explicitly infer both *Mental States* and *Physical States* (White et al., 2009). In this study, only the 8 *Mental States Stories* were used; accuracy scores therefore ranged from 0-16.

3.2.5 Self-reported Measures

Autistic traits were measured by the widely used Autism-Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, et al., 2001) and the Broad Autism Phenotype Questionnaire (BAPQ; Hurley et al., 2007), with higher scores indicating more autistic traits.

The AQ ranges between 0-50, Cronbach's α = 0.90; the BAPQ between 1-6, α = 0.94. The BAPQ was also employed as it was specifically designed in a sample of BAP parents (Hurley et al., 2007), and showed superior internal consistency when compared with the AQ (Ingersoll, Hopwood, et al., 2011). Social camouflaging (or compensatory) behaviours were measured by the Camouflaging Autistic Traits Questionnaire (CAT-Q; Hull et al., 2019), with higher scores indicating more strategies employed to cope with autistic characteristics during social interactions, ranging between 1-7, α = 0.92.

Anxiety traits were measured by the Spielberger State-Trait Anxiety Inventory Form Y-2 (STAI Y-2; Spilberger, 1983), with higher scores corresponding to more severe anxiety traits, ranging between 20-80, α = 0.92. Depression was measured by the Beck Depression Inventory (BDI; Beck et al., 1988), with higher scores indicating more severe depressive symptoms, ranging between 0-63, α = 0.90. Item 9 regarding Suicidal thoughts was removed for ethical reasons. Missing values (item *n* = 16, with the number of missing responses less than 25% of the total number of items on each of these measures) were imputed using the individual's mean scores of the scale or the sub-scale.

3.3 Results

All effects are reported as significant at *p* < .05, and two-tailed *p* values were reported throughout, if not specified. Statistical analyses were conducted using IBM SPSS Statistics (Version 29).

3.3.1 Validity of Implicit Mentalizing Tasks

One-sample *t*-tests were conducted on the false-belief and true-belief DLS of both implicit tasks in the *traits sample*. The results showed that both false-belief and true-belief DLS were significantly above zero in the prompt task: false-belief: $t(60) = 2.96$, $p = .004$, $d = 0.38$, true-belief: $t(60) = 8.65$, $p < .001$, $d = 1.11$ (see *Figure 3.5*). However, in the non-prompt task, only the DLS for the true-belief condition, but not for the false-belief condition, was significantly above chance: false-belief ($M = -0.05$, $SD = 0.25$): $t(59) = -1.44$, $p = .154$, $d = -0.19$, true-belief ($M = 0.09$, $SD = 0.27$): $t(59) = 2.64$, $p = .011$, $d = 0.34$. Since the non-prompt task therefore showed poor validity, all non-prompt data were excluded in the following analyses. A paired samples *t*-test on the false-belief and true-belief DLS of the prompt task revealed that the performance in the true-belief condition was significantly better than the false-belief in the *traits sample*, $t(60) = -4.79$, $p < .001$, $d = -0.89$ (see *Figure 3.5*).

3.3.2 Comparing the Mother Groups

Self-report measure: As expected, compared with non-BAP mothers, BAP mothers scored significantly higher in anxiety (marginal) and depression, but unexpectedly not in autistic traits and camouflaging behaviour (see Tables 3.2 & 3.3).

Table 3.3. ***Group-wise comparison between the BAP and non-BAP groups.***

	Inferential statistic
	BAP (*n* = 33) vs Non-BAP (*n* = 26)
Age	*t*(57) = 0.42, *p* = .674, *d* = 0.11
Handedness	$\chi^2(1) = 0.005$, *p* = .945
Education	$\chi^2(2) = 1.27$, *p* = .529
IQ (WASI-II[a])	*t*(57) = 0.25, *p* = .802, *d* = 0.07
Anxiety (STAI Y-2[b])	***t*(57) = 1.93, *p* = .059, *d* = 0.51 (marginal)**
Depression (BDI[c])	***t*(57) = 2.39, *p* = .020, *d* = 0.63**
Autistic traits (AQ[d])	*t*(57) = 0.25, *p* = .802, *d* = 0.15
Autistic traits (BAPQ[e])	*t*(55.04) = 1.16, *p* = .250, *d* = 0.29
Camouflaging (CAT-Q[f])	*t*(56.61) = 1.28, *p* = .206, *d* = 0.32

Note. [a]WASI-II = Wechsler Abbreviated Scale of Intelligence, Second Edition; [b]STAI Y-2 = Spielberger State-Trait Anxiety Inventory Form Y-2; [c]BDI = Beck Depression Inventory; [d]AQ = Autism-Spectrum Quotient; [e]BAPQ = Broad Autism Phenotype Questionnaire; [f]CAT-Q = Camouflaging Autistic Traits Questionnaire.

Implicit mentalizing. A two-way mixed-design analysis of variance (ANOVA) was conducted using the DLS as the outcome variable, Belief (false-belief, true-belief) as a within-subjects factor, and Group (BAP, Non-BAP) as a between-subjects variable. There were significant main effects of Belief, *F*(1, 57) = 29.88, *p* < .001, *partial* η^2 = .344, and Group, *F*(1, 57) = 5.23, *p* = .026, *partial* η^2 = .084, but no interaction. Similar to the *traits* *sample*, the true-belief condition had a higher DLS than the false-belief condition, but interestingly, BAP mothers scored higher than non-BAP mothers (see *Figure 3.5*).

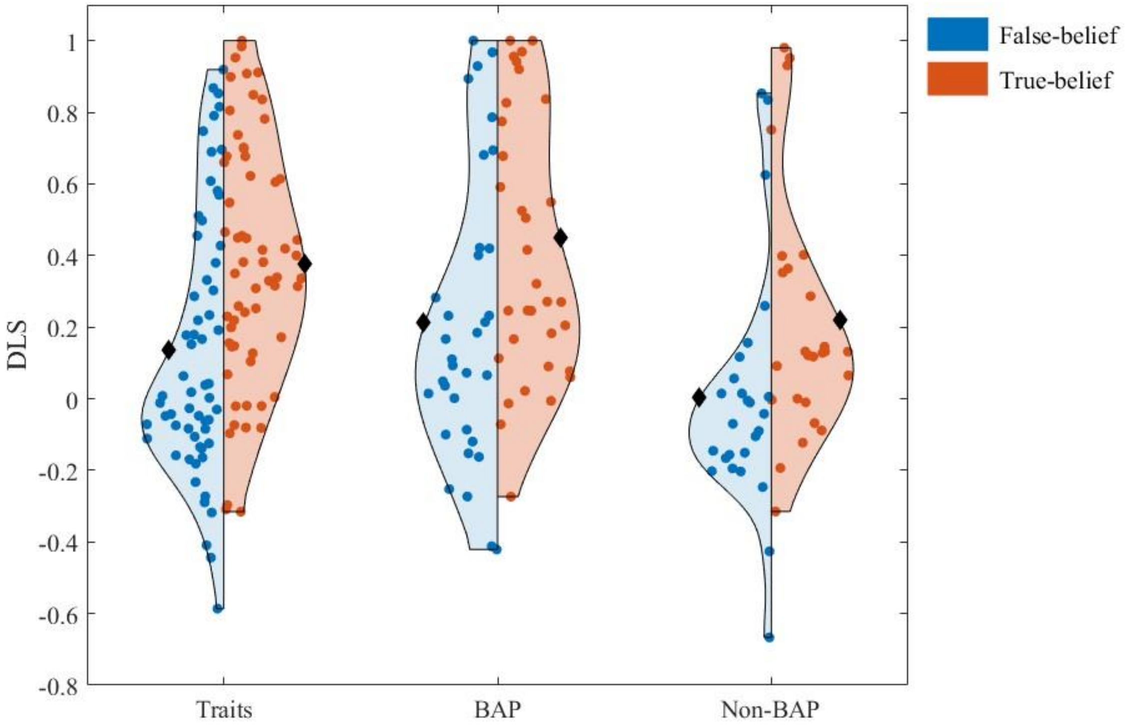


***Figure 3.5.* False-belief and True-belief DLS of the prompt task in the *traits* and *mother samples* (each dot represents the score of each participant); diamonds represent the mean of each condition.**

Explicit mentalizing. An independent samples *t*-test revealed that performance on the Strange Stories Task was comparable between the BAP (*M* = 12.72, *SD* = 2.49) and non-BAP (*M* = 13.15, *SD* = 1.80) groups, *t*(57) = -0.73, *p* = .466, *d* = -0.19.

3.3.3 Relationships

Given all participants in the *traits* and *mother samples* did not have an autism diagnosis, we combined the two samples to achieve an ideal statistical power for correlation and regression analyses. As the false-belief and true-belief conditions in the prompt implicit mentalizing task had a moderate-to-strong positive correlation, $r = .55$, $p < .001$, and there was no interaction between Belief and Group in the *mother sample*, these two conditions were merged by calculating the mean of each participant for the following analyses.

Correlations. Correlations were investigated among the performance on implicit mentalizing (prompt task DLS) and explicit mentalizing (Strange Stories task accuracy), individual differences in autistic traits (AQ, BAPQ), camouflaging (CAT-Q), anxiety (STAI Y-2) and depression (BDI), and age. Higher implicit mentalizing performance was significantly correlated with higher explicit mentalizing performance and with lower autistic traits (BAPQ) (see *Figures 3.6* & *3.7* and Table 3.4). Age was positively related to depression (see Table 3.4). However, these relationships would not withstand correction for multicomparison. As expected, self-reported autistic traits (AQ, BAPQ), camouflaging, anxiety and depression were highly correlated with each other (see Table 3.4). A relationship between implicit mentalizing and autistic traits was observed with the BAPQ, but not the AQ; the former was therefore considered more sensitive in detecting autistic traits in a non-clinical population, in keeping with the existing literature (Broderick et al., 2015; Ingersoll, Hopwood, et al., 2011), and so the BAPQ was employed in the following regression analysis.

Table 3.4. ***Correlations (r) among the mentalizing performances, individual differences in autistic traits, camouflaging, mental health, and age.***

	1	2	3	4	5	6	7	8
1 Implicit mentalizing (Prompt DLS)	-	**0.20***	-0.17	**-0.20***	-0.004	-0.04	-0.10	-0.10
2 Explicit mentalizing (Strange Stories)		-	-0.16	-0.16	-0.08	-0.02	-0.08	0.06
3 Autistic traits (AQ[a])			-	**0.83*****	**0.56*****	**0.53*****	**0.45*****	-0.05
4 Autistic traits (BAPQ[b])				-	**0.59*****	**0.62*****	**0.48*****	0.03
5 Camouflaging (CAT-Q[c])					-	**0.42*****	**0.29*****	-0.14
6 Anxiety (STAI Y-2[d])						-	**0.68*****	0.14
7 Depression (BDI[e])							-	**0.20***
8 Age								-

Note. *p < .05, ***p < .001. Pearson's correlation coefficients (r) are reported. [a]AQ = Autism-Spectrum Quotient; [b]BAPQ = Broad Autism Phenotype Questionnaire; [c]CAT-Q = Camouflaging Autistic Traits Questionnaire; [d]STAI Y-2 = Spielberger State-Trait Anxiety Inventory Form Y-2; [e]BDI = Beck Depression Inventory.

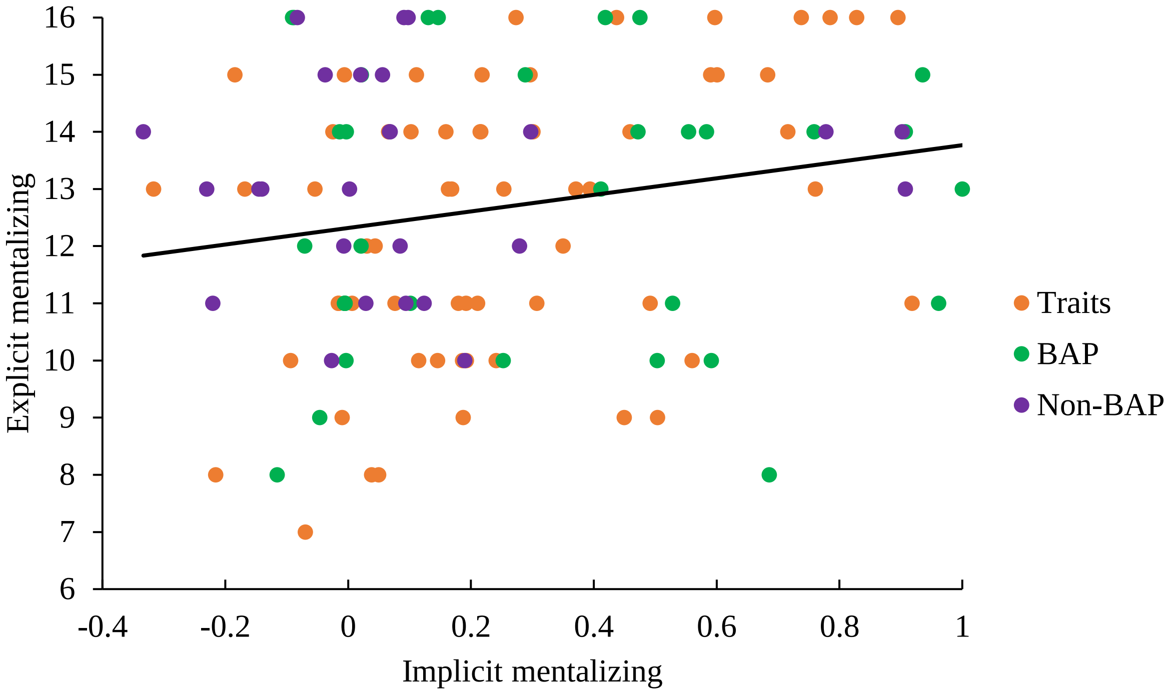


Figure 3.6. **Correlation scatter plot between the DLS of the prompt implicit mentalizing task and the accuracy of the strange stories task measuring explicit mentalizing ability (each dot represents a participant).**

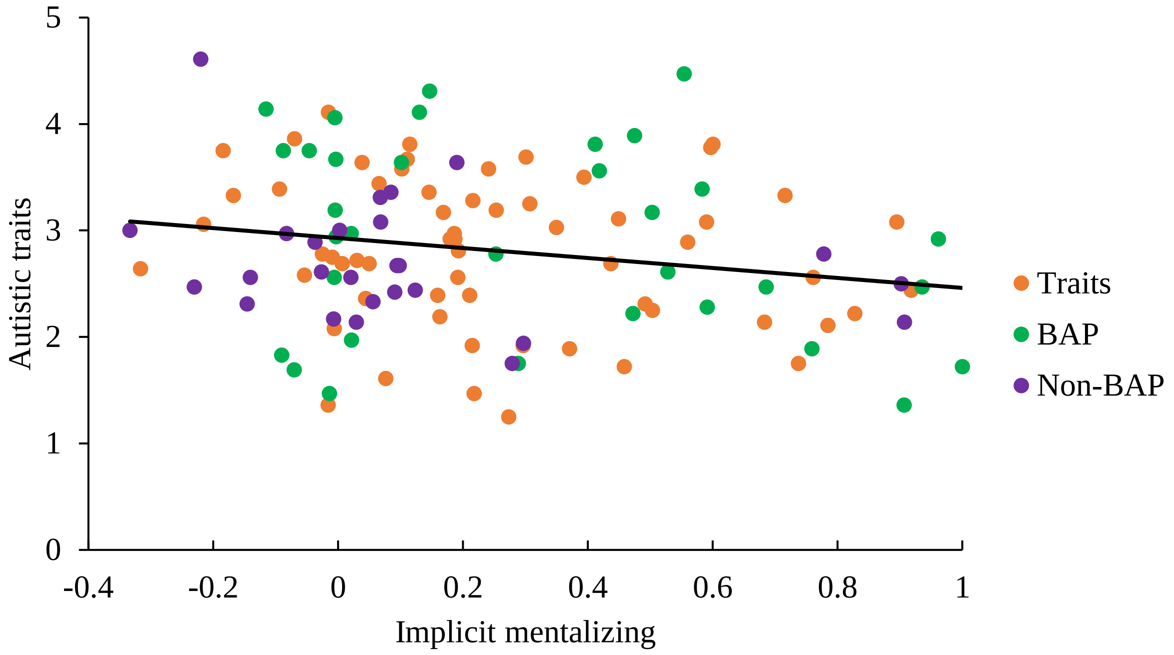


Figure 3.7. **Correlation scatter plot between the DLS of the prompt implicit mentalizing task and autistic traits measured by the BAPQ (each dot represents a participant).**

Regression. Multiple linear regression (enter method) was carried out with implicit mentalizing performance as the dependent variable and explicit mentalizing performance, age, autistic traits, camouflaging, anxiety, depression, and groups as potential predictors. Groups were coded as two dummy variables: BAP (BAP = 1, non-BAP & traits = 0), and non-BAP (non-BAP = 1, BAP & traits = 0), with *traits* as the reference category). The VIF values were below 3.89 and the tolerance statistics were above 0.26, which represents no multicollinearity. Results revealed that this model was significantly better at predicting the implicit DLS than using the mean of it, $F(8, 111) = 2.80$, $p = .007$, $R^2 = 0.17$. The individual predictors were examined and showed that autistic traits and explicit mentalizing (marginal) were significant predictors of implicit mentalizing (see Table 3.5).

Table 3.5. ***Multiple linear regression model of predictors of the implicit mentalizing performance (the prompt DLS), n = 120.***

	b	*SE b*	*β*	*t*(112)	*p*
Constant	0.196	0.238		0.824	.412
Explicit mentalizing (Strange Stories)	**0.023**	**0.012**	**0.166**	**1.862**	**.065 (marginal)**
Age	-0.005	0.005	-0.196	-1.148	.253
Autistic traits (BAPQ^{a})	**-0.139**	**0.055**	**-0.323**	**-2.521**	**.013**
Camouflaging (CAT-Q^{b})	0.045	0.036	0.142	1.241	.217
Anxiety (STAI Y-2^{c})	0.004	0.004	0.145	1.077	.284
Depression (BDI^{d})	-0.004	0.004	-0.103	-0.846	.399
Group BAP	0.185	0.117	0.259	1.576	.118
Group non-BAP	-0.046	0.120	-0.060	-0.386	.701

Note. R^2 = 0.15. *b* = Unstandardized B; *SE b* = Coefficients Standard Error; *β* = Standardized Coefficients Beta. ^{a}BAPQ = Broad Autism Phenotype Questionnaire; ^{b}CAT-Q = Camouflaging Autistic Traits Questionnaire; ^{c}STAI Y-2 = Spielberger State-Trait Anxiety Inventory Form Y-2; ^{d}BDI = Beck Depression Inventory.

3.4 Discussion

The current study aimed to develop a more evaluative implicit mentalizing paradigm by implementing a prompt question, a multi-trial design and matched true-belief conditions to improve task reliability and replicability and assess it in a non-autistic young adult sample. We then explored the relationship between implicit and explicit mentalizing abilities, autistic traits, compensatory tendencies and mental health outcomes in a non-clinical sample with sufficient statistical power. Third, we compared the aforementioned abilities and characteristics between BAP and matched non-BAP mothers.

3.4.1 Prompt Task Validation

Three main pieces of evidence indicate that the prompt implicit mentalizing task is valid, and may be better at facilitating mentalizing than the non-prompt version. First, both true-belief and false-belief conditions were performed significantly above chance in a group of non-autistic adults, meaning that participants showed a looking bias towards the *belief-congruent* AOI in this task, which conceptually replicated previous findings (Schneider, Bayliss, et al., 2012; Schneider et al., 2017; Schuwerk et al., 2016; Schuwerk et al., 2018). Accordingly, non-autistic people are able to predict the agent's behaviour by implicitly reasoning about her mental state. This indicates that the task is able to facilitate mentalizing and elicit belief-based action prediction in the general population, supporting the prompt task as a valid implicit mentalizing task.

On the other hand, the false-belief condition in the non-prompt version did not differ from chance. That is, participants did not show a preference for the *belief-congruent* location in the false-belief condition, which is consistent with some previous unsuccessful replications from Kulke and colleagues (Kulke, Johannsen, et al., 2019; Kulke, Reiß, et al., 2018; Kulke, von Duhn, et al., 2018; Kulke, Wübker, et al., 2019). This suggests that the non-prompt task was unable to elicit false-belief reasoning, indicating that it might not be a reliable paradigm, which is inconsistent with our findings in Chapter 2 with a similar non-prompt task.

In line with our hypothesis, this preliminary evidence seems to suggest that the more evaluative prompted task is indeed better at facilitating mentalizing than the less evaluative non-prompt task, which is consistent with Woo et al. (2023)'s proposal. However, as we created our own stimuli to conceptually replicate Southgate et al. (2007)'s paradigm, and these stimuli are also different from the stimuli we used in Chapter 2, we cannot rule out the possibility that small variations in the non-prompt task resulted in its invalidity. One such deviation is that we removed the delay phase in the familiarization trials between the end of the audio-visual cue and the onset of the actor's action, to make the task more realistic. Schuwerk et al. (2018) suggested that their unsuccessful replication might be because this phase was too long to build up the contingency between the cue and the action. Similarly, Kulke and Hinrichs (2021) reported adult participants noticed the artificial waiting time and suggested a shorter and more realistic delay should improve task reliability. Removing it altogether may have been too drastic, however, and a delay may in fact be needed to establish the contingency; future studies should modify the timing and further investigate the importance of context in mentalizing (Woo et al., 2023). However, this also indicates that the non-prompt paradigm may be more fragile than the prompt version, needing more strict criteria to elicit mentalizing, which further confirms our primary hypothesis.

Second, the performances of explicit and implicit mentalizing were positively correlated, and, although borderline, the former can affect the latter to a degree, which is consistent with our prediction. This suggests that the two tasks may tap into overlapping cognitive mechanisms, confirming that the prompt implicit mentalizing task was valid for measuring mentalizing. Although implicit mentalizing and explicit mentalizing are thought to work both complementarily and oppositionally (Frith & Frith, 2008), EEG and fMRI studies have revealed that implicit and explicit mentalizing are elicited at about the same time and have a shared neural network, including the medial prefrontal cortex and the temporoparietal junction (Hyde et al., 2015; Naughtin et al., 2017; Van Overwalle & Vandekerckhove, 2013). Given general cognitive factors, such as language, memory and attention likely influence explicit more than implicit mentalizing (Abell, 2000; Apperly & Butterfill, 2009; Gliga et al., 2014; Happé, 1995; Ullman & Pullman, 2015), it is perhaps unsurprising that some studies have not shown relationships between implicit and explicit mentalizing performance (Grosse Wiesmann et al., 2017; Kulke, Wübker, et al., 2019; Nijhof et al., 2016) when other factors play a more key role in particular tasks.

Third, we found that autistic traits were not only negatively associated with but also affected implicit mentalizing, which indicates that higher autistic traits may be a sign of poor implicit mentalizing in non-autistic populations. This result replicates previous observations of negative correlations between autistic traits and implicit mentalizing in both autistic (Deschrijver et al., 2016) and non-autistic populations (Nijhof et al., 2017) and is consistent with the idea that autistic people may have specific difficulties in implicit mentalizing, while some develop relatively good explicit mentalizing later in development (Abell, 2000; Eisenmajer & Prior, 1991; Frith, 2004; Happé, 1995; Hull et al., 2017; Livingston et al., 2019; Livingston & Happé, 2017; Senju et al., 2009; Ullman & Pullman, 2015). Accordingly, we can more confidently state that the prompt task is able to authentically measure implicit mentalizing.

However, we did not replicate the relationship between autistic traits and explicit mentalizing previously reported in autistic people (Stewart et al., 2020). Also, neither type of mentalizing ability was correlated with compensatory tendencies, anxiety, depression, and age in the entire sample, and none of these four factors could account for variance in implicit mentalizing performance. Thus, we did not replicate Livingston et al. (2019)'s finding in autism that weaker mentalizing and lower autistic traits are related to higher mental health problems because of compensation. Again, this might be because autism does not result from explicit mentalizing difficulties or lead directly to higher compensatory tendencies or mental health problems. Other factors might play more essential roles in the development of explicit mentalizing and compensatory tendencies, like executive function or language (Abell, 2000; Happé, 1995; Livingston et al., 2019), as well as mental health outcomes (Hull et al., 2019; Hull et al., 2017; Lai et al., 2017; Livingston & Happé, 2017). Together with the fact that autistic traits are relatively low in non-autistic populations, the lack of associations observed in our non-clinical sample is understandable.

It is also possible that self-reported inventories for assessing autistic traits, compensation and mental health might measure the awareness or the perceived social expectations of these characteristics instead of genuine individual differences (Scheeren & Stauder, 2008). Although self-reported questionnaires are the most common instruments, which are money- and time-saving, these measures may be influenced by the BAP (Rea et al., 2019; Rubenstein et al., 2017; Sasson et al., 2013) and unconscious compensatory mechanisms (Hull et al., 2017; Lai et al., 2017), thus, more objective measures are needed in future studies (Hurley et al., 2007; Livingston et al., 2019; Lord et al., 2012; Pruitt et al., 2018).

We also replicated Surian and Geraci (2012) and Wang and Leslie (2016), but not Kulke, Reiß, et al. (2018), that true-belief attribution was positively correlated with false-belief attribution, and true-belief conditions were consistently performed better than false-belief conditions in all samples. This relationship has also been observed in neuroimaging studies. Nijhof et al. (2018) observed that the right temporoparietal junction was recruited in both true-belief and false-belief reasoning, and more so during false-belief than true-belief conditions, in both implicit and explicit mentalizing. Similarly, Schneider et al. (2014) found the same pattern in the superior temporal sulcus, but not in the rest of the mentalizing network. We can assume therefore that both true-belief and false-belief reasoning recruit mentalizing to a degree, but given the differences in accuracy in our task and differences in brain activation in the literature, false-belief reasoning requires higher mentalizing abilities than true-belief reasoning.

3.4.2 BAP

Surprisingly, BAP mothers performed better in the implicit but comparably in the explicit mentalizing tasks compared with non-BAP mothers, which is not consistent with Gliga et al. (2014)'s study of infant BAP siblings. One potential explanation is that, because of a lack of group difference also in autistic traits, the BAP mothers in our sample did in fact have strong implicit mentalizing abilities. Unlike many infant siblings, it is possible that some or all BAP mothers do not possess autistic traits or autistic cognitive profiles, are not genetically predisposed to autism themselves and hence do not contribute to their child's genetic predisposition.

However, the lack of group differences in autistic traits may not necessarily mean BAP and non-BAP mothers are indistinguishable. An et al. (2021) found that BAP mothers had smaller grey matter volumes in the right middle temporal gyrus, temporoparietal junction, cerebellum, and parahippocampal gyrus than non-BAP mothers, even when group differences in autistic traits were absent. This might suggest the presence of subtle underlying neurological differences despite a lack of autistic traits, or alternatively that our BAP mothers were not representative of the wider BAP mother population and were totally unaffected at the behavioural, cognitive and neurological level.

Alternatively, it might be that an interaction between protective factors and autistic advantages boosted BAP mothers' performance in the prompt task. BAP parents are believed to reflect an underlying genetic liability for autism (Sasson et al., 2013), for example, the shared genetic overlap between BAP mothers and their autistic children has been observed to be associated with the mothers' autistic traits (Nayar et al., 2021). Notably, autism is not only associated with social difficulties but also with remarkable skills and talents (Happé, 2018; Happé & Vital, 2009), for example, a detail-focused cognitive style (Happé & Vital, 2009). Together with the finding that females require more inherited factors than males to exhibit autism (Lockwood Estrin et al., 2021), BAP mothers might possess some protective factors that mean they display fewer autistic traits than their children, but reserve some autism-like cognitive styles that predispose them to better develop certain cognitive abilities than non-BAP mothers (Happé & Vital, 2009).

A third explanation is that the BAP mothers may have possessed higher motivation to engage in the task because of their autistic children, and therefore, performed better in the more passive implicit task. However, in the explicit task, engagement might not enhance performance, as the already highly evaluative context (Woo et al., 2023) may mean participants are already fully engaged. Although Southgate et al. (2007)'s paradigm is well-known in the literature and presumably in autism communities, it is unlikely that the BAP group knew the task expectations beforehand, otherwise, they might have also performed well in the non-prompt version.

Although no group difference was found in self-reported autistic traits and compensatory tendencies, BAP mothers reported higher levels of depressive and marginally higher levels of anxious symptoms than non-BAP mothers. These results support the idea that the mental health difficulties in BAP mothers might be more related to their chronic stress from parenting and caring for autistic children (Bishop et al., 2007; Bitsika et al., 2013; Ekas et al., 2010; Su et al., 2018) than their own autistic traits (Bolton et al., 1998; Ingersoll & Hambrick, 2011; Ingersoll, Meyer, et al., 2011; Micali et al., 2004; Pruitt et al., 2018; Sucksmith et al., 2011) or the cost of compensation (Livingston & Happé, 2017). However, the current study cannot rule out a multi-risk model of mental health outcomes in BAP mothers, as the BAP is highly heterogeneous in relatives of autistic people (Bora et al., 2017; Rubenstein & Chawla, 2018). On all accounts, support is needed to alleviate mental health issues and develop psychological resilience in BAP mothers (Bitsika et al., 2013).

In addition, positive correlations were reported among autistic traits, compensatory tendencies and mental health problems in the merged large sample. These findings are consistent with the extant literature that individuals with more socio-cognitive difficulties (Baron-Cohen, Wheelwright, Skinner, et al., 2001; Green et al., 2019; Hurley et al., 2007) need to allocate more cognitive resources to compensate for their core difficulties, which is likely to compromise their mental health in both autistic and non-autistic populations (Hull et al., 2017; Lai et al., 2011; Lai et al., 2017; Livingston et al., 2019; Livingston & Happé, 2017).

3.4.3 Advantages & Limitations of This Study

One advantage of the current study is the use of a prompt question in the implicit mentalizing task. This adaptation seemed to increase the evaluability of the task context, which makes the prompt anticipatory paradigm more robust in facilitating implicit mentalizing and therefore improves the task reliability and replicability (Kulke & Hinrichs, 2021; Woo et al., 2023). However, a corresponding limitation of our task design is that the non-prompt and prompt task order could not be counterbalanced. If the prompt task was performed first, the non-prompt task would logically become a prompt version. Nevertheless, it seems unlikely that the fixed procedure can account for our primary findings.

Another advantage lies in directing attention towards the BAP, an area that still holds significant gaps in understanding. This may not only have significant implications for autism research (An et al., 2021; Billeci et al., 2016; Palmen et al., 2005) but also better support families with autistic children (Bitsika et al., 2013). Nonetheless, because of the female sample, our results cannot be generalized to the entire BAP community, particularly as recent studies have suggested that the BAP is more prevalent in BAP fathers than mothers (De la Marche et al., 2015; Rubenstein & Chawla, 2018). Accordingly, the lack of group differences between our BAP and non-BAP mothers, especially in autistic traits, seems to imply that our BAP mothers did not have autistic characteristics. Future studies should include both parents to reveal patterns in the whole family and sex- and gender-informed phenotypes of autism (Hull et al., 2019; Karst & Van Hecke, 2012; Pruitt et al., 2018; Rea et al., 2019; Su et al., 2018).

We acknowledge two additional limitations. The current study employed a cross-sectional design, so the direction of the association between mentalizing abilities, autistic traits, compensation and mental health in our correlation and regression analyses cannot be conclusively determined. Although implicit mentalizing was defined as dependent variable in our regression analysis, there are potential alternative ways in which the independent and dependent variables might plausibly relate to one another. It is also possible that some of the relationships may be bidirectional. Future research should incorporate a longitudinal design to investigate the causality of these relationships. Furthermore, we had relatively small samples, especially the non-BAP sample, which may compromise the power to detect group differences. Future research would benefit from recruiting larger samples.

3.5 Conclusion

In closing, the current study developed a more evaluative implicit mentalizing task which was proved to be robust in facilitating false-belief and true-belief reasoning (Woo et al., 2023). With the adapted prompt task, we found that both explicit mentalizing and autistic traits are associated with implicit mentalizing but not with each other, which supports the idea of two distinct but overlapping mentalizing systems (Apperly & Butterfill, 2009) and implicit but not explicit mentalizing difficulties in autistic adults (Frith, 2004; Senju et al., 2009). However, BAP mothers showed better implicit mentalizing and poor mental health than non-BAP mothers, but no other differences, which indicates the heterogeneity within the broader autism phenotype (Bora et al., 2017; Rubenstein & Chawla, 2018) as well as the need to support families with autistic members in terms of mental health and psychological resilience (Bitsika et al., 2013). Future studies are needed to further examine the prompt task reliability and validity and investigate associations among autism, mentalizing, compensation and mental health in more clinical and sub-clinical populations.

Chapter 4. Can Membership Modulate the Social Abilities of Autistic People? An Intergroup Bias in Smile Perception

Abstract

Autistic adults struggle to reliably differentiate genuine and posed smiles. Intergroup bias is a promising factor that may modulate smile discrimination performance, which has been shown in neurotypical adults, and which could highlight ways to make social interactions easier. However, it is not clear whether this bias also exists in autistic people. Thus, the current study aimed to investigate this in autism using a minimal group paradigm. Seventy-five autistic and sixty-one non-autistic adults viewed videos of people making genuine or posed smiles and were informed (falsely) that some of the actors were from an in-group and others were from an out-group. The ability to identify smile authenticity of in-group and out-group members and group identification were assessed. Our results revealed that both groups seemed equally susceptible to ingroup favouritism, rating ingroup members as more genuine, but autistic adults also generally rated smiles as less genuine and were less likely to identify with ingroup members. Autistic adults showed reduced sensitivity to the different smile types but the absence of an intergroup bias in smile discrimination in both groups seems to indicate that membership can only modulate social judgements but not social abilities. These findings suggest a reconsideration of past findings that might have misrepresented the social judgements of autistic people through introducing an outgroup disadvantage, but also a need for tailored support for autistic social differences that emphasizes similarity and inclusion between diverse people.

4.1 Introduction

4.1.1 Intergroup Bias in Smile Differentiation in Autism

In Chapter 3, we developed a more evaluative implicit mentalizing task by implementing a prompt question. With this paradigm, we primarily aimed to address the last challenge of Southgate et al. (2007)'s paradigm we identified (see Chapter 1, Section 1.1.4.2) and to explore the role of evaluative context in mentalizing. The more evaluative paradigm proved to be better than the less evaluative version (without the prompt question, similar to the original Southgate et al. (2007)'s paradigm) in facilitating belief reasoning. This finding indicates that context is capable of modulating mentalizing as discussed in Chapter 1, which confirms Woo et al. (2023)'s proposal of the importance of contextual factors in the facilitation of mentalizing. In addition to the evaluability of context, other possible contextual modulators include intra-personal factors, such as group membership. Thus, the current chapter sets out to explore whether intergroup bias can modulate social cognition in autism.

The current study was initiated and carried out during the COVID-19 pandemic, it was impossible to use the in-person false-belief reasoning task applied in the previous two chapters because of the constrained situation, as mentioned in Chapter 1. Therefore, it was a pressing issue to find alternative ways to index mentalizing. We decided to use a smile discrimination task that can be easily adapted to use online as an alternative way to index mentalizing on the basis of two robust findings in the literature. First, emotional expressions do not always reflect one's genuine feelings and intentions in social interaction and communication (e.g., Ekman, 2003; Hess et al., 1997; Lazarus, 1991; Niedenthal et al., 2010; Rosenberg & Ekman, 2020). Second, people spontaneously evaluate the authenticity behind others' emotional expressions (Cosme et al., 2021). Accordingly, mentalizing has been suggested to play an important role in distinguishing between genuine and posed emotional expressions (Cosme et al., 2021; Lavan et al., 2017; McGettigan et al., 2015; Szameitat et al., 2010). On ethical grounds, smiles were specifically chosen because they are non-hazardous and potentially have positive effects on releasing stress resulting from the worldwide lockdown.

As reviewed in Chapter 1, genuine smiles are considered to be spontaneous and associated with enjoyment emotions, while posed smiles are not necessarily congruent with the genuine emotional experience but act as purposeful communication tools in social situations (Ekman, 2003; Krumhuber et al., 2007; Rosenberg & Ekman, 2020). The ability to accurately differentiate between, and respond to, them is an essential social ability to effectively cope with the complexity of social interactions, with difficulties causing poor social communication and functioning (e.g., Blampied et al., 2010; Boraston et al., 2008; Ekman, 2003; Lazarus, 1991; Song et al., 2016; Young et al., 2015). Difficulties recognising and responding to others' emotional states (ICD-11; World Health Organization, 2018) have long been suggested to be a central feature of autism (Hobson, 1986). To the best of our knowledge, only Boraston et al. (2008) and Blampied et al. (2010) have investigated smile discrimination ability in autistic populations (i.e., autistic adults and autistic boys). Both studies found that autistic people were less accurate in discriminating the two smile types than non-autistic people. Because of the limited results in the literature, replication studies are needed to confirm this finding and explore whether smile discrimination ability contributes to social communication difficulties in autism.

One potential factor that may modulate smile discrimination is intergroup bias, as reviewed in Chapter 1 (Section 1.2.2). Intergroup bias (or ingroup favouritism) refers to the tendency to judge ingroup members more positively than outgroup members (Tajfel, 1982). An outgroup advantage has been detected in smile discrimination. Young (2017) found that people were more accurate and faster in differentiating genuine from posed smiles displayed by outgroup than ingroup members using a minimal group setting where the group boundary was arbitrary and the group allocation was random. Because posed smiles from ingroup members were more likely to be misidentified as genuine ones, Young (2017) suggested that ingroup favouritism may have biased people to interpret ingroup smiles as genuine even when smiles were posed, whereas being wary of outgroup members may have led to a more vigilant approach. However, although people conveyed higher identification with their ingroup than outgroup members, this was not related to their performance in determining smile authenticity.

Intergroup bias seems to be a compelling factor that may potentially modulate smile discrimination in autistic people. However, to the best of our knowledge, no study has assessed an intergroup bias in smile authenticity judgments in autism; moreover, the behaviour of autistic people in intergroup settings has rarely been examined (Qian et al., 2022). A few recent studies have suggested that intergroup bias is attenuated and even absent in autistic people and non-autistic adults with higher autistic traits (e.g., Bertschy et al., 2020; Hadad et al., 2019; Kang et al., 2020; Qian et al., 2022; Uono et al., 2021; Vaucheret Paz et al., 2020). However, others reported that autistic adults and children showed the same intergroup bias as their non-autistic counterparts (e.g., Wilson et al., 2011; Yi et al., 2016; Yi et al., 2015). Furthermore, none of the aforementioned studies measured group identification, so the potential effect of the subjective attitude of autistic people on their feelings about group affiliation is not clear. The details of these studies have been described and reviewed in Chapter 1 (Section 1.2.4).

4.1.2 Empathy & Alexithymia

Another factor that may modulate emotion recognition and may also relate to intergroup bias is empathy. Empathy enables people to understand and share another's emotions and feelings (De Vignemont & Singer, 2006; Singer et al., 2004) and plays a crucial role in emotion recognition (Dyck et al., 2001) and intergroup relations (Dovidio et al., 2010; Vanman, 2016). Although this is hotly debated, autistic people have been widely reported to show empathic differences (Baron-Cohen & Wheelwright, 2004; Bird & Viding, 2014; Smith, 2009). Using an empathy-for-pain paradigm, Gu et al. (2015) measured skin conductance responses (SCR) and behavioural responses in judging others' pain in autistic adults. They found heightened bodily (implicit) but reduced behavioural (explicit) empathy for pain in autism. Specifically, autistic adults showed enhanced SCR but reduced behavioural discriminability related to empathetic pain compared to non-autistic adults. Accordingly, Gu et al. (2015) proposed that the behavioural empathic differences in autism may be driven by imprecise interoception, instead of a lack of empathy. Importantly, these empathic differences in autistic people are related not only to their difficulties in identifying emotional expressions (Dyck et al., 2001; Sucksmith et al., 2013) but also to their attenuated favouritism towards ingroup members (Qian et al., 2022; Vaucheret Paz et al., 2020). In this regard, autistic people may show less ingroup identification and less intergroup bias in smile discrimination.

However, it should not be overlooked that the prevalence of alexithymia, referring to difficulties in identifying and describing one's own emotions, is significantly higher in autism than in the general population (Hill et al., 2004; Salminen et al., 1999; Shah et al., 2016). Although characterised as involving difficulties identifying one's own emotions, Bird et al. (2010) documented that alexithymic traits, but not autism, can predict empathic brain responses in the left anterior insula linking alexithymia rather than autism to difficulties representing other's emotions. Indeed, co-occurring alexithymia in autism has been suggested to be directly responsible for emotion recognition difficulties (Cook et al., 2013) and attenuated intergroup bias (Komeda et al., 2019). Therefore, limited empathy observed in autism may be related to alexithymia instead of autism per se (Bird et al., 2010; Komeda et al., 2019).

4.1.3 The Current Study

The current study was designed to investigate the effect of intergroup bias in discriminating between genuine and posed smiles in autism using a minimal group paradigm and more ecologically valid smiling videos compared with still images in some previous studies (e.g., Boraston et al., 2008). The core hypothesis, according to previous literature, predicted that autistic adults would show an intergroup bias on smile discrimination, rating ingroup smiles as more genuine than outgroup smiles, as shown in Young (2017) with non-autistic adults and based on the intergroup bias literature in autism. Second, this intergroup bias would be attenuated in autistic adults compared to non-autistic adults. Third, we sought to replicate Boraston et al. (2008)'s findings that autistic adults would show less sensitivity to smile types, rating genuine smiles and posed smiles as more similar compared to their non-autistic counterparts. Fourth, both autistic and non-autistic adults would show greater discrimination between genuine and posed smiles for outgroup than ingroup smiles. Fifth, when evaluating the effectiveness of the minimal group paradigm, we expected that both non-autistic and autistic adults would be more likely to identify with ingroup than outgroup members, but autistic adults would possess an attenuated intergroup identification compared to non-autistic adults. Sixth, the current study is also interested in the relationship of empathy to group identification and intergroup smile discrimination; we predicted that higher degrees of empathy would be associated with higher ingroup identification and more genuine ingroup smile ratings. As alexithymia in addition to autism is also likely to modulate intergroup bias in smile perception, it is necessary to measure alexithymia, so it can be subsequently controlled for in analyses to reveal the actual role of autism in this process. Since alexithymia was considered a confounding variable in the current study and was not of interest, we had no specific hypothesis about it.

4.2 Method

4.2.1 Participants

Across the two diagnostic groups, 151 adults (85 females, 66 males) were recruited. The sample size was calculated to be 41 per group, assuming a medium effect size (f = 0.25), by referring to Young (2017)'s and Boraston et al. (2008)'s studies, and a power of .80. However, given the current study was conducted online, we decided to increase the sample size by a further 50% in each group, to mitigate the noise that may be introduced by the lack of control over participants' hardware, software and environment (Rodd, 2023). Prolific (www.prolific.co) was used for recruitment and Gorilla (www.gorilla.sc) for creating and delivering the experiment. Participants were required to be fluent in English and have normal or corrected-to-normal vision. The autism criterion in Prolific in addition to a questionnaire, including autism diagnosis, age of diagnosis, and family history, was used to identify autistic and non-autistic participants. Participants were over recruited to allow for participants who might need to be excluded and so that we might ensure a close match for age and non-verbal reasoning between the groups.

Fifteen participants from the entire sample were excluded prior to data analysis, who were inconsistent in reporting their diagnosis, or who self-identified as autistic without a diagnosis. The resulting two groups (75 autistic and 61 non-autistic) were comparable for age, sex, educational level and non-verbal reasoning as measured by the Matrix Reasoning Item Bank (MaRs-IB; Chierchia et al., 2019), but, as expected, were significantly different in autistic traits, alexithymia, and empathic concern (see Table 4.1). Additionally, although both groups were predominantly white, the proportion in the autism group was significantly higher than in the non-autistic group. Given that emotion recognition sensitivity is independent of verbal ability (Blampied et al., 2010; Hobson, 1986) and the minimal group induction and the smile discrimination task involved relatively simple questions that did not require participants to give a verbal response, only non-verbal reasoning was measured and matched between groups. None of the non-autistic participants reported a diagnosis of psychiatric or neurodevelopmental conditions. All participants in the autism group stated that they had a diagnosis from a qualified clinician with an average diagnostic age of 18.16 years (*SD* = 11.10), ranging from 3-49 years. This study was approved by the local Research Ethics Committee. All participants gave informed consent and were reimbursed for their time and effort.

Table 4.1. ***Autistic and non-autistic participants' characteristics; Mean (Standard Deviation).***

	Autism (n = 75)	Non-autism (n = 61)	Inferential statistic
Sex (M : F)	35 : 41	28 : 33	χ^2(1) = 0.14, *p* = .706, odds ratio = 0.88
Age	28.27 (9.15)	29.05 (8.45)	*t*(134) = -0.51, *p* = .610, *d* = -0.09
Ethnicity	Asian (1.3%) Black (6.7%) White (82.7%) Mixed (9.3%)	Asian (4.9%) Black (24.6%) White (62.3%) Mixed (8.2%)	χ^2(3) = 10.77, *p* = .013
Education	High school (52%) UG[e] (33.3%) PG[f] (13.3%) Missing (1.3%)	High school (42.6%) UG[e] (37.7%) PG[f] (19.7%)	χ2(2) = 1.63, *p* = .443
Non-verbal reasoning (MaRs-IB[a])	0.58 (0.19)	0.58 (0.17)	*t*(134) = 0.45, *p* = .964, *d* = 0.008
Autistic traits (AQ-10[b])	6.71 (2.40)	3.36 (1.82)	*t*(133.38) = 9.25, *p* < .001, *d* = 1.55
Alexithymia (TAS-20[c])	60.12 (10.96)	48.62 (12.68)	*t*(119.33) = 5.59, *p* < .001, *d* = 0.98
Empathy (IRI-EC[d])	18.25 (6.29)	20.67 (4.71)	*t*(133.14) = -2.56, *p* = .011, *d* = -0.43

Note. [a]MaRs-IB = Matrix Reasoning Item Bank; [b]AQ-10 = 10 item Autism-Spectrum Quotient; [c]TAS-20 = Toronto Alexithymia Scale; [d]IRI-EC = Interpersonal Reactivity Index (empathic concern subscale); [e]UG = undergraduate; [f]PG = postgraduate.

4.2.2 Procedure

Participants started the session by completing a dot-estimation task as an induction for setting minimal groups, then a smile discrimination task. This was followed by the MaRs-IB and finished with a series of questionnaires measuring: ingroup and outgroup identification; individual differences in autistic traits, alexithymia and empathic concern; and demographic information. Participants were then fully debriefed. The overall duration of the experiment was one hour.

4.2.3 Participatory Research

The current study was finalized after consultation with autistic community members. Six autistic adults (3 females, 3 males) were invited from Prolific to contribute to the design of the study before collecting data. They completed the key sections (i.e. the minimal group induction, the smile discrimination task and the group identification questionnaire) of the procedure and were fully informed of the research aim of the present study. Then, they were interviewed individually about their thoughts regarding the research direction and the study design. A 20-minute semi-structured interview was conducted focussing on: the importance of the study, the smoothness of the procedure, features that could be changed or improved, elements that were unclear or inappropriate or made them feel uncomfortable, and any topics that they would like to be further investigated in research.

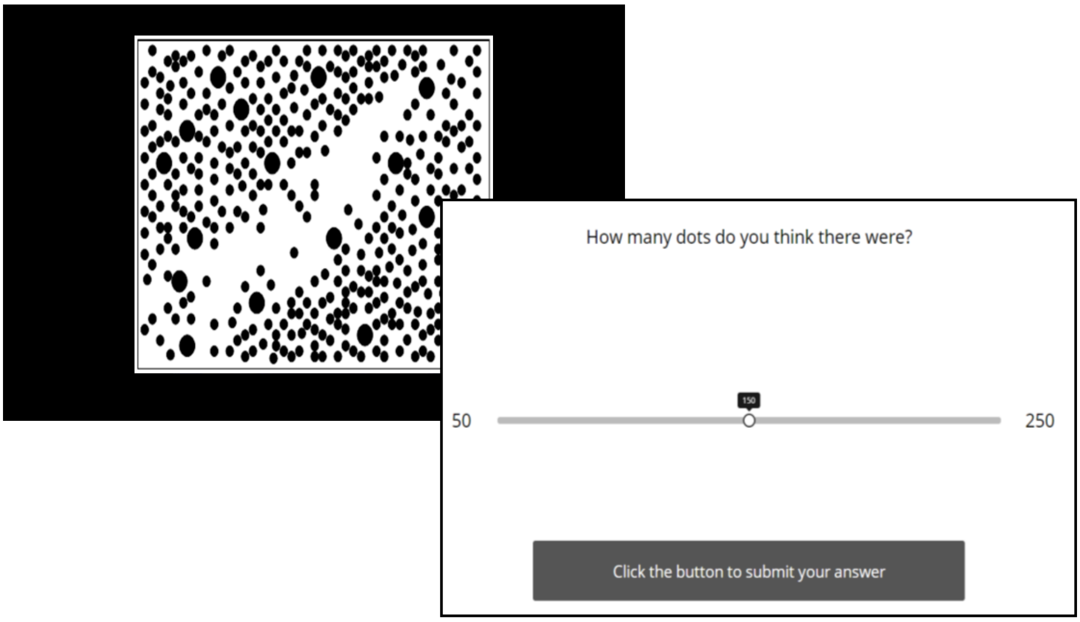
Four from this autistic group considered that it would be helpful to better understand facial emotion recognition in autism. All of them thought the instructions and procedure were clear and easy to follow. Two changes were made according to their responses. First, as all of them mentioned it was difficult for them to concentrate for such a long time and was tiring without an explicit break (even though they were free to take breaks anytime), the duration was reduced to one hour from 1.5 hours through refining the procedure and using brief versions of questionnaires, and four countdown breaks were added in the smile task. Second, one member did not remember their minimal group, so two questions were added to help with remembering group allocation, one after the group allocation to consolidate their memory, and the other after the smile task to check if the minimal group was remembered correctly. Noone suggested any topics for future research.

4.2.4 Minimal Group Induction

A dot-estimation task adapted from Howard and Rothbart (1980) was used, which served as a minimal group induction to randomly categorize participants into two groups: overestimators and underestimators. Participants were instructed that, according to previous studies, people tend to consistently overestimate or underestimate the number of objects they have seen, which also relates to their personality. They were also told they would later watch some videos of overestimators and underestimators, so it was important to remember their group.

Ten pictures each containing 50-250 dots were presented, each for 2000ms (see *Figure 4.1* for an illustration). Participants were asked to estimate the number of dots after each picture on a slider bar. After the ten trials, participants were told their scores were being calculated, and after a 2000ms delay they were informed that they were either an overestimator or an underestimator. To encourage participants to believe they were similar to their in-group members, they were told this was based on their estimation of the dots; however, the group allocation was fully randomized. Participants were given either yellow or green as an indicator of their group membership, which was reinforced by some positive personality traits of their ingroup members. The same colour badge would appear in each video later in the smile discrimination task, indicating the group membership of the smiler.

For counterbalance, approximately half of the autistic and non-autistic participants were assigned to each minimal group and therefore to each colour.



How many dots do you think there were?

Figure 4.1. **Illustration of the dot-estimation task.**

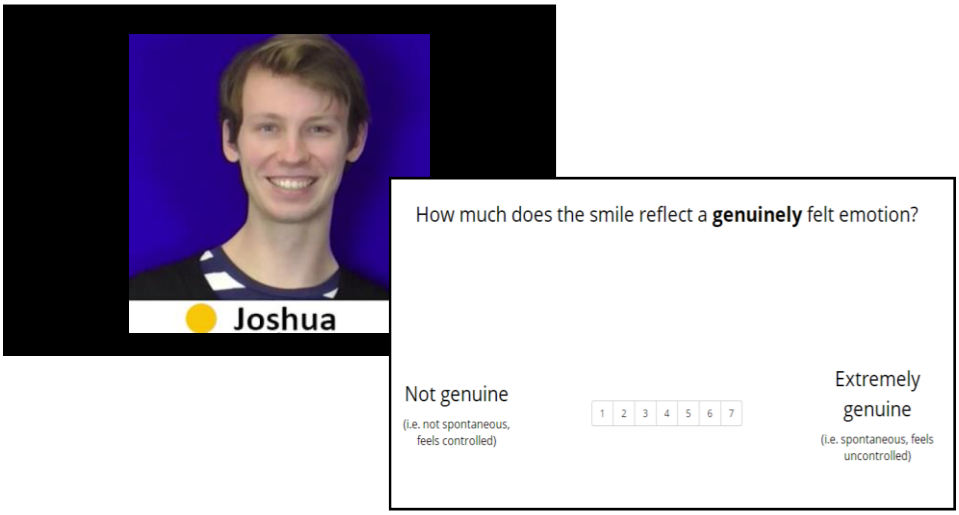
4.2.5 Smile Discrimination Task

Stimuli. The 20 colour videos used in Young (2017) were adopted, which have been validated to detect intergroup differences in smile discrimination (retrieved from https://www.bbc.co.uk/science/humanbody/mind/surveys/smiles/). Each smiler only presented one of two smile types (i.e., genuine or posed), 13 were males and 7 females, with a range of races (e.g. White, Black, Asian) and ages, and presumed to be non-autistic. To improve task reliability and sensitivity, the present study set out to increase the number of trials by employing a second set of 64 colour videos taken from Farmer et al. (2021). We intended to improve the signal-to-noise ratio and increase power, allowing for a better estimation of individual performance. This set of stimuli contained eight actors, half male and half female, all White young adults, and presumed to be non-autistic. Each smiler provided four genuine and four posed smiles. Therefore, the total number of videos was 84, half genuine and half posed. These videos were determined to be valid emotional expressions through previous studies (e.g., Young et al., 2015) and independent ratings (Farmer et al., 2021).

To match the two sets of stimuli, each clip was edited to the same size (i.e., 354px*360px) and length (i.e., 2000ms), to begin with a neutral facial expression and end with a fully expressed smile, using Adobe Premier Pro 2020. Each smiler was given either a yellow or green badge to indicate their group membership (overestimator or underestimator) as well as a name (e.g., Joshua), and both were placed along the bottom of each clip (see *Figure 4.2*). Half of the clips were randomly preselected to always be labelled as overestimators and the other half as underestimators. Colour (i.e., green vs. yellow) and minimal group type (i.e., overestimator vs. underestimator) were counterbalanced in both participants and smilers.

Setup. Participants were told they would watch a series of videos of underestimators' and overestimators' emotional facial expressions made in response to some funny things and would be required to make judgements of authenticity, contagion, valence and intensity after each recording, although only the authenticity ratings are analysed here. Participants responded on a 7-point Likert scale, with 1 = not genuine (i.e., not spontaneous, feels controlled) and 7 = extremely genuine (i.e., spontaneous, feels uncontrolled). Each trial began with a 500ms central fixation cross, with a 100ms blank screen before and after this. The video clip then played automatically only once, followed immediately by the authenticity question (see *Figure 4.2*). There was unlimited time for participants to make their judgements.

The two sets of videos were presented separately, split into two blocks. The first block contained two sub-blocks (10 trials each), and the second block contained four sub-blocks (16 trials each). Each sub-block only presented faces from one minimal group (half genuine half posed), and the group type was presented to participants at the beginning of the sub-block. Sub-blocks within each block and trials within each sub-block were randomly presented. According to the interview response (see section 2.2.1), four 15000ms countdown breaks were included before the second block and between its sub-blocks to prevent fatigue. Participants were asked about the group membership of the smiler after the first and second trials of each sub-block to check and help maintain their attention. They were also asked about their own group membership at the end of the entire task to verify whether they had correctly remembered their minimal group affiliation.


How much does the smile reflect a **genuinely** felt emotion?

Joshua
Not genuine
(i.e. not spontaneous, feels controlled)
1 2 3 4 5 6 7
Extremely genuine
(i.e. spontaneous, feels uncontrolled)

Figure 4.2. **Illustration of the smile discrimination task and an authenticity judgement.**

Analysis. Item-wise analysis was applied to analyse the main effects and interactions of autism diagnosis, group membership and smile type on the authenticity rating in the smile discrimination task in two mixed analyses of variance (ANOVAs). In the current study, it was assumed that the variance between different smilers was greater than the variance between different judges, so we took the average rating per video within autistic and non-autistic participants and for ingroup and outgroup smiles. Accordingly, we treated each video as an independent item, even though some smilers provided multiple videos. Diagnosis (autism vs. non-autism) and Group (ingroup vs. outgroup) were therefore treated as within-subject variables, while Smile type (genuine vs. posed) was a between-subjects variable.

4.2.6 Self-reported Measures

Following the smile discrimination task, group identification (GI) was measured by rating the applicability of eight statements (i.e., four ingroup and four outgroup) covering three areas (i.e., cognition, evaluation and affection) adapted from Doosje et al. (1995): (1) “I feel strong ties to overestimators [underestimators]”, (2) “I see myself as a member of the overestimator [underestimator] group”, (3) “I identify with the members of the overestimator [underestimator] group”, (4) “I am glad to be a member of the overestimator [underestimator] group”. The group type was highlighted with the corresponding colour. Each statement was rated on a 7-point Likert scale (1 = not at all, 7 = very true). The average GI score for ingroup and outgroup for each participant was calculated across the four questions.

Autistic traits were measured by the ten-item Autism-Spectrum Quotient (AQ-10; Allison et al., 2012), with higher scores indicating more autistic traits, ranging between 0-10. Empathic concern was measured by the empathic concern scale of the Interpersonal Reactivity Index (IRI-EC; Davis, 1980), with higher scores indicating a greater tendency to experience feelings of concern, compassion and warmth for others, ranging between 0-28; it should be noted that this measure does not directly assess the tendency to experience the feelings of others. Alexithymia was measured by the twenty-item Toronto Alexithymia Scale (TAS-20; Bagby et al., 1994), with higher scores indicating more difficulties identifying one's own emotions, ranging between 20-100. This was included in order to control for alexithymia, to test whether this condition could explain any group differences, given its high co-occurrence rate in autism. Demographic information was collected at the end of the experiment, including age, sex, education, ethnicity, autism diagnosis and age at diagnosis (if applicable).

4.3 Results

All the data were analysed using IBM SPSS Statistics (Version 29).

4.3.1 Smile Discrimination

Authenticity ratings. A 2x2x2 mixed-design ANOVA was conducted using the authenticity rating as the outcome variable, with Diagnosis (autism vs. non-autism) and Group (ingroup vs. outgroup) as within-subjects variables, and Smile type (genuine vs. posed) as a between-subjects factor. The results indicated main effects of Diagnosis, $F(1, 82) = 194.14$, $p < .001$, partial $\eta^2 = .703$, Group, $F(1, 82) = 26.11$, $p < .001$, partial $\eta^2 = .242$, and Smile type, $F(1, 82) = 124.82$, $p < .001$, partial $\eta^2 = .604$, and an interaction between Diagnosis and Smile type, $F(1, 82) = 36.75$, $p < .001$, partial $\eta^2 = .309$. Importantly, there was no interaction between Diagnosis and Group predicted by our second hypothesis, $F(1, 82) = 0.21$, $p = .646$, partial $\eta^2 = .003$, nor between Group and Smile Type predicted by our fourth hypothesis, $F(1, 82) = 0.07$, $p = .792$, partial $\eta^2 = .001$, nor 3 way interaction.

Specifically, non-autistic adults considered smiles overall as more genuine than autistic adults; smiles from ingroup members were rated as more genuine than those from outgroup members; and genuine smiles were rated as more genuine than posed smiles (see *Figure 4.3*).

To further investigate the interaction between Diagnosis and Smile type, ingroup and outgroup were collapsed, and then two paired samples *t*-tests were carried out, one for the autism group and one for the non-autism group, comparing the two smile types. Post-hoc tests revealed that both diagnostic groups rated genuine smiles as significantly more genuine than posed smiles: autism, *t*(82) = 9.88, *p* < .001, *d* = 2.16; non-autism, *t*(82) = 12.12, *p* < .001, *d* = 2.64. This is consistent with the main effect of Smile type. However, although the effect sizes are larger for the non-autistic participants, both are large effect sizes and there is not a robust statistical analysis to compare effect sizes. Hence, it is still unknown whether the two groups discriminate smiles differently.

Due to the format of the data in the item-wise analysis, there is no simple way to carry out an analysis to further explore the interaction between Diagnosis and Smile type that shows the diagnostic group difference in smile discrimination (i.e., difference between genuine and posed smiles). Thus, a conventional participant-wise analysis was applied to analyse the difference between autistic and non-autistic participants on the difference of authenticity rating between genuine and posed smiles in which group membership was collapsed. An independent samples *t*-test showed that the rating difference between genuine and posed smile is significantly smaller for autistic (*M* = 1.38, *SD* = 0.81) than non-autistic (*M* = 1.67, *SD* = 0.91) participants, *t*(134) = -2.00, *p* = .048, *d* = 0.86, which indicates that autistic people are to a lesser extent capable of discriminating genuine from posed smiles than non-autistic people.

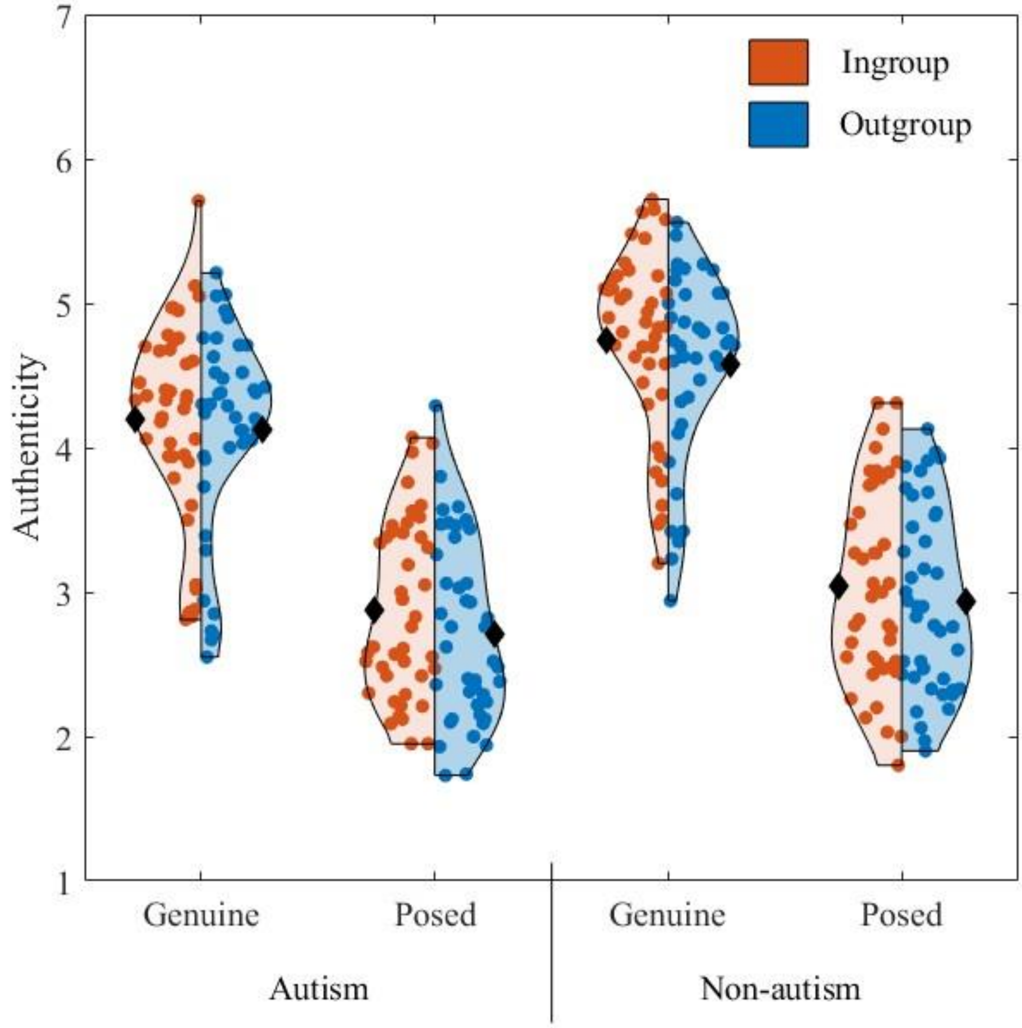


Figure 4.3. **The smile authenticity ratings by Diagnosis, Group and Smile type (each dot represents the mean rating of each smile); black diamonds represent the mean of each condition.**

Adjusted authenticity ratings. Given the greater prevalence of alexithymia in autism and the group difference on the TAS-20 in the current sample (see section 2.1), it is possible that the observed effect of diagnosis, to some extent, is driven by alexithymia rather than autism (see Section 1.3). To control for alexithymia characteristics, a simple linear regression model was used to predict the authenticity ratings of each smile item across all participants based on the TAS-20 scores, then we calculated the difference between the observed value of the authenticity rating and the value of the rating predicted from the regression line (i.e., the standardized residual). The residual, or the adjusted authenticity rating, was entered in the same analysis as the unadjusted rating, a 2x2x2 mixed-design ANOVA. There were main effects of Diagnosis, $F(1, 82) = 155.64$, $p < .001$, partial $\eta^2 = .655$, Group, $F(1, 82) = 25.53$, $p < .001$, partial $\eta^2 = .237$, and Smile type, $F(1, 82) = 10.96$, $p = .001$, partial $\eta^2 = .118$, and an interaction between Diagnosis and Smile type, $F(1, 82) = 36.07$, $p < .001$, partial $\eta^2 = .306$, but no other significant interactions (see Table 4.2 for descriptive statistics). The results therefore remain the same after controlling for alexithymia, which indicates that alexithymia cannot explain the main variance observed in the smile discrimination task.

Table 4.2. ***Standardised residuals of smile authenticity rating after controlling for alexithymia, Mean (Standard Deviation).***

		Ingroup	Outgroup
Autism	Genuine	-0.114 (0.15)	-0.161 (0.10)
	Posed	-0.014 (0.13)	-0.116 (0.10)
Non-autism	Genuine	0.190 (0.11)	0.086 (0.13)
	Posed	0.065 (0.14)	-0.003 (0.09)

4.3.2 Group Identification

Group identification scores were analysed using a 2x2 mixed-design ANOVA, with Group (ingroup vs. outgroup) as a within-subjects variable, and Diagnosis (autism vs. non-autism) as a between-subjects factor. There were significant main effects of Diagnosis, $F(1, 134) = 16.45$, $p < .001$, partial $\eta^2 = .109$, and Group, $F(1, 134) = 162.66$, $p < .001$, partial $\eta^2 = .548$, and an interaction between Diagnosis and Group, $F(1, 134) = 5.12$, $p = .025$, partial $\eta^2 = .037$. Specifically, non-autistic participants were more likely to identify with others than autistic participants; and participants identified more strongly with ingroup members than outgroup members. Post-hoc tests (with Bonferroni correction for multiple comparisons, α-level adjusted to $p = .025$) indicated that non-autistic adults reported greater group identification with their ingroup members than autistic adults, $t(134) = -4.07$, $p < .001$, $d = -0.70$, but no group difference was observed in outgroup identification, $t(134) = -1.66$, $p = .099$, $d = -0.29$ (see *Figure 4.4*).

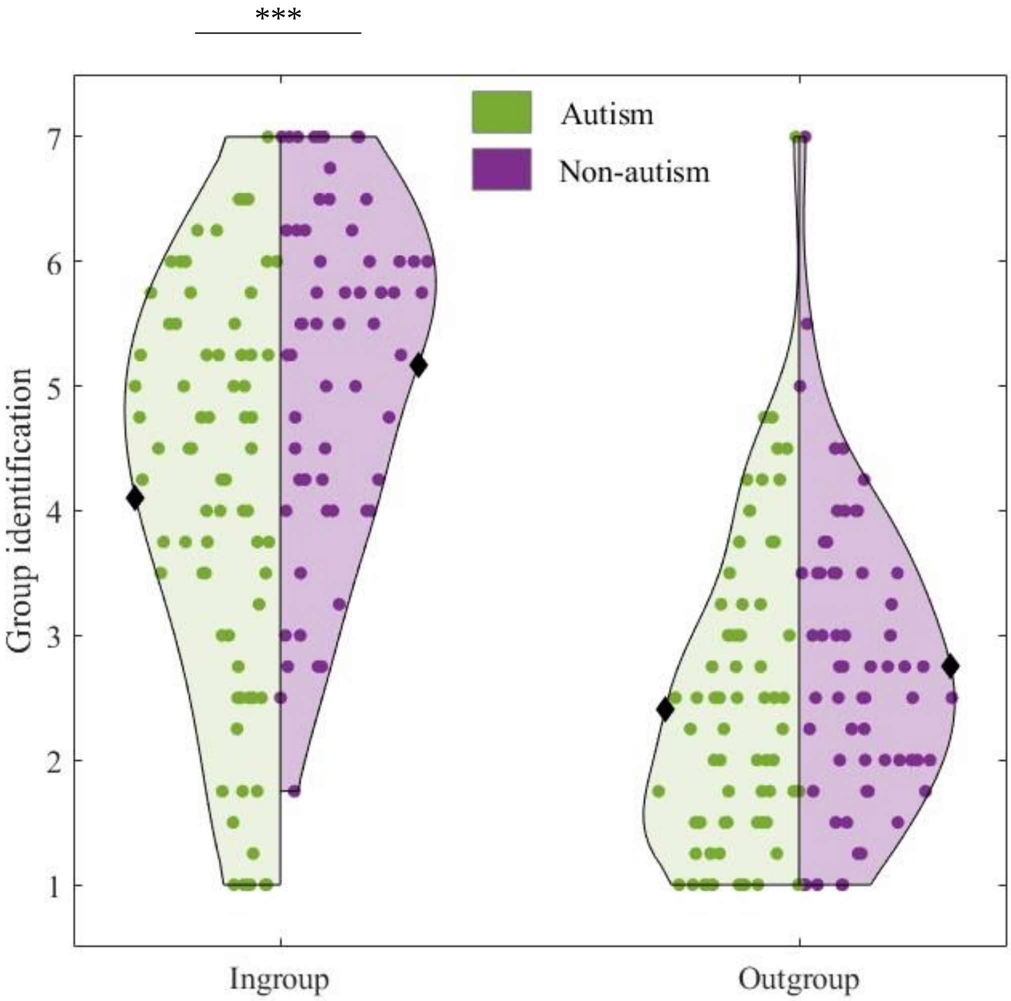


Figure 4.4. **Diagnosis x Group interaction on group identification scores (each dot represents the score of each participant); black diamonds represent the mean of each condition.**

4.3.3 Correlations

In both autistic and non-autistic adults, correlations were conducted to determine whether empathic concern was associated with intergroup identification and intergroup bias in smile discrimination, and whether smile discrimination ability contributes to social communication difficulties. As this was exploring individual differences, we recalculated the smile judgement ratings by taking the average rating per participant for ingroup and outgroup smiles. In autistic adults, empathic concern was positively correlated with authenticity ratings of ingroup smiles, *r* = .28, *p* = .015, and outgroup smiles, *r* = .27, *p* = .017, but not with the group identification scores, ingroup *r* = .05, *p* = .670, outgroup, *r* = .05, *p* = .677; and autistic traits were negatively correlated with authenticity ratings of posed smiles, *r* = -.27, *p* = .018, but not genuine smiles, *r* = -.14, *p* = .238. In contrast, in non-autistic adults, empathic concern was correlated with the ingroup identification, *r* = .44, *p* < .001, but not with the outgroup identification, *r* = -.152, *p* = .243, nor with ingroup smiles *r* = .13, *p* = .322, nor outgroup smiles, *r* = .13, *p* = .338; and autistic traits were not correlated with authenticity ratings of genuine, *r* = -.10, *p* = .468, nor posed smiles, *r* =.18, *p* = .177. Additionally, the difference between ingroup and outgroup identification was not associated with the difference between ingroup and outgroup smile judgements in either autistic, *r* = .14, *p* = .223, or non-autistic people, *r* = -.21, *p* = .111.

4.4 Discussion

To the best of our knowledge, this is the first study to investigate whether an intergroup bias can modulate the perception of genuine and posed smile authenticity among autistic adults. We found that group membership did affect authenticity judgements similarly in autistic and non-autistic adults, but did not modulate the ability to differentiate genuine from posed smiles in either diagnostic group.

4.4.1 Intergroup Bias & Group Identification

As expected, ingroup favouritism on smile authenticity identification was found not only in non-autistic adults, replicating Young (2017)'s findings, but also in autistic adults. Specifically, ingroup smiles were rated as more genuine than outgroup smiles in our minimal group setting. This indicates that intergroup bias can indeed influence how autistic people perceive smile authenticity. Furthermore, autistic people seemed to be as susceptible as non-autistic people to this intergroup bias, because no interaction was observed between Diagnosis and Group. Accordingly, autistic adults' sensitivity to intergroup bias on smile judgements seems not to be attenuated. Considering results in the recent literature, this is consistent with some autism studies on the cross-race effect (Wilson et al., 2011; Yi et al., 2016; Yi et al., 2015). It is possible that these latter studies involved tasks that autistic people struggled more with, such as judging social norms (Qian et al., 2022) and direct gaze aversion (Uono et al., 2021), making intergroup modulation harder to detect, whilst we were using a task that autistic were capable of, albeit to a lesser extent than non-autistic people (see below). Regardless, the current results indicate that ingroup members might be perceived as more authentic and therefore interaction with them might be more rewarding (Shore & Heerey, 2011) and enjoyable (Krumhuber et al., 2007) for both autistic and non-autistic people.

We also observed that autistic adults generally rated smiles as less authentic than non-autistic adults. It is possible that autistic adults are generally less trusting of unfamiliar people, given their increased likelihood to have experienced victimisation (Sterzing et al., 2012), and therefore judge all smiles to be less genuine. Relatedly, we found that those autistic adults who gave lower ratings of smile authenticity also reported lower empathic concern, but it is not possible from our data to know whether or how this might relate to reduced trust. Alternatively, given autistic adults were susceptible to a minimal social group manipulation, they could presumably also be influenced by pre-existing social groups with whom they are more likely to share similar cognitive styles, such as autism vs non-autism groupings. If the diagnostic-group identification also causes intergroup effects, and if our autistic adults assumed in the absence of evidence to the contrary that all the videos contained people from the non-autistic majority, this could account for the generally lower ratings made by the autism group. Indeed, the idea of diagnostic-ingroup favouritism has been partially supported by Sasson et al. (2017) and Alkhaldi et al. (2019), who both reported that non-autistic people rated autistic people less favourably than other non-autistic people, without knowing who was autistic. If a diagnostic intergroup bias does account for the generally lower ratings given by our autistic participants, we might have failed to fairly measure how autistic adults judge smile authenticity. This could also be said for the many studies in the literature assessing social judgements in autism, which presumably used non-autistic protagonists (Gernsbacher et al., 2017). This might suggest a need to re-evaluate past findings of social perception in autism and consider whether any of those studies might have misrepresented the social judgements of autistic people through introducing an outgroup disadvantage. Future studies could test this possibility directly by including autistic as well as non-autistic protagonists.

Certainly, when we asked participants how closely they identified with each minimal group at the end of the experiment, although both diagnostic groups reported higher group identification towards their ingroup than outgroup, autistic adults reported identifying less with the actors than non-autistic adults, and this was especially the case for ingroup members. This attenuated self-reported group identification may indicate that autistic people are less likely to have a sense of belonging and internal safety and security provided by ingroup identification. This means they might miss the opportunities of supports, benefits and resources from ingroup members (e.g., Balliet et al., 2014; Qian et al., 2022). In contrast, they may be less likely to discriminate or be prejudiced against and even dehumanize outgroup members, which could potentially reduce intergroup conflict in society (e.g., Balliet et al., 2014; Leyens et al., 2007; MacLachlan, 2020). Thus, maybe autistic people are inherently more inclusive of diversity. Whatever the cause, this reduction in identification is likely to have been related to the lower authenticity ratings given by autistic adults. In addition, the group identification difference between ingroup and outgroup was not related to the smile genuineness rating difference in both diagnostic groups, consistent with Young (2017). Although intergroup bias can modulate smile judgement, it might work differently to the feeling of closeness with ingroup members. People might not feel they are close to a group, but they might still treat ingroup and outgroup differently, and vice versa.

4.4.2 Smile Discrimination

Consistent with Boraston et al. (2008) and Blampied et al. (2010)'s findings but here using dynamic stimuli, we found autistic adults rated genuine smiles and posed smiles as more similar compared to non-autistic adults. This indicates that while autistic people are capable of discriminating genuine from posed smiles, this is to a lesser extent than non-autistic adults. Importantly, the results remained the same after controlling for alexithymia, so smile authenticity judgements must rely on autism-specific cognitive processes. Autistic adults may be less sure of the authenticity of others, perhaps due to differences in reasoning about mental states (Boraston et al., 2008), which could subsequently affect their social communication.

As well as differing in reliance on mentalizing, genuine and posed smile judgements rely on attention to different parts of the face. Only genuine smiles involve muscle contraction around the eyes, especially the AU6 (Duchenne & de Boulogne, 1990; Ekman et al., 1990; Ekman & Friesen, 1982), so a lack of attention to the eye region during smile judgements could explain the reduction in smile discrimination in autism (Boraston et al., 2008). Future studies using eye-tracking techniques could reveal the fixation pattern and attention distribution of autistic people when judging smiles, to explore the information they tend to use from smiling faces; it would also be of interest to understand the use of these muscles in autistic smile production. This might give a deeper insight into the mechanisms underlying subtle facial expression recognition differences in autism.

An alternative seemingly plausible interpretation of this reduction in smile discrimination in autistic adults comes from a neurodiversity perspective. Following from the 'double empathy problem' (Milton, 2012) which hypothesises that autistic social interaction and communication difficulties are bidirectional. That is, if autistic people struggle to navigate in a non-autistic society, it should be equally difficult for non-autistic people to fit into an autistic society. In reality, more than 97% of the general population is non-autistic (e.g., Brugha et al., 2009; Chown, 2014; Li et al., 2022), so the sense of being disfavoured by the non-autistic majority is likely to be harmful (Milton, 2012; Mitchell et al., 2021), which could discourage autistic people from interacting with others (Mitchell et al., 2019). It has been suggested that autistic people can more easily decode social cues and reason about the mental states of other autistic people than about non-autistic people, and the opposite would be true for non-autistic people (e.g., Fletcher-Watson & Happé, 2019; Komeda et al., 2019; Sheppard et al., 2016). Indeed, there is evidence showed that non-autistic people are less successful in understanding autistic than non-autistic targets' behaviours (Sheppard et al., 2016). However, Young (2017)'s study of smile discrimination would indicate that we should expect increased smile discrimination for outgroup members. Having said this, Young (2017) failed to replicate this increased outgroup smile discrimination effect in his second experiment, as did we in ours – there was no interaction between smile type and intergroup membership, nor a 3-way interaction with diagnostic group, indicating that an intergroup bias did not modulate smile discrimination ability. It therefore seems unlikely that a diagnostic intergroup bias could explain the diagnostic group difference in smile discrimination ability.

4.4.3 Empathy

Regarding the relationship of empathy to intergroup identification and intergroup smile authenticity judgment, higher empathic concern in autistic adults was related to more genuine rating on all smile types no matter which group they belonged to but not to group identification, while higher empathic concern in non-autistic adults was only associated with a greater tendency to identify with their ingroup. We replicated the modulation effect of empathy on emotion recognition in autism (Dyck et al., 2001; Sucksmith et al., 2013) and that on group identification in non-autism (Dovidio et al., 2010; Vanman, 2016) but not vice versa. That is, higher empathy in autistic people can generally increase the perceived authenticity of smiles, consistent with Dyck et al. (2001), but not enhance their ingroup identification. On the other hand, more empathic non-autistic adults are more likely to identify with ingroup members (Dovidio et al., 2010; Vanman, 2016), but they seem to not rely on empathy to differentiate genuine from posed smiles, so other factors instead of empathic concern play a more central role in smile discrimination for them.

4.4.4 Advantages & Limitations of This Study

More generally, our use of a minimal group paradigm to generate intergroup bias meant we were able to minimize the potential effects of other forms of intergroup bias and elucidate that even arbitrary labels can induce ingroup favouritism in both autistic and non-autistic people, quite apart from groupings that are associated with social stigmatism (Milton, 2012). Further, conducting the study online was advantageous during the COVID-19 pandemic (Tsantani et al., 2022; Türközer & Öngür, 2020) and for the inclusion of autistic people who might not be have been able to participate in laboratory experiments. Our findings hold promise that it is feasible and valid to assess smile perception and more generally implement minimal group paradigms online. In fact, it is possible that minimal group paradigms might have stronger effects online than in lab-based studies, as there is little other contextual information to guide them online and hence the assigned membership would be more prominent than in lab-based environments.

However, we are also aware that an online approach also has limitations – less control over the environment, the monitor and the integrity of participants during testing (Tsantani et al., 2022). Similarly, because of ethics and feasibility, we could not verify participants' diagnoses, although our autistic adults showed significantly higher autistic traits than the non-autism group, so we believe that our findings are a valuable addition to current autism research. Additionally, all of our autistic adults possessed average-to-high non-verbal reasoning ability – future studies should confirm these results in a laboratory setting and recruit autistic people with diverse cognitive abilities.

Of course there are large individual differences in genuine and posed smile expressions, which are usually interpreted in more ambiguous and varied social interaction contexts (Heerey, 2014). Thus, our findings may require evaluation under more naturalistic settings. However, given the videos of genuine and posed smiles produced by actors were differentiable even in a remote online situation, their fundamental differences could be more salient and therefore more likely to be identified in face-to-face interaction. Thus, we believe that our findings are useful for understanding subtle expression discrimination under intergroup settings in autism.

4.5 Conclusion

In conclusion, the current study contributes to a better understanding of autism through demonstrating autistic sensitivity to social group categories despite a tendency to judge all smiles as less genuine and difficulties in differentiating subtle facial emotion expressions under minimal group settings. We propose that this might be due to reduced identification with, empathy for or trust in unfamiliar or diagnostic outgroup members, in combination with mentalizing or social attention differences. As autistic people perceive ingroup members to be more authentic, this is likely to give rise to more rewarding and more comfortable interactions. This has implications for designing tailored support and policies that emphasize similarities and inclusion between autistic and non-autistic people to avoid intergroup conflicts (Mitchell et al., 2021), rather than focusing on how they might be different (Baron-Cohen, 2017). This might facilitate autistic people in navigating the social world more effectively and make society more inclusive.

Chapter 5. Neural and Facial Mechanisms of Intergroup Bias and ToM in Smile Discrimination: A fNIRS Study

Abstract

Intergroup bias has been found to modulate genuine and posed smile discrimination. Facial mimicry has been suggested to facilitate social cognition. However, as using behavioural measurements only may not fully capture the mechanisms involved in intergroup bias in social cognition, the neuroimaging method would be helpful to unpack the neural correlates underpinning this process. Thus, the current study aimed to investigate the neural and facial mechanisms of intergroup bias in smile discrimination. Thirty-three adults viewed videos of people making genuine or posed smiles and were informed (falsely) of the group membership of the actors. The ability to differentiate genuine and posed smiles (ToM) and male and female actors (non-ToM) of in-group and out-group members and group identification were assessed, and participants' facial expressions were recorded. Interestingly, although the behavioural results did not reveal evidence of intergroup bias, I found that the medial frontal gyrus and dorsolateral prefrontal cortex were more activated during ingroup smiles, while the inferior frontal gyrus was more activated during outgroup smiles. Additionally, ToM conditions were harder than non-ToM conditions, indicated by lower accuracy and longer reaction time. No evidence was found for mimicry at both the behavioural and neural levels. These findings extend the current understanding of the neural mechanisms underpinning intergroup bias in social cognition and have implications for understanding the complexity of the human brain in response to multiple higher-order cognitive modulations.

5.1 Introduction

In Chapter 4, by using a minimal group paradigm, I demonstrated that although intergroup bias might not be able to improve the social cognitive ability to distinguish between genuine and posed smiles, it can modulate people's perception of others, which allows the smiles of ingroup members to be judged as more genuine in both non-autistic and autistic adults. I discussed some potential mechanisms that may be engaged during intergroup bias in social cognition, including mentalizing, attention, empathy, familiarity and group identification. In this chapter, I focus on the neural correlates of intergroup bias in social cognition and how these relate to behavioural performance.

Additionally, as smile authenticity is considered a spectrum rather than dichotomized, I used a 7-point Likert scale to measure the smile identification ability in Chapter 4, unlike Young (2017)'s binary question. One possibility for the absence of intergroup bias modulation effect in social ability is that the 7-point Likert scale used in Chapter 4 may have a higher tolerance for classification error than the binary question, as a modest rating difference between genuine and posed smiles could be statistically significant and therefore be considered capable of smile discrimination.

5.1.1 Neural Mechanisms of Social Cognition in Relation to Intergroup Bias

As using behavioural measurements only cannot fully capture the mechanisms involved in intergroup processes, a growing body of neuroscience research has begun to unpack the neural correlates underpinning intergroup bias on a wide range of perceptions, attitudes and behaviours. Substantial research evidence has been found that people perceive and respond differently to ingroup and outgroup information at the neural level (reviewed in Molenberghs & Louis, 2018; Moradi et al., 2020). In particular, both the literature and

Chapter 4 have found that intergroup bias can modulate mentalizing by using smile discrimination tasks. Thus, there should be somewhere in the brain that group membership interacts with mentalizing. However, little is known about which regions of the brain are involved in this interaction, and how. Moreover, although some neuroimaging studies have looked at prototypical emotion identification in relation to group membership, the smile task might be a more ecologically valid and interesting way to tap into mentalizing and intergroup bias. This section is going to discuss some functional brain systems that have been suggested to be involved in identifying and evaluating others' emotional and/or mental states in intergroup settings.

As reviewed in Chapter 1 (Section 1.2.2), the brain functional systems that are associated with Theory of Mind (ToM; or mentalizing) and executive control (in particular attentional control) have been suggested to be involved in social cognition in intergroup settings (e.g., Moradi et al., 2020; Mullen et al., 1992; Schupp et al., 2003). The core nodes of the mentalizing network involve the *temporoparietal junction (TPJ), posterior superior temporal sulcus (pSTS)* and *medial prefrontal cortex (mPFC)* (Frith & Frith, 2006; Frith & Frith, 2003; Schurz et al., 2014). The main regions of the executive control network contain but are not limited to the *dorsolateral prefrontal cortex (dlPFC)* and *middle frontal gyrus (MFG)* (e.g., Decety & Lamm, 2007; Eberhardt, 2005; Moradi et al., 2020; Mullen et al., 1992; Schupp et al., 2003; Seeley et al., 2007; Smith et al., 2019).

Existing literature has provided significant insights into the underlying neural representations of intergroup bias during emotion processing. For example, Lin et al. (2018) examined intergroup social influence on prototypical emotion processing. Participants' group-free baseline rating of images showing people engaged in emotional contexts was compared with the second stage rating where the same images were shown along with how their ingroup and outgroup members rated them using functional magnetic resonance imaging (fMRI). They found a tendency for people to assimilate their emotional states more with ingroup than outgroup members. Paralleled with these behavioural findings, they also observed greater activation when aligning with the ingroup over the outgroup in the *medial prefrontal cortex (mPFC), pSTS, lateral prefrontal cortex (lPFC), temporal pole (TP), ventromedial prefrontal cortex (vmPFC), ventral striatum (VS), amygdala,* and *insula*, whereas no regions showed more activation during the opposite alignment. Lin et al. (2018) claimed that intergroup social influence on emotion processing engages mentalizing, executive function, reward processing, and salience detection.

Besides emotion processing, intergroup bias has also been observed in altering mental state decoding. Adams Jr et al. (2010) looked at racial intergroup bias in decoding emotions from only the eye region by using a mentalizing task, the Reading the Mind in the Eyes Task (Baron-Cohen, Wheelwright, Hill, et al., 2001) using fMRI. They found both Asian and Caucasian participants were more accurate in detecting mental emotional states from their ingroup, and that greater activation in the *TPJ* area associated with mentalizing when decoding the emotion of ingroup than outgroup members.

Mentalizing seems to not only be influenced by intergroup bias but also modulate intergroup bias. As mentioned in Chapter 1 (Section 1.2.1), Park and Young (2020) showed that people updated fewer impressions of ingroup members, compared with outgroup members, when both engaged in the same negative behaviour. This tendency was associated with having closer relationships in their social life and with a reduction in *TPJ* activity in response to ingroup members' negative behaviour. They concluded that ingroup favouritism can bias people to make and update impressions more positively on ingroup than outgroup members, which may potentially be beneficial to maintaining relationships with their ingroup members. Likely, the tendency to discount harmful mental states (e.g., intentions) of ingroup members seems to be achieved through selectively “turning down” the mentalizing network or biasing mentalizing outcomes, as indexed by *TPJ* activity. Given the *TPJ* has been associated with updating social impressions based on the detected gap between the estimation of a target’s mental states and the observed reality (Koster-Hale & Saxe, 2013; Thornton & Mitchell, 2018), the failure to recruit the mentalizing network may underlie the failure to update negative impression of ingroup members (Hughes, Ambady, et al., 2017; Kliemann et al., 2008).

Another example of racial intergroup bias was found in non-verbal behaviour reasoning by Katsumi and Dolcos (2018). Participants were presented with non-verbal social encounters between a racial ingroup or outgroup guest-host character, including approach and avoidance, and were asked to rate the host’s competence and their own interest in engaging in follow-up interaction with the hosts. The authors found that participants rated ingroup more positively than outgroup members. Although the main effect of intergroup bias was not observed in brain activity, they found the *mPFC* and *pSTS* showed greater activation when observing ingroup than outgroup approach behaviour. It is not surprising that the mentalizing network was recruited in a social cognitive context where mentalizing was necessary or plausible. Notably, the *pSTS* and *TPJ* areas are also associated with bottom-up processes, reorienting attention to salient stimuli (Decety & Lamm, 2007). Accordingly, the greater activation in the *pSTS* during observing ingroup social interaction may reflect not only greater mentalizing engagement but also attentional modulation processing (Katsumi & Dolcos, 2018). Indeed, it has been suggested that intergroup bias is potentially underpinned by variations in attentional saliency (e.g., Moradi et al., 2020; Mullen et al., 1992; Schupp et al., 2003).

As reviewed in Chapter 1 (Section 1.2.2), the executive control network has also been found to be involved in social cognition during intergroup settings. For example, Katsumi and Dolcos (2018) compared social to non-social control scenes where one of the agents was replaced with an object. They found that brain regions involved in visual perception of human action and social cognition were more engaged in the social than non-social conditions, including the *extrastriate visual cortex (EVC)*, *pSTS* and *dlPFC*. As the *dlPFC* has been primarily related to the voluntary regulation of racial intergroup bias based on the literature (e.g., Bartholow & Henry, 2010), Katsumi and Dolcos (2018) concluded that the attenuated *dlPFC* activation during the ingroup non-social condition may indicate a reduction in executive control and regulatory processes. This context-based mechanism seems to be in line with the idea investigated in Chapter 3 that social evaluative context facilitates mentalizing proposed by Woo et al. (2023). Specifically, when social interaction is absent, the non-social context provides observers less reason to monitor and regulate the information.

The *dlPFC* activity has been linked to several cognitive functions related to intergroup bias in the literature (Zhang et al., 2023), such as attentional control (e.g., Seeley et al., 2007; Smith et al., 2019), cost-benefit analysis (e.g., Hosokawa et al., 2013), updating beliefs about others’ risk preferences and adapting to them (e.g., Suzuki et al., 2016), impulse control and inhibition (e.g., Shackman et al., 2009; Steinbeis et al., 2012), complying with social norms to facilitate cooperation (e.g., Spitzer et al., 2007; Stallen et al., 2018), and perceiving social dominance hierarchies (e.g., Qu et al., 2017). Hence, it might be not straightforward to map the activation pattern of the *dlPFC* to a single or a set of cognitive functions in intergroup settings. Moreover, differential activation in the *dlPFC* when processing and evaluating information from ingroup and outgroup members has been identified in various social contexts (e.g., Amodio, 2014; Eberhardt, 2005; Rilling et al., 2008; Zhang et al., 2023). For example, increased *dlPFC* activity has been observed when processing stimuli from ingroup versus outgroup members, which has been linked to more top-down monitoring and regulation of predisposed intergroup bias (reviewed in Amodio, 2014).

In a hyperscanning study using functional near-infrared spectroscopy (fNIRS), Zhang et al. (2023) concurrently measured leader and follower neural responses in the *dlPFC* during intergroup conflicts. They found that more leader contribution led to greater group survival during ingroup defence than outgroup attacks, which was linked to their increased *dlPFC* activity. Meanwhile, more synchronization of *dlPFC* activity between leader and follower was associated with higher cooperation between them when leaders organized an ingroup defence; however, their behaviour and neural activity aligned poorly when launching an outgroup attack.

Notably, although neural intergroup bias typically manifests as greater activation for processing ingroup than outgroup information (e.g., Katsumi & Dolcos, 2018), there are occasions when outgroup information is prioritized (Moradi et al., 2020). For example, Rilling et al. (2008) measured brain activation when participants interacted with an ingroup versus an outgroup partner using a minimal group paradigm. They found that people who reported that they felt no difference when interacting with both partners showed higher *dlPFC* activation during outgroup than ingroup interactions. It has been suggested that stronger *dlPFC* activation when observing racial outgroup faces may indicate a top-down neural correlate of cognitive exertion (reviewed in Eberhardt, 2005; Moradi et al., 2020). Thus, it may be possible to actively modulate the saliency of outgroup members and make more cognitive effort to process outgroup information, which could potentially prevent the predisposed intergroup bias tendency. Rilling et al. (2008) explained that their findings could indicate that more executive control efforts are needed to overcome negative biases against outgroup members.

Taking together all the reviewed studies in the current section, there seems no single brain region or network consistently responsible for intergroup biases in social cognition. Neural intergroup biases are more likely to manifest as modulations of functional neural networks that are widely involved in social-emotional and -cognitive processes (reviewed in Molenberghs, 2013). Potentially, the entire brain could be involved in intergroup biases, but with specific neural networks and patterns based on the implicated group boundary, input modality, and outcome bias (reviewed in Molenberghs, 2013; Molenberghs & Louis, 2018). A combination of multiple group boundaries may result in a stronger intergroup bias (Molenberghs & Louis, 2018), but is also possible to reduce or even offset the outcome bias. Thus, I would like to extend the findings in Chapter 4 to capture the neural correlates of intergroup bias in social cognition in the current chapter.

5.1.2 Recognition and Mimicry of Facial Expressions

Recognising facial emotions is challenging and some theories suggest mimicry may contribute to this (e.g., Lee et al., 2023; Niedenthal et al., 2010; Wang & Hamilton, 2012). As introduced and discussed in Chapter 1 (Section 1.2.3), human mimicry is intrinsic and often unconscious in social encounters (reviewed in Chartrand & Van Baaren, 2009). Simulation theories have been proposed to explain the function of mimicry in recognising and understanding facial expressions (e.g., Gallese, 2007, 2009; Niedenthal et al., 2010). In particular, Niedenthal et al. (2010) proposed the simulation of smiles (SIMS) model and suggested that facial mimicry may be especially beneficial for judging subtle or ambiguous emotional expressions, such as the authenticity of smiles. The SIMS model suggests that the meaning of a smile is uncertain which can be positive (e.g., genuine smiles) or negative (e.g., dominance smiles – associated with feelings of superiority and pride for maintaining social status). To understand a smile, it needs to be first noticed through eye contact, which results in a rewarding or negative affect and motor mimicry, then this bodily experience leads to the interpretation of the target's intentions or feelings. Empirical evidence in favour has reported that the degree of mimickers' smile muscles (i.e., AU6 & AU12) contraction predicted their smile authenticity judgments of the mimicked targets (Korb et al., 2014), and disrupting mimicry impairs smile discrimination ability (Rychlowska et al., 2014), see Chapter 1 (Section 1.2.3) for more details.

As discussed in Chapter 1 (Section 1.2.3), the discovery of the mirror neuron system (MNS) has primarily supported simulation theories, which have been suggested to be essential for social mimicry (e.g., Bastiaansen et al., 2009; Heyes, 2011; Krautheim et al., 2019; McLellan et al., 2012; Olsson & Ochsner, 2008; Shamay-Tsoory, 2011; Spunt & Lieberman, 2012; Wang & Hamilton, 2012). The MNS is putatively located in the *inferior frontal gyrus (IFG), superior temporal sulcus (STS), and inferior parietal cortex* (Iacoboni et al., 1999; Rizzolatti & Craighero, 2004). These specialized neurons fire both when the same action is acted and observed (Rizzolatti, 2005; Rizzolatti & Craighero, 2004; Rizzolatti & Sinigaglia, 2016). The activation of the MNS is sensitive to the authenticity of expressions. For example, McLellan et al. (2012; n = 7) found greater neural activity in response to genuine compared to posed facial expressions in the *IFG*, Lee et al. (2023; n = 44) found that the *IFG* was activated when accurately identifying genuine from posed smiles (see Chapter 1 for more details).

As mentioned in Chapter 1, both social mimicry and the MNS can be modulated by social contextual factors, for example, intergroup bias. Ingroup members are more likely to be mimicked than outgroup members (e.g., Bourgeois & Hess, 2008; Mondillon et al., 2007;

Peng et al., 2020; Peng et al., 2021). Likewise, enhanced neural activity in the MNS has been found for perceiving ingroup compared to outgroup members' facial expressions (e.g., Krautheim et al., 2019).

However, several studies did not find mediation effects of facial mimicry on emotion processing, such as smile authenticity judgments using avatar faces (Korb et al., 2014), smile function recognition (i.e., reward, affiliative, and dominance; Orlowska et al., 2018), and the accuracy of prototypical emotion recognition (Blairy et al., 1999). Potentially explaining these findings, visual theories raise an obvious objection to simulation theories, as discussed in Chapter 1 (e.g., Allison et al., 2000; Kanwisher, 2000). Visual theories propose that people understand an action by visually analysing each element that consists of it and the interactions between elements. Thus, action understanding may not require simulation. According to visual theories, the *STS* and *EVC* are primarily associated with action understanding (Allison et al., 2000; Kanwisher et al., 1997). Visual theories also propose that the MNS reflects action understanding per se, instead of contributing to action understanding (Csibra, 2008; Hickok, 2013).

Taken together, there is evidence supporting either simulation theories or visual theories. Thus, it is necessary to look at whether mimicry plays a role in social cognition and compare simulation theories and visual theories at both the behavioural and neural levels. Accordingly, the current chapter also examines whether facial imitation can facilitate smile discrimination and be modulated by intergroup bias by measuring both the facial movements and the neural activation in the MNS.

5.1.3 fNIRS and ROIs

In order to measure mimicry, participants should be able to make facial movements freely, thus I adopted a multimodal approach. Accordingly, a neuroimaging technique was also required that was feasible for multimodal measurements and had a good tolerance of motion artefacts. Thus, functional near-infrared spectroscopy (fNIRS) was chosen to investigate the neural and facial mechanisms of intergroup bias in smile discrimination, which has been widely used in understanding the neural mechanisms of social cognition (reviewed in Pan et al., 2019; Pinti et al., 2019; Pinti et al., 2020). fNIRS is a non-invasive optical neuroimaging technique that records the hemodynamic response (reviewed in Scholkmann et al., 2014). This is achieved by shining NIR light into the head, with wavelengths in the range of 700-1000 nm, from the source, and then the attenuation rate of the NIR light can be measured at the detector after passing through the local cortex. Both sources and detectors are located on the cap put on participants' heads. The amount of NIR light that comes back tells us something about the blood flow in the brain that is associated with the target neural activation.

fNIRS has a number of advantages compared with other neuroimaging modalities, such as fMRI and electroencephalogram (EEG). One of the major advantages of fNIRS lies in its good tolerance for motion artefacts, as mentioned earlier. Once the cap is well-positioned, participants are allowed to move freely and fNIRS systems can still provide a good signal (reviewed in Herold et al., 2017). Second, fNIRS is silent, with no safety concerns, at a low cost, and can be portable, which not only allows for a large range of possible tasks in various contexts but also becomes an ideal choice when participants' safety and comfort is prioritized (reviewed in Pan et al., 2019; Pinti et al., 2020). Thus, fNIRS has also been documented to be feasible to work with a wide range of populations, including infants, the elderly, and people with clinical conditions (e.g., Lloyd-Fox et al., 2014; Obrig, 2014; Pu et al., 2008; Zhang & Roeyers, 2019; Zhang et al., 2023). Third, unlike functional magnetic resonance imaging (fMRI), fNIRS measures the changes in concentration of both oxygenated hemoglobin (HbO_2) and deoxygenated hemoglobin (HbR) (reviewed in Liu et al., 2015; Pinti et al., 2019; Pinti et al., 2020), and the combination of the two signals using the correlation-based signal improvement (CBSI) method allows more accurate estimation of functional neural activation (Cui et al., 2010).

Fourth, fNIRS possesses a relatively good balance between temporal and spatial resolutions. fNIRS has sampling rates up to 100 Hz, typically between 1-10 Hz which results in a better temporal resolution than fMRI (1-3 Hz) (Pinti et al., 2020; Quaresima & Ferrari, 2019). However, fNIRS can only measure activity at the local cortical surface at a spatial resolution of 2-3 cm (Pinti et al., 2020). Although it is superior to EEG (5-9 cm), it is impossible to access the subcortical or deeper regions. Therefore, I decided to focus on the cortical areas that are potentially associated with intergroup bias in judging the emotional mental states of others, but not those relevant subcortical regions (e.g., amygdala, insula; see Section 5.1.1). Additionally, as the *medial prefrontal cortex (mPFC)* is folded in the great longitudinal fissure from which it is potentially difficult to get effective signals through fNIRS, despite its relevance to the current study (see Section 5.1.1), I decided not to attempt to measure its activity in the current study, to give more coverage to the other relevant brain regions.

For this study, 8 brain regions of interest (ROIs) on the cortical surface were carefully identified that are likely to be associated with the effect of intergroup bias and ToM on smile discrimination (see *Figure 5.1*): the *TPJ, postcentral gyrus (PCG),* contains the *primary somatosensory cortex), EVC, middle frontal gyrus (MFG), IFG, temporal sulcus (TS),* *superior temporal cortex (STC), dlPFC*. These ROIs are strongly linked to four functional systems: the mentalizing system, MNS, attentional-saliency network, and executive control network, as discussed in Sections 5.1.1-5.1.2.

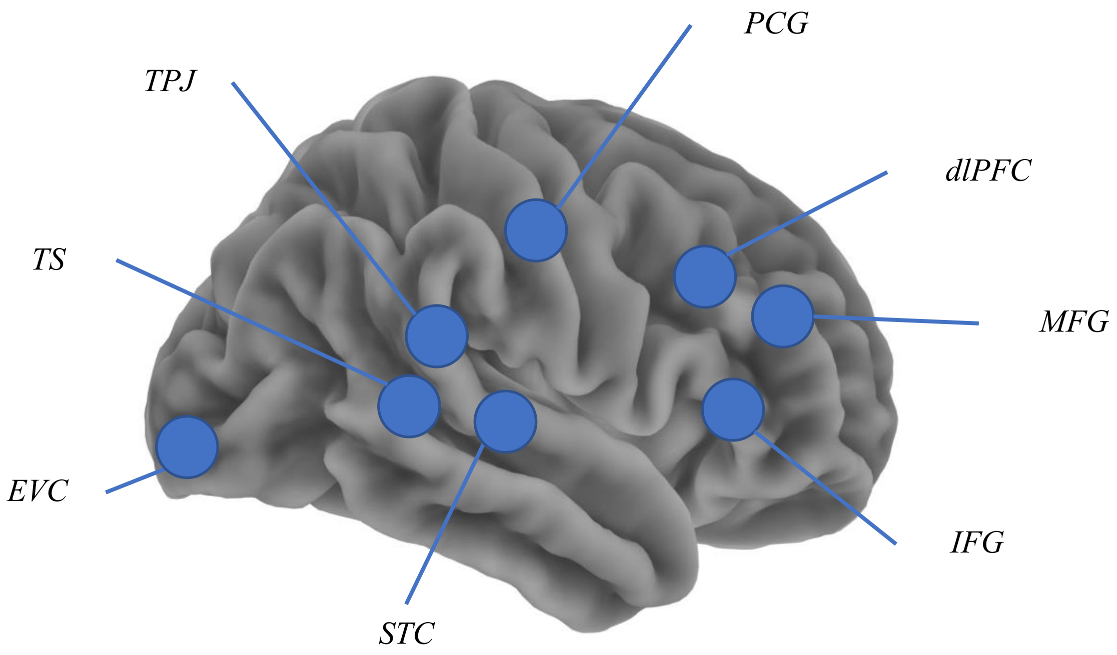


***Figure 5.1*. Illustration of brain regions of interest (ROIs): *temporoparietal junction (TPJ), postcentral gyrus (PCG), extrastriate visual cortex (EVC), middle frontal gyrus (MFG), inferior frontal gyrus (IFG), temporal sulcus (TS), superior temporal cortex (STC), dorsolateral prefrontal cortex (dlPFC)*.**

5.1.4 The Current Study

The current study first intended to extend the findings in Chapter 4 about ingroup favouritism in smile judgements to capture the underlying neural representations of intergroup bias and mental state decoding during a smile discrimination task under a minimal group setting using fNIRS. Based on the paradigm used in Chapter 4, I added a non-ToM control condition, asking the smilers' gender, to be able to identify the specific neural mechanism of ToM in smile identification. A 2 x 2 factorial design was adopted where the factors are Group (ingroup vs. outgroup) and ToM (ToM vs. non-ToM), see Section 5.2.4 for more details. According to the literature, an enhanced activation broadly in the identified ROIs to ingroup compared with outgroup smiles would be predicted, while the ToM condition would increase activity in the mentalizing system in comparison with the Gender condition. Meanwhile, if intergroup bias is absent at the behavioural level, stronger activation in the *dlPFC* for outgroup than ingroup members would be expected, based on Rilling et al. (2008), if cognitive control is used to overcome intergroup discriminatory tendencies.

Second, I was also interested in examining whether facial imitation would facilitate smile discrimination and be modulated by intergroup bias, through recording participants' facial behaviours and their neural activation in the MNS. The use of fNIRS gave us the flexibility to capture facial action while simultaneously recording neural activity. According to simulation theories, it would be hypothesized that more facial mimicry and greater activation in the MNS would be observed in the ToM than Gender conditions that may facilitate subtle facial emotional recognition; as well as when observing the ingroup over outgroup smilers, indicating people's tendency for spontaneously mimicking ingroup members. However, no difference in facial movement and neural activation in the MNS based on visual theories between the ToM and Gender conditions would be observed, as behavioural mimicry is not necessary for social cognition in these models.

Last but not least, given Chapter 4 did not replicate the modulation effect of intergroup bias in social abilities, the current study aimed to conduct a closer replication that directly measured smile discrimination in conjunction with measuring brain activity. So, another difference I made to the previous paradigm was to change the response approach from a 7-point Likert scale to Young (2017)'s binary response model. As discussed in Section 5.1, there was no specific prediction for the intergroup bias effect, as each direction was plausible. Regarding the ToM effect, given mental state reasoning could be more sophisticated than judging actors' gender, the ToM condition would be less accurate but involve more time to response than the non-ToM (Gender) control condition.

5.2 Method

5.2.1 Participants

Thirty-four participants (17 females, 22 Asian) were recruited through a local participant database, and advertisements placed around the local community. Participants were required to be fluent in English, have normal or corrected-to-normal vision, and range in age from 18 to 35 years. A questionnaire, including autism diagnosis, age of diagnosis, and family history, was used to identify autistic and non-autistic participants, the former of which would be excluded in the current study. None of the participants reported a diagnosis of psychiatric or neurodevelopmental conditions (see Table 5.1). One participant from the recruited sample was excluded from the analysis, whose accuracy was 3 standard deviations away from the group mean in the non-ToM control conditions, presumably indicating a lack of attention. The demographics of the resulting sample (*n* = 33) are reported in Table 5.1.

Individual differences in autistic traits, alexithymia, and empathic concern were measured (see Table 5.1). Specifically, autistic traits were measured by the Autism-Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, et al., 2001), with higher scores indicating more autistic traits, ranging between 0-50. Alexithymia was measured by the twenty-item Toronto Alexithymia Scale (TAS-20; Bagby et al., 1994), with higher scores indicating more alexithymic traits, ranging between 20 and 100. Empathy was measured by the empathic concern scale of the Interpersonal Reactivity Index (IRI-EC; Davis, 1980), with higher scores indicating a greater tendency to experience feelings of concern, compassion and warmth for others, ranging between 0-28. This study was approved by the UCL Research Ethics Committee. Data collection took place during a time of COVID-19 restrictions, and all methods were performed in accordance with the approved guidelines and regulations. Written informed consent was obtained from all participants. All participants were reimbursed for their time and effort.

Table 5.1. ***Participants' demographics; Mean (Standard Deviation).***

	Participants (n = 33)
Sex (M : F)	16 : 17
Age	26.15 (3.76)[d]
Handedness	Right (84.8%), Left (9.1%), Both (3.0%), Missing (3.0%)
Ethnicity	Asian (66.7%), White (24.2%), Mixed (9.1%)
Education	High school (18.2%), UG[e] (30.3%), PG[f] (51.5%)
Autism diagnosis	No (100%), Yes (0%)
Family history of autism	No (81.8%), Yes (9.1%), Not sure (9.1%)
Autistic traits (AQ[a])	18.81 (8.70)
Alexithymia (TAS-20[b])	49.48 (11.75)
Empathic concern (IRI-EC[c])	20.03 (5.42)

Note. [a]AQ = Autism-Spectrum Quotient; [b]TAS-20 = Toronto Alexithymia Scale; [c]IRI-EC = Interpersonal Reactivity Index (empathic concern subscale); [d]Age for two participants was missing who should be within the range of 18-35; [e]UG = undergraduate; [f]PG = postgraduate.

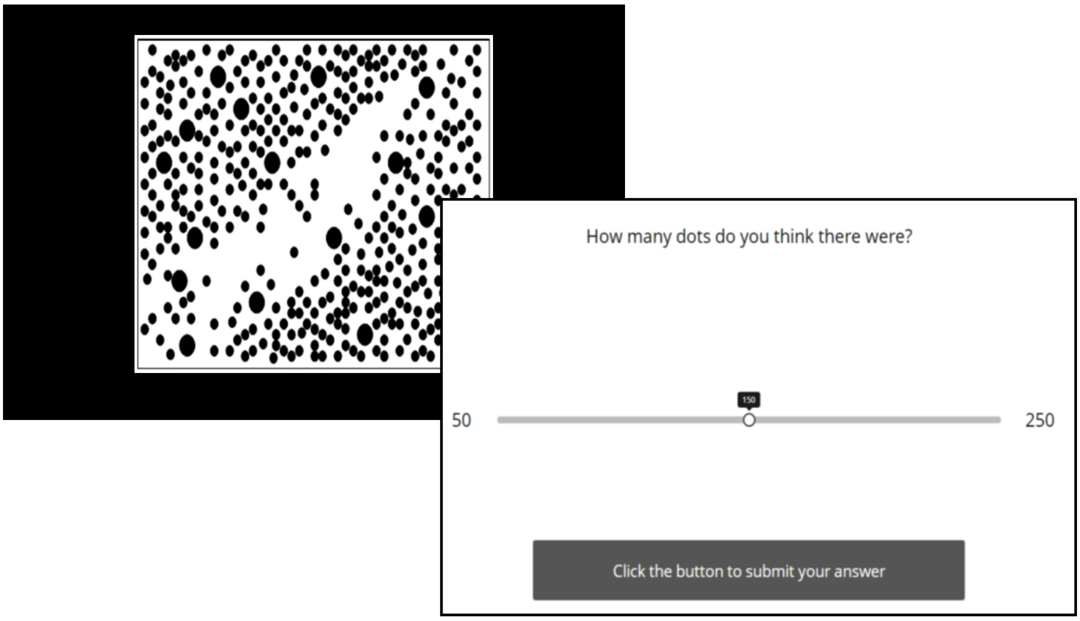
5.2.2 Procedure

Participants started the session by completing a dot-estimation task as an induction for setting minimal groups and four practice trials for each condition of the smile discrimination task, created and delivered through Gorilla (www.gorilla.sc). This was followed by the smile discrimination task, written and delivered using the Psychophysics Toolbox extensions (Brainard & Vision, 1997; Kleiner et al., 2007; Pelli & Vision, 1997), in MATLAB (Mathworks, Natick, MA), during the NIRS phase. Smile videos were presented via a Dell 27-inch monitor. Participants sat approximately 70cm from the screen and were instructed to sit still throughout the assessment to reduce motion artefacts in the neural signals. Then, the session finished with a series of questionnaires measuring: ingroup and outgroup identification; individual differences in autistic traits, alexithymia, and empathic concern; and demographic information. Participants were then fully debriefed. The overall duration of the experiment was two hours. Testing was conducted at the Institute of Cognitive Neuroscience, University College London.

5.2.3 Minimal Group Induction & Group Identification

As in Chapter 4, a dot-estimation task adapted from Howard and Rothbart (1980) was used, which served as a minimal group induction to randomly categorize participants into two groups: overestimators and underestimators. Participants were instructed that, according to previous studies, people tend to consistently overestimate or underestimate the number of objects they have seen, which also relates to their personality. They were also told they would later watch some videos of overestimators and underestimators, so it was important to remember their group.

Ten pictures each containing 50-250 dots were presented, each for 2000ms (see *Figure 5.2* for an illustration). Participants were asked to estimate the number of dots after each picture on a slider bar, ranging from 50-250. After the ten trials, participants were told their scores were being calculated, and after a 2000ms delay, they were informed that they were either an overestimator or an underestimator. To encourage participants to believe they were similar to their in-group members, they were told this was based on their estimation of the dots; however, the group allocation was fully randomized. Participants were given either a yellow or green sticker to wear in a visible place as an indicator of their group membership to encourage them to better affiliate with their minimal group, which was further reinforced by some positive personality traits of their ingroup members. Later in the smile discrimination task, a yellow or green badge would appear below each video, indicating the group membership of the smiler. The colour (i.e., yellow, green) and group type (i.e., underestimator, overestimator) allocated to participants and the colour (i.e., yellow, green) assigned to smile videos were all counterbalanced.



How many dots do you think there were?

***Figure 5.2.* Illustration of the dot-estimation task.**

Group identification (GI) was measured to assess the validity of the minimal group induction by rating the applicability of eight statements (i.e., four ingroup and four outgroup) covering three areas (i.e., cognition, evaluation and affection) adapted from Doosje et al. (1995): (1) “I feel strong ties to overestimators [underestimators]”, (2) “I see myself as a member of the overestimator [underestimator] group”, (3) “I identify with the members of the overestimator [underestimator] group”, (4) “I am glad to be a member of the overestimator [underestimator] group”. The group type was highlighted with the corresponding colour. Each statement was rated on a 7-point Likert scale (1 = not at all, 7 = very true). The average

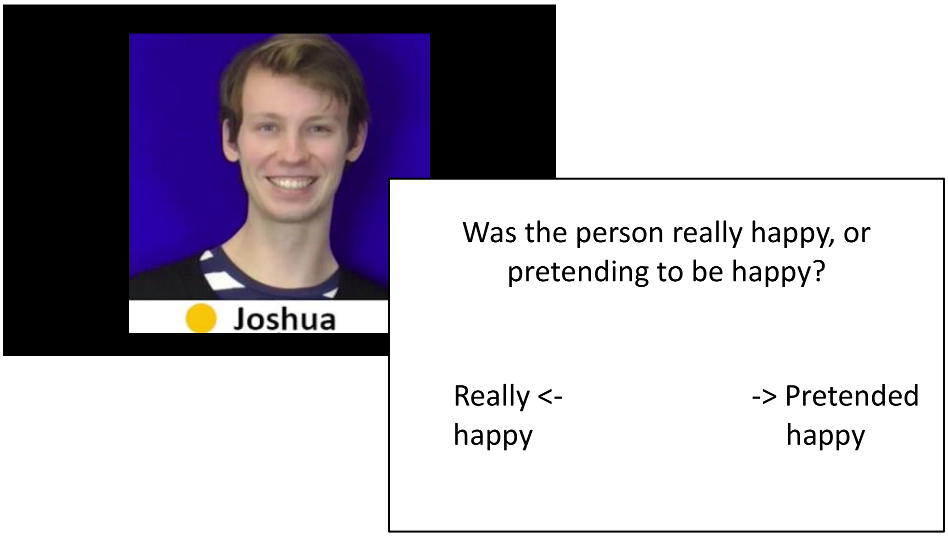
GI score for ingroup and outgroup for each participant was calculated across the corresponding four questions.

5.2.4 Smile Discrimination Task

Two sets of colour videos were adopted, which had been validated to detect intergroup differences in smile judgements in Chapter 4. Each of the 20 smilers in Young (2017) presented either a genuine or posed smile, including 14 males and 6 females, with a range of ethnicity (e.g. White, Black, Asian) and age (retrieved from https://www.bbc.co.uk/science/humanbody/mind/surveys/smiles/). The 64 colour videos taken from Farmer et al. (2021) contained eight actors, half male and half female, all White young adults, and each provided four genuine and four posed smiles. To improve task reliability and sensitivity for better detection of the task-related brain mechanism, the current study set out to further increase the number of trials by presenting each actor’s smile of each type four times, including two presentations of the each video and two presentations of the mirrored version of each video. Therefore, the total number of trials was 144, half genuine and half posed, portraying 56 female and 88 male smiles. These videos were determined to be valid emotional expressions through previous studies (e.g. Young et al., 2015) and independent ratings in Farmer et al. (2021).

Each video clip was edited to the same size (i.e., 354px*360px) and length (i.e., 2000ms), to begin with a neutral facial expression and end with a fully expressed genuine or posed smile, using Adobe Premier Pro 2020 and Shotcut. Each smiler was given either a yellow or green badge to indicate their group membership (overestimator or underestimator) as well as a name (e.g. Joshua), and both were placed along the bottom of each clip (see *Figure 5.3*). Half of the clips were randomly preselected to be labelled as overestimators and the other half as underestimators. Colour (i.e., green vs. yellow) and minimal group type (i.e., overestimator vs. underestimator) were counterbalanced in both participants and smilers.

Participants were instructed that they would watch some emotional facial expressions of underestimators and overestimators responding to funny things and indicate whether the person was really happy or pretending to be happy (ToM question) or whether the person was a male or female (Gender control question). Thus, there were four conditions: Ingroup-ToM, Ingroup-Gender, Outgroup-ToM, and Outgroup-Gender. Four additional smile videos, one for each condition, selected from Pexels (retrieved from https://www.pexels.com/), were used in the practice session to familiarize participants with the task before moving to the NIRS phase. The 144 videos were split into 32 blocks, 8 blocks per condition. There was a 10000ms inter-block interval between every two consecutive blocks with a fixation cross presented in the middle of the screen. Each block contained four Farmer et al. (2021) trials or five BBC trials, half genuine, half posed smiles and with a similar gender ratio. At the beginning of each block, an information page (4000ms) would inform participants of the question (i.e. ToM or Gender) that was going to be asked and of the group membership (i.e. overestimator or underestimator) of the smilers in the following block, which was highlighted by the corresponding group colour. In each trial, the video clip was played automatically only once, followed immediately by the allocated question of the current block (see *Figure 5.3*). Questions were answered via keystroke, with the left arrow key mapping onto ‘really happy’ or ‘male’ and the right arrow key mapping onto ‘pretend happy’ or ‘female’. Participants had up to 4000ms to answer the question, so they were instructed to respond as quickly as they could. There was a 2500ms inter-trial interval between every two consecutive trials with a fixation cross on the screen.



Was the person really happy, or pretending to be happy?

Really <- happy

-> Pretended happy

Figure 5.3. **Illustration of the smile discrimination task and an authenticity judgement.**

Accuracy of each condition (i.e., Ingroup-ToM, Ingroup-Gender, Outgroup-ToM, and Outgroup-Gender) was calculated by dividing the number of accurate trials (e.g., calling a genuine smile ‘really happy’, calling a female ‘female’) by the total trial number in that condition. Accuracy ranged from 0-1, with higher values indicating better accuracy. In each condition, reaction time (RT) was calculated by averaging the RT of all the accurate trials, for use in the behavioural analysis, and general RT was calculated by averaging the RT of all the trials (including accurate and inaccurate trails, but not missing trials) to use with the neuroimaging data. To investigate response bias, base rates of responding (i.e., the proportion of the time participants selected “genuine”, regardless of accuracy) were also calculated in Ingroup-ToM and Outgroup-ToM conditions respectively.

5.2.5 Face recording

During the smile discrimination task, in addition to accuracy and RT, brain oxygenation and haemodynamic signals, screen recording, facial behaviour, eye movement and physiological data were also recorded; the last two are not analysed in this chapter however, according to the aims of the current study. The Open Broadcaster Software (OBS) Studio, software for video recording and live streaming, was used to record the screen during the entire smile detection task. A digital clock was displayed in the bottom right corner of the screen in order to record the actual time of the key events to accurately synchronize all the measurements together, including neural, psychophysiological, neurocognitive, and behavioural measures in post-processing.

The Windows 10 Camera Application and a connected Logitech C920 HD Pro Webcam (30fps, 1920x1080) was used to track participants' facial behaviours. The recorded video was further processed with OpenFace (Baltrušaitis et al., 2015; Baltrušaitis et al., 2013; Baltrušaitis et al., 2018; Zadeh et al., 2017). The OpenFace algorithms on facial action unit (AU) recognition are based on the Facial Action Coding System (FACS; Ekman & Friesen, 1976) that deconstructs human facial movements in the tonus of specific facial muscles labelled as the corresponding AUs and taxonomizes facial behaviour accordingly. A subset of 18 facial AUs can be recognized by OpenFace, including muscles around brows, eyes, nose, cheeks, lip, jaw and chin (i.e. AU01, AU02, AU04, AU05, AU06, AU07, AU09, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26, AU28, AU45). This provided information about the presence and intensity of activity in each of these AUs for each frame of the recorded video. The intensity of all the AUs was averaged and the mean for each participant in each condition was calculated for both behavioural analysis and the analysis of the neural signals, ranging from 0 to 5, with higher values indicating higher intensities.

Gaze data was recorded by the Tobii Eye Tracker 5, a screen-based eye-tracker system, with a sampling rate at 60Hz (Tobii, Sweden). The Tobii Platform Development Kit (PDK) integrated with the eye-tracker, provided access to the eye tracking software. A 3-

point calibration was performed before starting the task. The eq02+ LifeMonitor, a wearable monitor that provides physiological data, was used to continuously monitor participants' ECG data (i.e. heart rate and breathing rate) at 256 Hz (Equivital, UK; https://www.equivital.com/heart-rate-and-breathing-rate-monitor). The Equivital eqManager software was used to manage, extract and transform the recorded data.

5.2.6 Neural Signal Acquisition

5.2.6.1 Signal Acquisition

Brain oxygenation and haemodynamic signals were recorded using a continuous-wave functional near-infrared spectroscopy (fNIRS) system (LABNIRS, Shimadzu Corp., Kyoto, Japan). fNIRS data acquisition used 54 optodes, including 26 sources and 28 detectors, arranged in an alternated configuration, creating 80 measurement channels as shown in *Figure 5.4* (i.e. source-detector pair; length of 3 cm) measuring the haemodynamic changes in the cerebral cortex, and 2 short channels (length of 1.5 cm) measuring the haemodynamic changes in the local scalp and skull that would not be analysed in the current study. Each light source emitted light at three wavelengths (780, 805 and 830 nm), and raw intensity signals of their reflectance were recorded by the corresponding detector at a sampling frequency of 23.81 Hz. Triggers were added through MATLAB to mark the beginning of each block in the fNIRS data, which ensured accurate identification of each experimental block and alignment with the time parameters of the behavioural datasets.

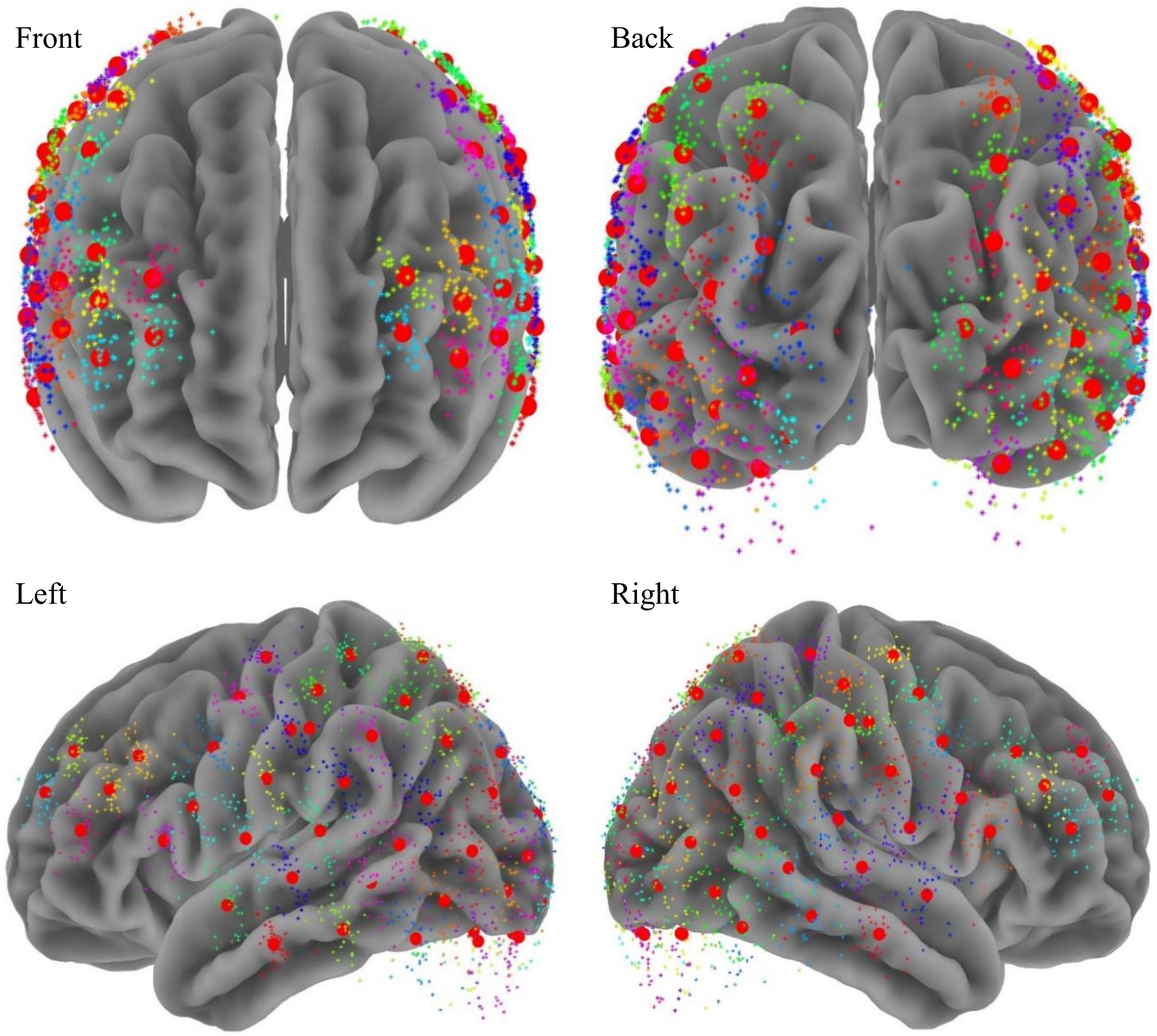


Figure 5.4. **NIRS channels configuration. Mean locations of channel centroids (big red dots) across all participants and channel localizations for each participant (small colourful dots) are represented on the front and back sides and the right and left hemispheres of a single rendered brain. Each nominal channel is assigned to one colour.**

Participants were fitted with a cap embedded with optode holders. The cap and optodes were carefully placed in the same way for each participant, and rubber bands were used to adjust the cap size to accommodate individual differences in head size. In order for the optodes to properly touch participants’ scalps to maximize the transmission of light through their scalp, a lighted fibre-optic ear scoop was used to move hair away from underneath the optode before placing it inside the holder. These operations ensured that the fNIRS signals were of good quality. Prior to the start of recording, initial measurements were conducted to check the light intensity for each channel. The signal quality was then optimized accordingly by appropriately adjusting the detector gains by the system automatically or manually replacing the optodes or moving any hair blocking the light away to maximize the optical coupling between the optodes and the scalp. The fNIRS recording proceeded when each detector was assured to detect sufficient reflected light from the paired source (i.e. between 60-150 db) or when each channel signal reached its maximum limit that could not be improved any more.

5.2.6.2 Pre-processing

The raw fNIRS data were pre-processed using the HomER2 toolbox (Huppert et al., 2009). The pre-processing pipeline steps I adopted followed the standardization of fNIRS analysis procedures developed by Pinti et al. (2019). Specifically, raw intensity data from all channels were first visually inspected to assess the signal quality. Channels with detector saturation, substantial motion artefacts or poor optical coupling shown as an absence of the heartbeat oscillation (frequency at 1-1.5 Hz) in the signal's power spectrogram were excluded from further analyses. The number of channels with good signal quality that were included in the following analyses are reported in Table 5.2. Then, raw intensity signals were converted into changes in optical density using the *hmrInteensity2OD* function. Motion artefacts were corrected according to the wavelet-based method (see Molavi & Dumont, 2012) using the *hmrMotionCorrectWavelet* function (*iqr* = 1.5). A band-pass filter in the frequency range [0.01 0.4] Hz was applied using the *hmrBandpassFilt* function (3rd filter order) to remove physiological noise such as heart rate, low-frequency noise and slow drifts in the data. After that, the concentration changes of oxygenated hemoglobin (HbO_2) and deoxygenated hemoglobin (HbR) were calculated based on the modified Beer-Lambert law (Kocsis et al., 2006) using the *hmrOD2Conc* function, assuming a fixed differential path-length factor of 6 that is typically used for continuous-wave fNIRS (Yücel et al., 2016). To localize functional activation on the basis of one signal including the contribution of both HbO_2 and HbR, the correlation-based signal improvement (CBSI) method (Cui et al., 2010) was used to combine the pre-processed HbO_2 and HbR into the activation signal. Tachtsidis and Scholkmann (2016) suggested that this approach has the potential to reduce false positives in statistical inference analyses.

Table 5.2. ***Channel centroid Montreal Neurological Institute (MNI) coordinates.***

Channel	MNI Coordinates[b]			Channel	MNI Coordinates[b]		
number[a]	x	y	z	number[a]	x	y	z
1 (1)	-33	57	23	41 (18)	24	-98	16
2 (27)	-33	47	35	42 (15)	26	-98	-10
3 (1)	-48	45	12	43 (19)	31	-86	35
4 (27)	-49	36	24	44 (20)	36	-94	5
5 (25)	-50	27	34	45 (24)	34	-91	-20
6 (23)	-59	19	9	46 (21)	33	-74	52
7 (23)	-60	10	19	47 (22)	44	-82	24
8 (27)	-64	-1	-10	48 (25)	44	-87	-5
9 (22)	-59	3	37	49 (18)	34	-61	63
10 (4)	-66	-7	10	50 (13)	51	-68	40
11 (25)	-67	-16	-22	51 (16)	53	-76	8
12 (20)	-52	-5	51	52 (15)	48	-79	-20
13 (15)	-65	-14	27	53 (11)	52	-55	51
14 (17)	-69	-22	-2	54 (7)	60	-62	24
15 (24)	-45	-14	63	55 (6)	58	-68	-7
16 (9)	-63	-21	42	56 (21)	46	-38	64
17 (16)	-69	-31	12	57 (19)	63	-44	42
18 (19)	-67	-38	-17	58 (9)	65	-54	11
19 (26)	-56	-30	54	59 (18)	61	-60	-17
20 (14)	-67	-39	27	60 (16)	57	-27	55
21 (19)	-67	-47	-3	61 (17)	68	-36	30
22 (32)	-44	-40	64	62 (21)	69	-45	1
23 (27)	-61	-47	40	63 (7)	46	-12	63
24 (20)	-64	-56	8	64 (6)	65	-19	44
25 (26)	-58	-61	-20	65 (0)	70	-29	15
26 (26)	-49	-58	50	66 (23)	69	-37	-14
27 (24)	-59	-65	21	67 (9)	54	-4	52
28 (21)	-56	-70	-9	68 (0)	67	-12	30
29 (32)	-30	-63	63	69 (1)	71	-21	1
30 (22)	-49	-71	38	70 (14)	61	4	38
31 (23)	-51	-79	6	71 (0)	68	-6	13
32 (22)	-45	-81	-21	72 (27)	69	-16	-19
33 (26)	-30	-76	51	73 (5)	62	10	21
34 (23)	-42	-85	23	74 (17)	66	-1	-7
35 (22)	-41	-90	-6	75 (15)	52	27	35
36 (21)	-28	-88	35	76 (14)	62	19	11
37 (24)	-33	-97	4	77 (26)	51	36	25
38 (26)	-30	-94	-20	78 (22)	36	48	35
39 (23)	-19	-99	16	79 (0)	51	45	12
40 (28)	-22	-101	-10	80 (0)	36	57	22
81 (25)	-64	-28	42	82 (18)	66	-25	45

Note. [a]Number of good datapoints out of the total 33 datapoints for each channel in brackets.

[b]Coordinates are based on the MNI system in mm, (-) indicates left hemisphere.

5.2.6.3 Channel Digitization and Localization

Conventionally, it is assumed that the same probe would fall in the same location across participants in non-invasive functional neuroimaging studies. However, this assumption seems to not hold in the current study. As shown in *Figure 5.4*, the locations of the same channel between participants (i.e. small dots with the same colour) have significant variability. Despite carefully placing the cap and optodes in the same way for each participant, the same channels may not overlay the same cortical regions for everyone, potentially because of individual differences in head size and shape and systematic errors. Nevertheless, such inconsistency of channel positions across participants does not only exist in the current study, it is a common issue in fNIRS studies, which may undermine fNIRS spatial resolution and the estimation accuracy of group-level effects (Tak et al., 2016; Zimeo Morais et al., 2018).

To minimize the negative impact of this issue, I decided to take into account the contribution of each participant's channel locations by also looking at channels that are close to the identified cortical regions, instead of assuming specific channels overlay those regions for all participants. Specifically, digitization was conducted for all participants before starting fNIRS data acquisition. A Liberty 3D electromagnetic tracking system (Polhemus, Colchester, VT) was used to determine anatomical locations of fNIRS optodes in relation to head landmarks based on the 10-20 electrode placement system, including Nasion, Inion, right and left preauricular points, and Vertex (or Cz). Then, the fNIRS channel locations from real space, specifically to each individual, were co-registered onto a standard brain template, and the Montreal Neurological Institute (MNI) coordinates (Mazziotta et al., 2001a, 2001b) of each channel were estimated for each participant using the NIRS-SPM package (Ye et al., 2009) with MATLAB (Mathworks, Natick, MA). This allowed the locations of each channel to be compared across participants, thus dealing with the substantial individual variability. The corresponding anatomical locations of each channel are presented on a brain surface rendered in MNI coordinates (see *Figure 5.4*; small colourful dots) using a collection of MATLAB functions (*simpleBrainSurface*; https://github.com/robertreingit/simpleBrainSurface). The mean of the resulting MNI coordinates for each channel centroid obtained from the entire sample were calculated (see Table 5.2 & *Figure 5.4*; big red dots).

5.2.6.4 Regions of Interest (ROIs): Channel Allocation

Given the regions of interest (ROIs) in the current study were not demarcated in terms of individual channels, I describe and discuss the ROIs and results in relation to the anatomical location of the activation. Functional ROIs that would potentially engage in ToM, intergroup bias or visual processing of facial expressions, which are of interest to the aims of the current study, were identified based on data available in the Neurosynth database (https://neurosynth.org/). Any brain regions that could engage in these processes but would not be able to be detected by NIRS technology were not considered in the current study.

For each ROI, a spherical space was created for which the point with the strongest positive activation was identified as the centre of the ROI with a radius of 2 cm. For each participant, all the channels within this area were averaged as the functional activation signal of the corresponding ROI for the participant. The MNI coordinates of 18 ROIs were identified for left and right hemispheres (see Table 5.3) and included in the following inferential statistical analysis: *temporoparietal junction* [181 studies], *temporal sulcus* [518 studies], *superior temporal cortex* [1422 studies], *dorsolateral prefrontal cortex* [1049 studies], *postcentral gyrus* [184 studies], *extrastriate* [246 studies], *middle frontal* [682 studies], and *inferior frontal* [1890 studies]. As this was an exploratory study, ROIs with good data quality were considered for the group-level statistical analysis. Following this process, ROIs had on average 41 allocated channels from on average 21 participants.

Table 5.3. ***ROI coordinates.***

ROI	Hemisphere	MNI Coordinates[a]			Participant	Channel
		x	y	z	number	number
Temporoparietal junction	Left	-54	-56	22	26	54
	Right	58	-56	18	16	23
Postcentral gyrus	Left	-58	-18	34	27	73
	Right	60	-8	20	7	8
Extrastriate visual cortex	Left	-50	-76	4	29	76
	Right	50	-72	0	24	50
Middle frontal gyrus	Left	-48	22	20	29	72
	Right	50	22	12	19	28
Inferior frontal gyrus	Left	-50	30	-8	8	8
	Right	50	22	4	12	14
Temporal sulcus	Left	-50	-56	12	27	48
	Right	52	-40	6	9	10
Superior temporal cortex	Left	-58	-16	0	24	41
	Right	60	-32	4	15	19
Dorsolateral prefrontal cortex	Left	-46	34	32	29	72
	Right	42	38	32	28	59

Note. [a]Coordinates are based on the MNI system in mm, (-) indicates left hemisphere.

5.2.6.5 Contrast Effects Analysis

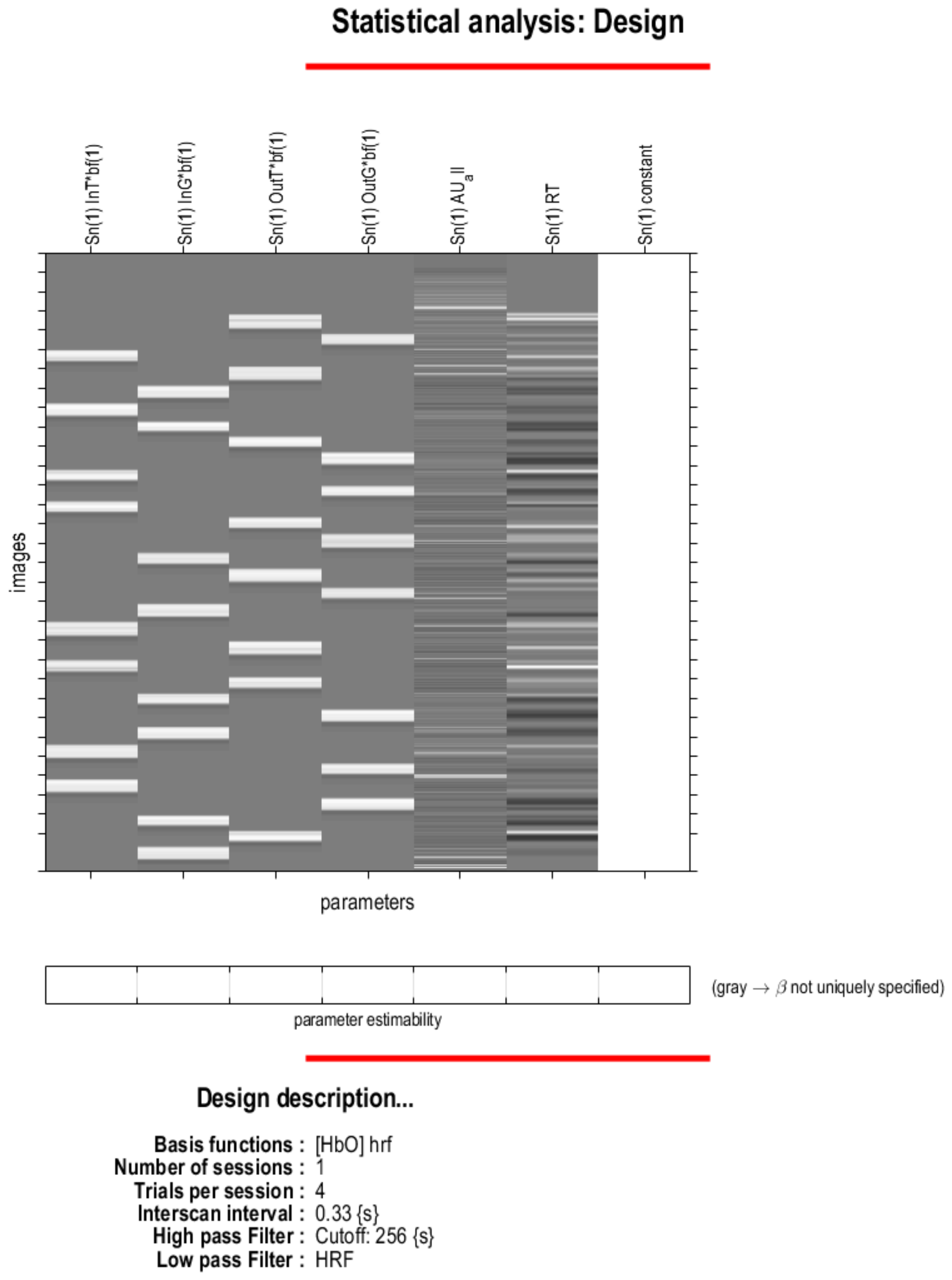
A first-level (or single-subject) channel-wise general linear model (GLM; Friston et al., 1994) was built for each participant to fit the fNIRS activation signals, down-sampled to 3 Hz, using the SPM for fNIRS toolbox (https://www.nitrc.org/projects/spm_fnirs/) to localize functional brain activity occurring in response to the task. For each participant, the design matrix (see *Figure 5.5* as an example) included four categorical regressors modelling the four corresponding task conditions and two additional parametric regressors that accounted for the AU action (the production of participants' own facial movements) and the reaction time of each trial (RT) respectively: ToM-Ingroup (ToM-In), Gender-Ingroup (Gender-In), ToM-Outgroup (ToM-Out), Gender-Outgroup (Gender-Out), AU_all and RT. Single-subject beta values were estimated for each of the six regressors.

To generate the AU_all regressor, a column was added to the design matrix for each participant to model the amount of facial AU movement of the participant over the entire fNIRS recording. To synchronize the AU data and the fNIRS activation signals, I trimmed the former to the same length as the latter for each participant. The specific clock time of the first block information page in the screen recording which is also the first onset of the brain signals was compared with the camera video onset time; any excess was trimmed. The rest of the data was first down-sampled to 3fps, the same frequency as the processed fNIRS data, and then I compared its length with the fNIRS data; the difference was either trimmed or imputed using the mean of the AU data after trimming the beginning. Any missing data in other time points was also imputed using the mean.

Similarly, to generate the RT regressor, a column was added to the design matrix for each participant to model the underlying neural/cognitive processes that related to the button press and the short rest period after the button press which could not be captured in the current study. To fit the RT regressor with the fNIRS activation signals, an RT parameter was created with the same length and timeline as the processed fNIRS data; each datapoint along the timeline is the total RT of the corresponding trial. Any non-trial datapoints were imputed with the grand mean of the RT, and any missing data (10 out of 4752) was imputed using the corresponding condition means. Then, the grand mean was adjusted to zero for each participant, and the data was convolved with the Hemodynamic Response Function (HRF).

To localize brain activation at the group level, specific contrasts were generated among the six regressors (i.e. ToM-In, Gender-In, ToM-Out, Gender-Out, AU_all and RT):
Contrast 1 – Main effect of ToM: [ToM-In + ToM-Out] vs. [Gender-In + Gender-Out]
Contrast 2 – Main effect of Group: [ToM-In + Gender-In] vs. [ToM-Out + Gender-Out]
Contrast 3 – Simple Group effect in ToM question: [ToM-In] vs. [ToM-Out]
Contrast 4 – Simple Group effect in Gender question: [Gender-In] vs. [Gender-Out]
Contrast 5 – Simple ToM effect in Ingroup: [ToM-In] vs. [Gender-In]
Contrast 6 – Simple ToM effect in Outgroup: [ToM-Out] vs. [Gender-Out]
Contrast 7 – ToM x Group interaction: [ToM-In + Gender-Out] vs. [ToM-Out + Gender-In]
Contrast 8 – Main effect of face movement (AU)
Contrast 9 – Main effect of reaction time (RT)

For each contrast comparison, one-sample *t*-tests were conducted as the inferential statistical approach on the beta estimates for each ROI. Since the current study was exploratory, there was no correction for multiple comparisons.



Statistical analysis: Design

Design description...

Basis functions :	[HbO] hrf
Number of sessions :	1
Trials per session :	4
Interscan interval :	0.33 {s}
High pass Filter :	Cutoff: 256 {s}
Low pass Filter :	HRF

Figure 5.5. **A first-level GLM design matrix. As the presentation sequence, AU action and RT are varied across participants, the design matrix should be unique for each participant.**

5.3 Results

All effects are reported as significant at $p < .05$, and two-tailed p values are reported throughout, if not specified. IBM SPSS Statistics (Version 29) was used to conduct statistical analyses for behavioural data; MATLAB R2022b was used for fNIRS data.

5.3.1 Behavioural Performance

5.3.1.1 Group Identification

Group identification scores were analysed using a paired samples *t*-test. Participants identified more strongly with ingroup members ($M = 4.52$, $SD = 1.24$) than outgroup members ($M = 3.13$, $SD = 0.97$), $t(32) = 5.06$, $p < .001$, $d = 0.881$, indicating that the minimal group manipulation had worked.

5.3.1.2 Group and ToM Effects

To test the effects of intergroup bias and ToM on smile identification, I conducted a two-way repeated-measures analysis of variance (ANOVA) for accuracy and RT respectively, with Group (ingroup vs. outgroup) and ToM (ToM vs. non-ToM) as within-subjects variables. For both of the outcome variables, there was a main effect of ToM, accuracy: $F(1, 32) = 418.30$, $p < .001$, partial $\eta^2 = .929$, RT: $F(1, 32) = 34.28$, $p < .001$, partial $\eta^2 = .517$, but no Group effect, accuracy: $F(1, 32) = 2.90$, $p = .099$, partial $\eta^2 = .083$, RT: $F(1, 32) = 0.62$, $p = .438$, partial $\eta^2 = .019$, nor interaction, accuracy: $F(1, 32) = 0.78$, $p = .385$, partial $\eta^2 = .024$, RT: $F(1, 32) = 1.80$, $p = .189$, partial $\eta^2 = .053$. Specifically, accuracy was lower, and RT was higher in the ToM question than in the Gender control question (see *Figures 5.6 & 5.7*). Accordingly, the varied RT between questions might reflect neural/cognitive processes that the current task cannot capture, so the RT for all the answered trials was added as a regressor to model these processes in the contrast effects analysis, as mentioned in Section 5.2.6.5.

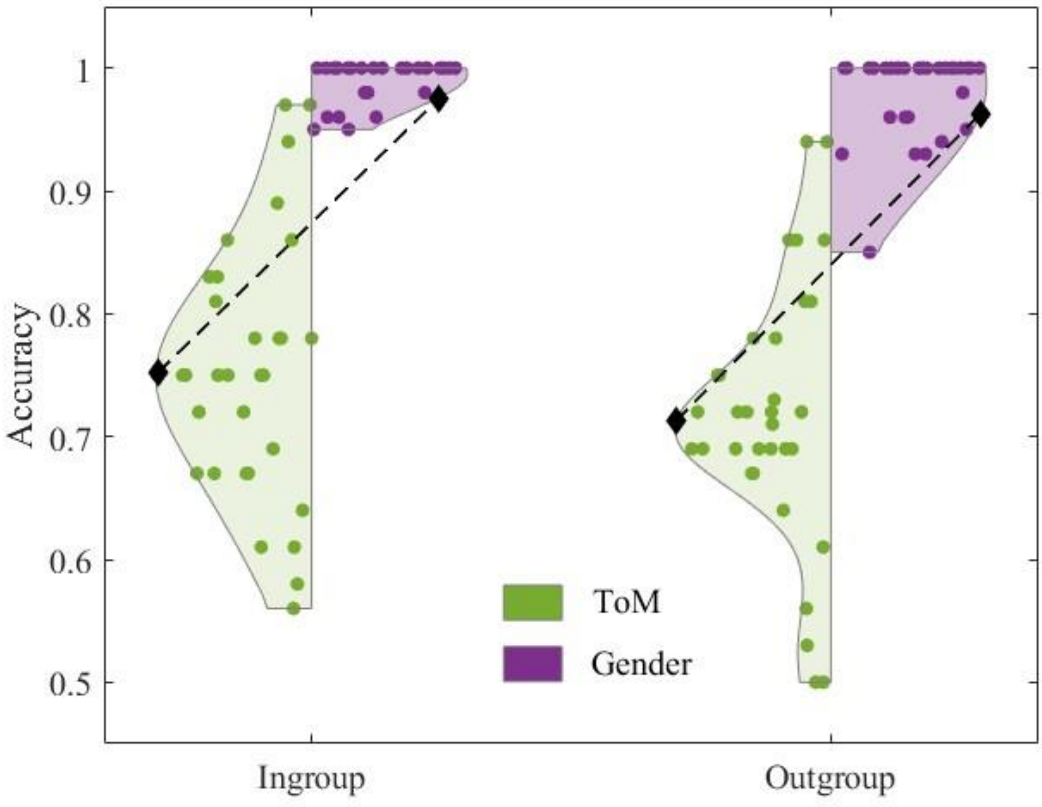


Figure 5.6. **The smile discrimination accuracy by ToM and Group (each dot represents the mean accuracy of each participant); black diamonds represent the mean of each condition.**

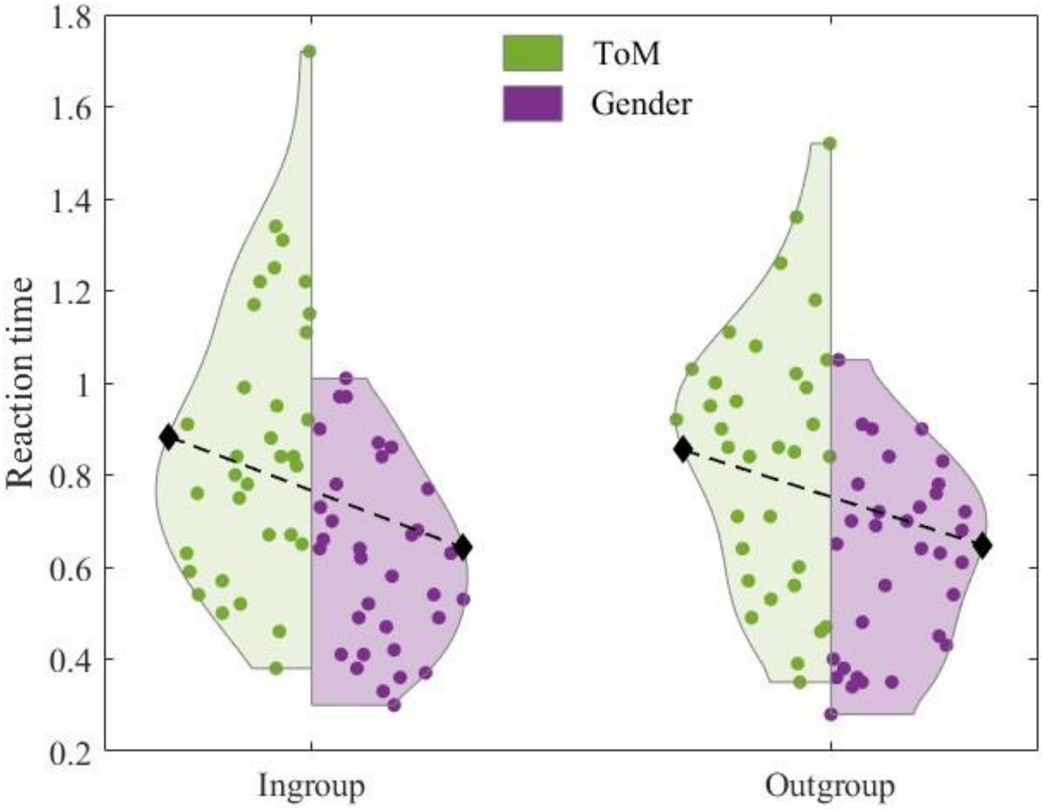


Figure 5.7. **The smile discrimination reaction time (RT) by ToM and Group (each dot represents the mean RT of each participant); black diamonds represent the mean of each condition.**

5.3.1.3 Smile Type and Group Effects

To test response bias, one-sample *t*-tests were conducted genuine smile proportion in Ingroup-ToM and Outgroup-ToM conditions respectively, with 0.5 as the test chance value. The results showed that the proportions of the time participants gave genuine response were significantly below the chance level in the Outgroup-ToM condition (*M* = 0.45, *SD* = 0.09), *t*(32) = -3.49, *p* = .001, *d* = -0.61; but this tendency was only marginally significant in the

Ingroup-ToM condition (*M* = 0.46, *SD* = 0.11), *t*(32) = -1.98, *p* = .056, *d* = -0.35. A paired samples *t*-test revealed no proportion difference of selecting genuine smile between the two conditions, *t*(32) = 1, *p* = .325, *d* = 0.17. These indicated that participants had a response bias towards judging smiles as posed smiles and this tendency was not affected by their group membership.

I further examined the effects of smile type and its interaction with intergroup bias in smile discrimination in ToM and non-ToM (Gender) conditions separately on RT. Given incorrect trials might indicate a failure of the desired perceptual/cognitive processes, that is ToM, smile perception and discrimination, only the RT for accurate trials, trials that have been correctly answered, were included. A two-way repeated-measures ANOVA was conducted using RT as the outcome variable respectively, with Group (ingroup vs. outgroup) and either Smile type (genuine vs. posed) or Gender type (female vs. male) as within-subjects factors. In the ToM condition, neither main effect nor interaction was found, Smile type, *F*(1, 32) = 2.03, *p* = .164, partial η^2 = .060, Group, *F*(1, 32) = 2.54, *p* = .121, partial η^2 = .074, interaction, *F*(1, 32) = 0.29, *p* = .591, partial η^2 = .009 (see *Figure 5.9*). In the non-ToM (Gender) condition, the results showed a main effect of Gender type, *F*(1, 32) = 32.67, *p* < .001, partial η^2 = .505, specifically, participants spent more time identifying female faces than male faces. But, there was no main effect of Group, *F*(1, 32) = 0.25, *p* = .875, partial η^2 = .001, nor interaction, *F*(1, 32) = 1.59, *p* = .217, partial η^2 = .047 (see *Figure 5.9*).

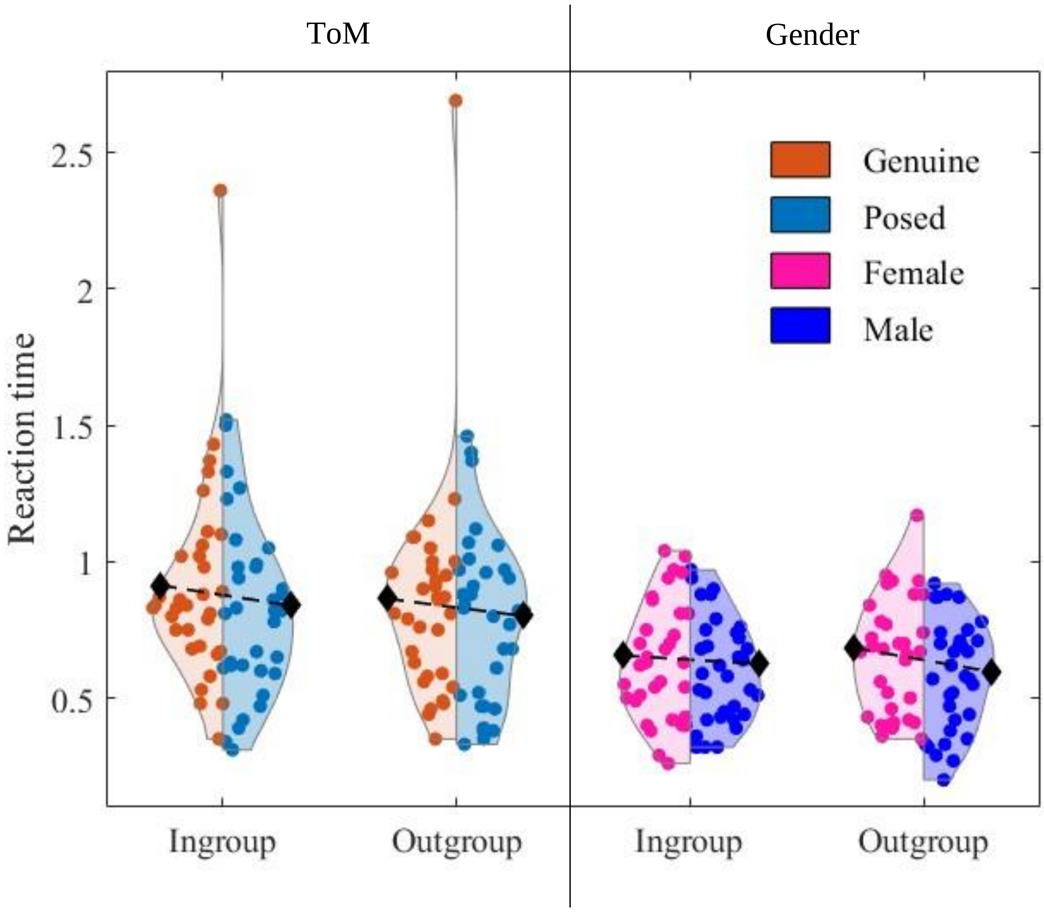


Figure 5.8. **The smile discrimination reaction time (RT) by Smile/Gender type and Group (each dot represents the mean RT of each participant); black diamonds represent the mean of each condition.**

Given both actors and participants including females and males, there could be a gender intergroup bias that confounded the targeted minimal intergroup bias. Thus, I decided to run the same analysis but add the participants' own gender as a between-subjects factor in the Gender model. The results remained the same and everything related to this factor was not significant, indicating that participants' own gender cannot explain the variance observed in the task (See Table 5.4).

Table 5.4. ***GLM in Gender condition with participants' gender as a between-subjects factor.***

	Effects	Inferential statistics
Accuracy	Group	$F(1, 31) = 1.77$, $p = .194$, partial $\eta^2 = .054$
	Gender type	$F(1, 31) = 0.31$, $p = .581$, partial $\eta^2 = .010$
	Participant gender	$F(1, 31) = 0.61$, $p = .440$, partial $\eta^2 = .019$
	Group*Gender type	$F(1, 31) = 2.30$, $p = .139$, partial $\eta^2 = .069$
	Gender type*Participant gender	$F(1, 31) = 0.05$, $p = .826$, partial $\eta^2 = .002$
	Group*Participant gender	$F(1, 31) = 1.19$, $p = .283$, partial $\eta^2 = .037$
	3-way interaction	$F(1, 31) = 0.02$, $p = .903$, partial $\eta^2 < .001$
RT	Group	$F(1, 31) = 0.03$, $p = .861$, partial $\eta^2 = .001$
	Gender type	**$F(1, 31) = 33.08$, $p < .001$, partial $\eta^2 = .516$**
	Participant gender	$F(1, 31) = 0.06$, $p = .808$, partial $\eta^2 = .002$
	Group*Gender type	$F(1, 31) = 1.53$, $p = .226$, partial $\eta^2 = .047$
	Gender type*Participant gender	$F(1, 31) = 1.07$, $p = .308$, partial $\eta^2 = .033$
	Group*Participant gender	$F(1, 31) = 0.41$, $p = .526$, partial $\eta^2 = .013$
	3-way interaction	$F(1, 31) = 0.04$, $p = .840$, partial $\eta^2 = .001$

5.3.1.4 Item-wise Analysis

To evaluate the validity and quality of the materials in the smile discrimination task, I carried out an item-wise analysis for each actor in both ToM and Gender conditions. For each participant, I averaged the accuracy for each actor in ToM and Gender conditions separately and then calculated the mean of the accuracy of all participants seeing each actor in each condition. One-sample *t*-tests were conducted on genuine, posed, female and male smilers, with 0.5 as the test chance value. The results showed that the accuracies of all four smiler types were significantly above 0.5; genuine (*M* = 0.69, *SD* = 0.21): *t*(17) = 3.90, *p* = .001, *d* = 0.92, posed (*M* = 0.78, *SD* = 0.18): *t*(17) = 6.56, *p* < .001, *d* = 1.55, female (*M* = 0.99, *SD* = 0.03): *t*(14) = 65.62, *p* < .001, *d* = 17.54, male (*M* = 0.99, *SD* = 0.03): *t*(21) = 77.03, *p* < .001, *d* = 16.42. However, although the majority of genuine and posed smilers can be identified above the chance level 0.5, there were three genuine smilers (i.e., “Emily”, “Jacob” and “Amanda”; see *Figure 5.10*) and two posed smilers (i.e., “Rachel” and “Paul”; see *Figure 5.11*) who were correctly identified below or at chance level; some of them were even consistently recognized as the opposite category. As shown in *Figure 5.12*, all actors’ gender was correctly identified significantly above chance; however, as 34 out of 36 were totally accurate or very close to that, “Alex” and “Ashley” seemed to create confusion for a few participants occasionally.

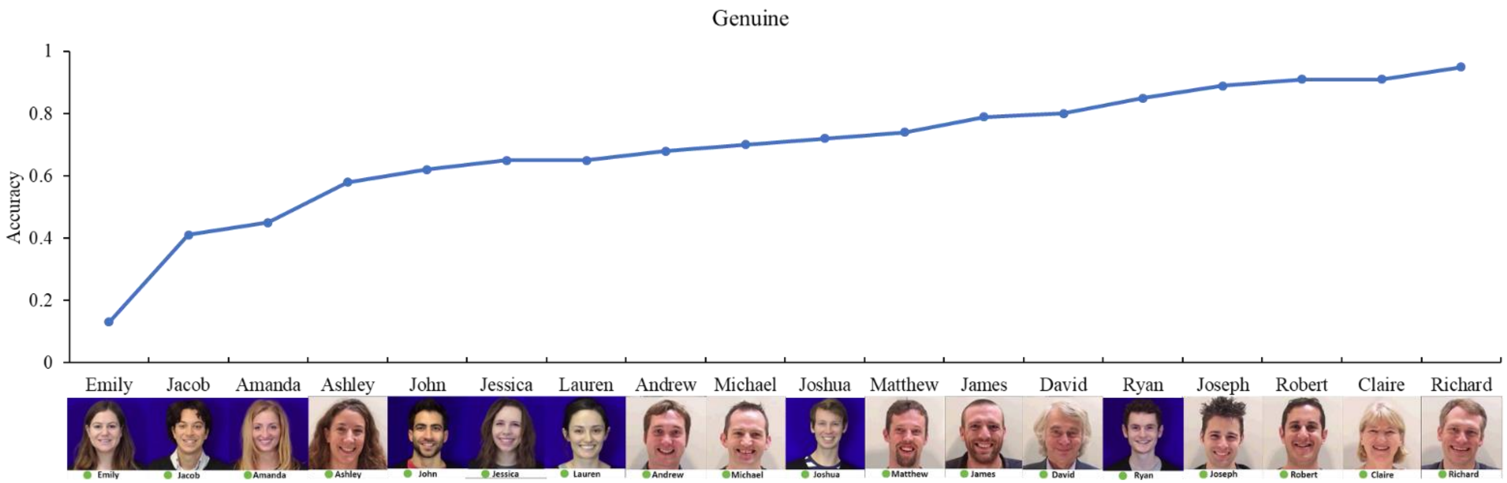


Figure 5.9. **Mean smile discrimination accuracy across participants for all the actors provided genuine smiles.**

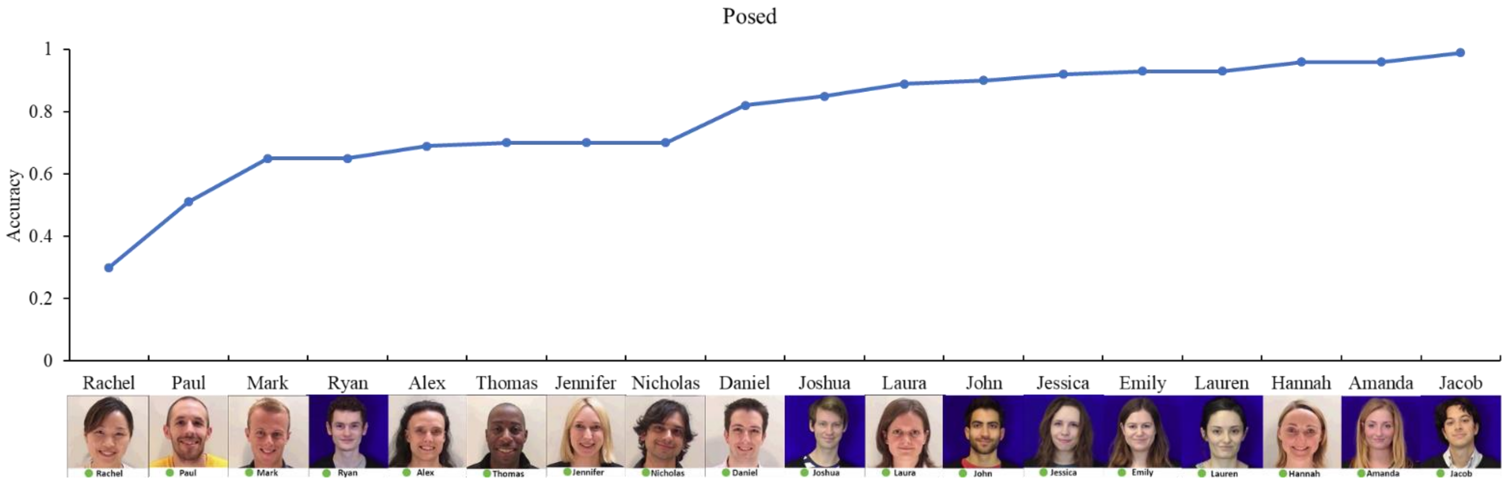


Figure 5.10. **Mean smile discrimination accuracy across participants for all the actors provided posed smiles.**

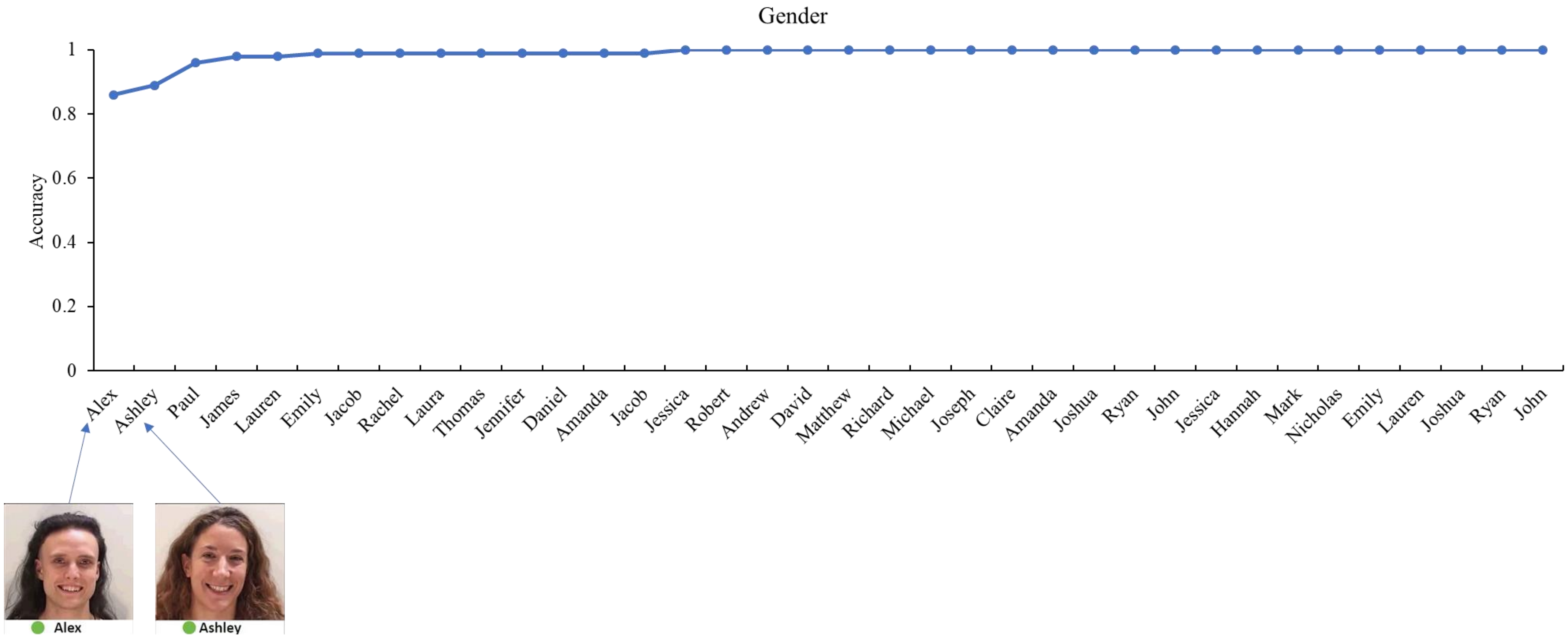


Figure 5.11. **Mean gender identification accuracy across participants for all the actors.**

5.3.1.5 Facial Movement

To test the effects of ToM and intergroup bias on participants' facial movements, I aligned participants' face recording and the task screen recording data and calculated the total AU intensity of each condition for each participant, as mentioned in Section 5.2.5. A two-way repeated-measures ANOVA was conducted on the AU intensity, with Group (ingroup vs. outgroup) and ToM (ToM vs. non-ToM) as within-subjects variables. There was no main effects nor interaction: ToM: $F(1, 32) = 0.58$, $p = .453$, partial $\eta^2 = .018$, Group: $F(1, 32) = 0.82$, $p = .371$, partial $\eta^2 = .025$, interaction: $F(1, 32) = 0.09$, $p = .764$, partial $\eta^2 = .003$ (see *Figure 5.13*).

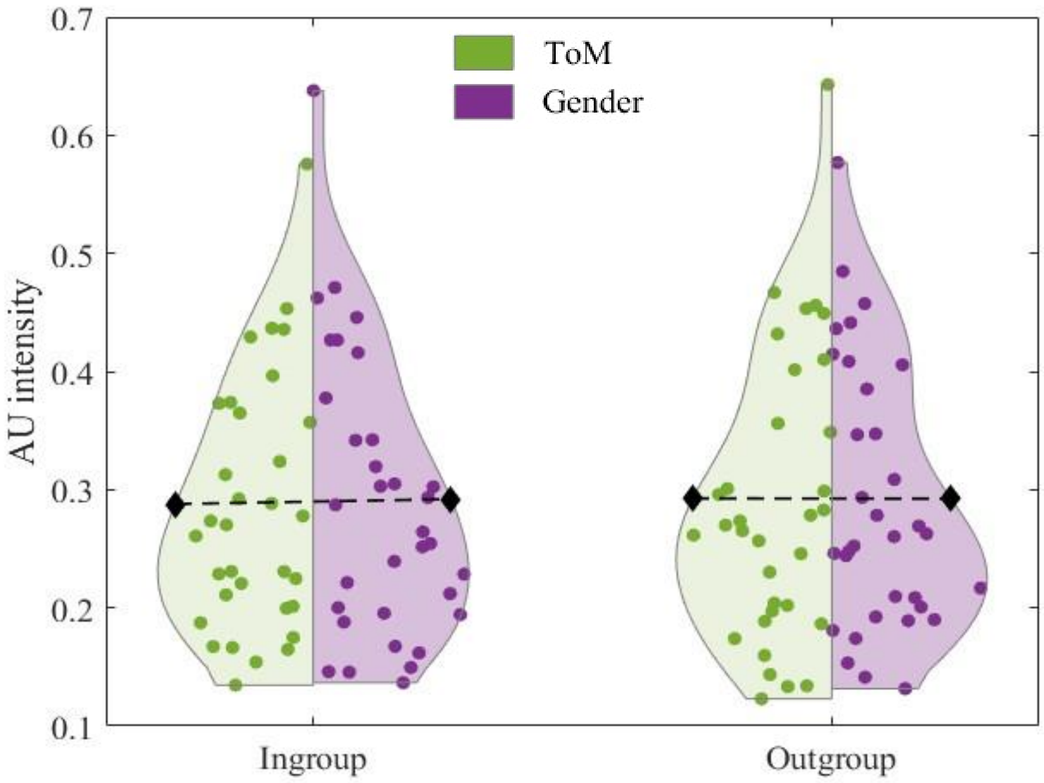


Figure 5.12. **The AU intensity by ToM and Group (each dot represents the mean intensity of each participant); black diamonds represent the mean of each condition.**

5.3.2 Brain Activation

This section reports the results of the group-level GLM analysis conducted on the fNIRS neural activation signals (i.e. the CBSI – the combination of HbO_2 and HbR signals) of the brain ROIs specified in Section 5.2.6.4 for the contrasts listed in Section 5.2.6.5. Specifically, six regressors (i.e. ToM-In, Gender-In, ToM-Out, Gender-Out, AU_all and RT) were included in the design matrix to fit the fNIRS data and estimate the beta values for each participant (see Section 5.2.6.5). These beta values were used to measure how intergroup bias, ToM, AU, RT, and the interaction between group membership and ToM modulate brain activity in each ROI. In *Figures 5.14-5.16*, the group averaged beta values of the significant contrasts for $p < .05$ are marked with asterisks, and the corresponding ROIs are circled in red.

5.3.2.1 Main and Simple Group Effects (Contrasts 2, 3 and 4)

Concerning the main effect of Group, I found that ingroup faces were associated with an increase in brain haemodynamic activity in the *left MFG (lMFG), t* = 2.15, *p* = .040, *left dlPFC (ldlPFC), t* = 2.47, *p* = .020, and *right dlPFC (rdlPFC), t* = 2.62, *p* = .014 (see *Figure 5.14A, B & C*), but a decrease in the *right IFG (rIFG), t* = -2.47, *p* = .031, compared to outgroup faces (see *Figure 5.15A*). Similarly, for simple Group effects, in the ToM condition, an increased level of activation was observed in the *lMFG, t* = 2.47, *p* = .020, and *ldlPFC, t* = 2.43, *p* = .022, during ingroup over outgroup faces (see *Figure 5.14A & B*); in the Gender condition, *ldlPFC, t* = 2.04, *p* = .051, and *rdlPFC, t* = 3.09, *p* = .005, were significantly more engaged when watching ingroup faces (see *Figure 5.14B & C*).

5.3.2.2 Main and Simple ToM Effects (Contrasts 1, 5 and 6)

To investigate the effect of ToM, I compared the ToM and Gender conditions as well as within both ingroup and outgroup faces. Interestingly, for the main effect of ToM, a decrease in the ToM compared to Gender conditions was observed in the *left PCG (lPCG), t* = -2.76, *p* = .010, and *left TS (lTS), t* = -2.32, *p* = .028 (see *Figures 5.15B & 5.16A*). Similarly, for simple effects of ToM during ingroup faces, the *rdlPFC, t* = -2.16, *p* = .040, the *lPCG, t* = -2.14, *p* = .042, and *right EVC (rEVC), t* = -2.38, *p* = .026, were more activated in the Gender condition (see *Figures 5.14C, 5.15B & 5.16C*). On the other hand, an increase in the ToM condition was only observed in the *rIFG, t* = 2.76, *p* = .019, and *right STC (rSTC), t* = 2.18, *p* = .047, during outgroup faces (see *Figures 5.15A & 5.16B*).

5.3.2.3 Interaction Between Group and ToM (Contrast 7)

An interaction between the ToM and Group factors was observed in the *rdlPFC*, *t* = -2.35, *p* = .026. This is likely driven by the ToM effect during ingroup faces (see Section 5.3.2.2), the main effect of Group and the Group effect in the Gender condition (see Section 5.3.2.1) in this region. Specifically, ingroup faces triggered higher activation than outgroup faces in the Gender condition but not in the ToM condition (see *Figure 5.14C*).

5.3.2.4 AU and RT Effects (Contrasts 8 and 9)

Finally, I tested whether any variance of brain activation can be explained by the effects of participants' own facial behaviour (AU) and their reaction time (RT) during the task. No significant effect was associated with the two factors in any of the ROIs.

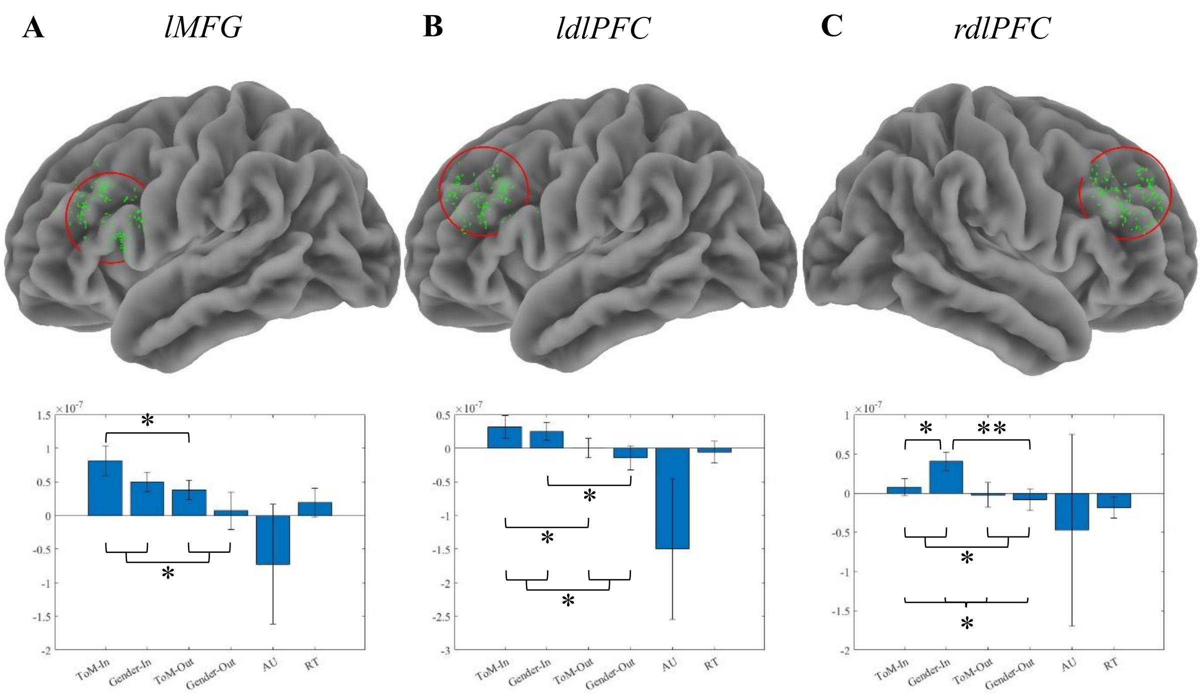


Figure 5.13. **The group-level GLM results on the significant ROIs fNIRS activation signals for the nine contrasts of interests. Green dots indicate the included channels from all participants, the corresponding ROIs are circled in red. The group-averaged beta values for all regressors (i.e. ToM-In, Gender-In, ToM-Out, Gender-Out, AU, and RT, same order in the X-axis) are presented in the bar charts. ****p*** **< .05, *****p*** **< .01.** ***lMFG = left middle frontal gyrus, ldlPFC = left dorsolateral prefrontal cortex, rdlPFC = right dorsolateral prefrontal cortex*****.**

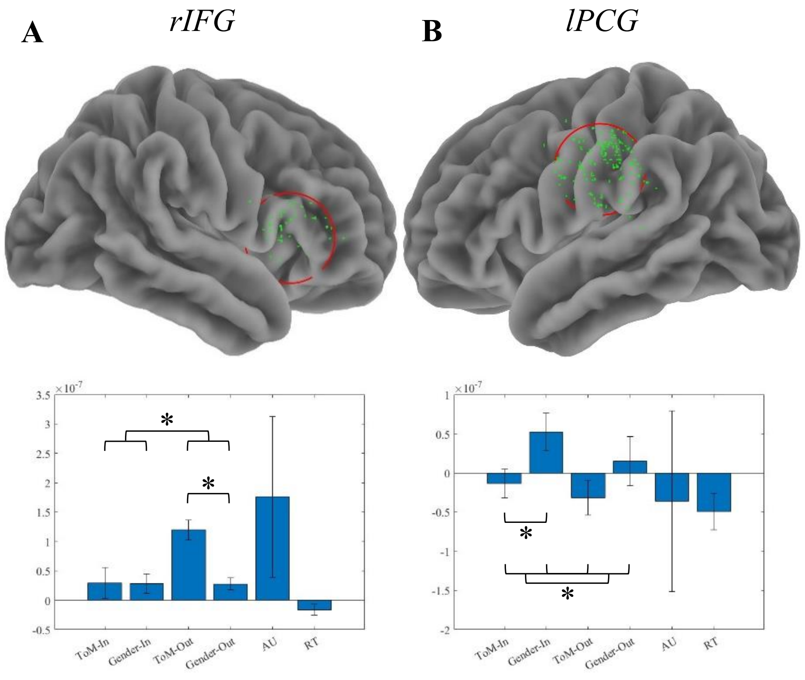


***Figure 5.14.* The group-level GLM results on the significant ROIs fNIRS activation signals for the nine contrasts of interests. Green dots indicate the included channels from all participants, the corresponding ROIs are circled in red. The group-averaged beta values for all regressors (i.e. ToM-In, Gender-In, ToM-Out, Gender-Out, AU, and RT, same order in X-axis) are presented in the bar charts. **p* < .05. *rIFG = right inferior frontal gyrus, lPCG = left postcentral gyrus*.**

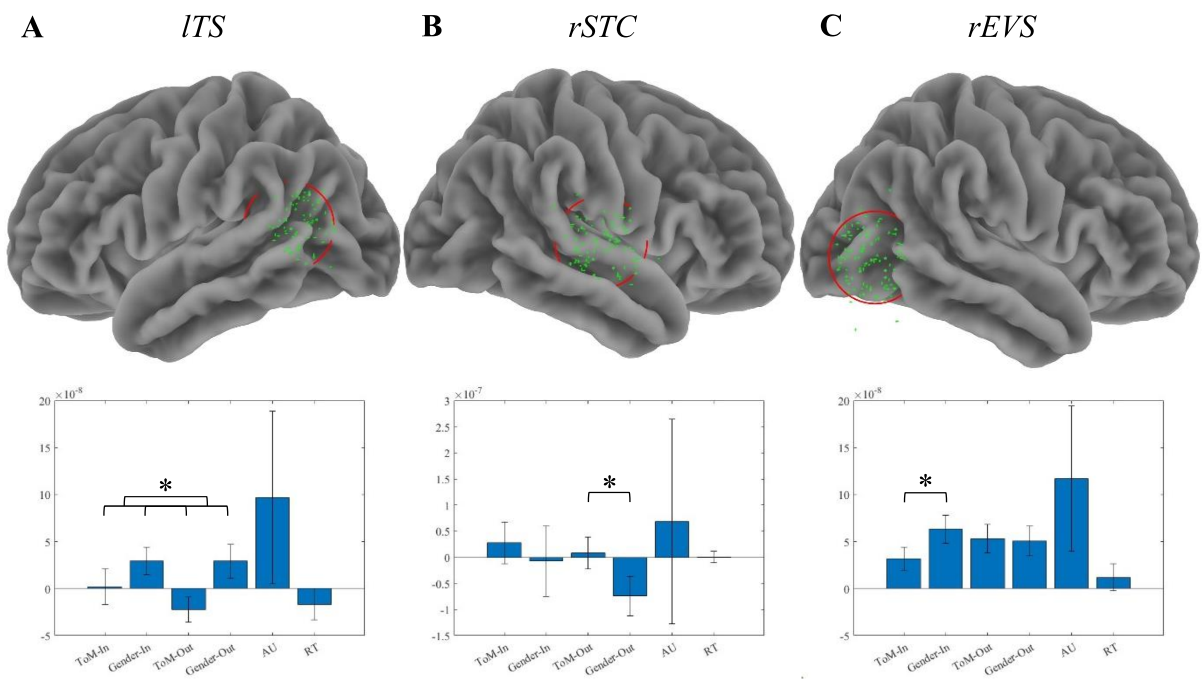


Figure 5.15. **The group-level GLM results on the significant ROIs fNIRS activation signals for the nine contrasts of interests. Green dots indicate the included channels from all participants, the corresponding ROIs are circled in red. The group-averaged beta values for all regressors (i.e. ToM-In, Gender-In, ToM-Out, Gender-Out, AU, and RT, same order in X-axis) are presented in the bar charts. *p < .05. *lTS = left temporal sulcus, rSTC = right superior temporal cortex, rEVC = right extrastriate visual cortex*.**

5.4 Discussion

The current study primarily aimed to investigate the underlying neural mechanisms involved in intergroup bias during the smile discrimination task, as well as the effect of intergroup bias on social mimicry. I secondarily aimed to replicate the outgroup advantage in Young (2017) that people would be more accurate in distinguishing between genuine and posed smiles from outgroup members than from ingroup members. The neural results revealed a main effect of intergroup bias in smile discrimination but not in mimicry. Specifically, the *lMFG* and *bilateral dlPFC* were more activated during ingroup smiles, while the *rIFG* was more activated during outgroup smiles. Additionally, a simple effect of ToM was found in outgroup conditions: the *rIFG* and *rSTC* showed greater activation in the ToM conditions compared with Gender conditions. The behavioural results showed a main effect of ToM, that participants were less accurate and spent more time in ToM conditions than in non-ToM (Gender) control conditions. However, there was no indication of the ToM effect in facial mimicry. Moreover, the results did not show any effect of intergroup bias in smile discrimination and facial mimicry at the behavioural level.

5.4.1 Group Identification

Replicating the findings in Chapter 4, ingroup identification was higher than outgroup, which is in line with the literature that people are more likely to tie and be willing to identify themselves with ingroup members despite the completely arbitrary grouping (e.g., Doosje et al., 1995; Tajfel, 1970; Young & Hugenberg, 2010). This result also indicates that the minimal group manipulation was valid, which made it possible to next explore and discuss the experimental aims of the current study.

5.4.2 Intergroup Bias

5.4.2.1 Intergroup Bias in Behaviour

Unexpectedly, there was no accuracy or reaction time difference between differentiating ingroup and outgroup smiles during ToM conditions. Accordingly, the current study failed to replicate the behavioural outgroup advantage in both ToM (Adams Jr et al., 2010) and smile discrimination (Young, 2017) even with a method that more closely replicated past findings. However, this result replicates the findings in Chapter 4. Thus, I may reasonably suspect that intergroup bias might facilitate positive evaluation towards one's ingroup but cannot modulate the absolute ability to differentiate genuine from posed facial emotional expressions.

Another possibility for these results is that the current paradigm might not easily facilitate intergroup modulation at the behavioural level. First, the stimuli were prototypical genuine or posed smiles (Farmer et al., 2021; Young, 2017), which might leave little space for group membership to influence accuracy. In other words, if a smile is unambiguously prototypical, its category is unlikely to be modulated by a moderate intergroup effect. However, as participants showed a bias to respond that smiles were posed, the stimuli seemed to be ambiguous to some extent and thus able to assess smile discrimination ability and facilitate intergroup modulation during the process. It should be noted that performance was above chance for both smile types, indicating that participants were able to differentiate between genuine and posed facial expressions. Moreover, as mentioned in Chapter 4, minimal group paradigms might have attenuated effects in lab-based studies compared with online studies. Presumably, there is more contextual information to distract participants in the lab, such as the company of at least one researcher and the implementation of the fNIRS equipment. Hence, the assigned membership could be less prominent than in online environments.

Lastly, the Asian-dominant sample (67%) in the current study might introduce a racial intergroup bias, especially when most of the actors were White people. As participants could be more influenced by pre-existing social groups, like racial groups (e.g., Adams Jr et al., 2010; Katsumi & Dolcos, 2018), with whom they share more similar sociocultural backgrounds, this multi-intergroup environment may prevent the current study from observing the implemented minimal intergroup bias. Indeed, there was only one Asian female in the actor pool, and the item-wise analysis showed that the posed smile from this actor (accuracy of 30%) tended to be consistently identified as genuine across participants. Considering the current Asian-dominant sample, this might indicate a racial ingroup favouritism. Moreover, Xie et al. (2019) suggested that an ingroup disadvantage in recognising subtle facial expressions may only exist in Chinese people but not in Western people. Thus, group membership may have distinct modulation effects in different sociocultural groups (e.g., Cheon et al., 2011; Han, 2018; Mathur et al., 2010).

5.4.2.2 Ingroup vs. Outgroup Neural Processing

The current study found some evidence of differential processing between ingroup and outgroup at the neural level in both the ToM and non-ToM (Gender) conditions. Neural activation in the *middle frontal gyrus (MFG)* and *dorsolateral prefrontal cortex (dlPFC)* was enhanced for perceiving ingroup compared to outgroup smiles. I created ROIs in both the *MFG* and *dlPFC* because the literature suggested that they might be both relevant to intergroup bias, but in practice, the two ROIs anatomically overlapped (see *Figure 5.14A & B*) and these two brain areas showed very similar patterns of activation. Thus, the *MFG* and *dlPFC* would be considered as if they are the same region in the following discussion, named *dlPFC-MFG*. Consistent with the literature, this enhanced *dlPFC-MFG* activation is likely to be associated with the top-down regulation of intergroup bias on attention and evaluation processes (Bartholow & Henry, 2010; Katsumi & Dolcos, 2018; Zhang et al., 2023). Group membership may modulate the saliency of ingroup and outgroup information, accordingly, attention could be reoriented to more salient ingroup members through executive control (e.g., Decety & Lamm, 2007; Eberhardt, 2005; Moradi et al., 2020; Mullen et al., 1992; Schupp et al., 2003).

I found greater *IFG* activity for the outgroup compared to the ingroup facial observation. This is also inconsistent with Peng et al. (2021). The *IFG* has been recognized as part of the MNS and is linked to facial emotional perception and social cognition processes (e.g., Brunet et al., 2000; Herwig et al., 2010; Krautheim et al., 2019; McLellan et al., 2012; Molenberghs et al., 2009). One potential explanation of this unexpected finding is that although the *IFG* is associated with facial emotional processing, it may contribute to mental state reasoning (e.g., Arioli et al., 2021; Dal Monte et al., 2014; Hooker et al., 2008; Molenberghs et al., 2016). Accordingly, the greater *IFG* activity for the outgroup may represent reduced mentalizing about the ingroup due to the ingroup favouritism, as mentioned in Park and Young (2020) and Hughes, Zaki, et al. (2017). They both found that people tend to engage in less mentalizing about ingroup than outgroup members, especially for negative mental states. These differences in brain activity for ingroup and outgroup members might be potential neural mechanisms for the human tendency to more readily favour ingroup members and more readily be vigilant and discriminate against outgroup members.

Notably, while the aforementioned neural activities provided evidence for minimal intergroup bias in smile perception-related brain activity, the behavioural smile discrimination results did not show any intergroup bias in the current study. Here I used the discrimination task which might be less sensitive than the judgement task used in Chapter 4. Thus, this could be why there is a brain effect but no behavioural effect in the current study. The brain effect might reflect the subtler smile judgement as shown in Chapter 4. Group membership may not bias smile discrimination, especially when the certainty of smile type is high, but it may modulate smile judgements as I found in Chapter 4. Accordingly, the Group effects may not indicate the neural correlates of the intergroup bias in smile discrimination. Those group modulation effects at the neural level are more likely in smile judgements. Hence, the current study is not able to confirm the underlying neural substrates of intergroup bias in smile discrimination, as the inference that specific brain areas are involved in this process cannot be made when there is no statistical difference between ingroup and outgroup smiles at the behavioural level. However, it is also possible that the observed effects of group membership did in fact represent the bias in smile discrimination at the neural level, as neural response tends to occur rapidly and implicitly, whereas smile discrimination performance requires deliberate and explicit reasoning. The modulation effect of group membership, if any, might have been overridden at the behavioural level (Han, 2018). Future studies should investigate this dissociation between behaviour and the neural correlates of intergroup bias in smile perception with more ambiguous stimuli.

5.4.3 ToM

5.4.3.1 ToM & Smile Discrimination in Behaviour

Consistent with the literature on ToM and the hypotheses, ToM tasks that recruit more advanced social cognitive skills, especially mentalizing ability, are more challenging than non-ToM tasks (e.g., White et al., 2011), like recognising simple features (e.g., gender) based on others’ appearance. Specifically, the results showed that identifying others’ emotional mental states from their smile videos not only took more time but was less accurate than identifying their gender.

In the non-ToM (Gender) conditions, males were identified much quicker than females despite their comparable accuracies. Gender classification has been suggested to correspond with masculine (e.g., beard, short hair, receding hairline, wide and angular facial contour, large and long lower face) and feminine (e.g., long hair, soft and round facial contour) facial features (Hoss et al., 2005; Mitteroecker et al., 2015). Accordingly, the shorter response time in male videos may be due to masculine features being more evident and salient than feminine features, at least in the particular stimuli used here. Indeed, Hoss et al. (2005) showed that high masculine facial features in males, but not high feminine features in females, can facilitate gender identification. This possibility is also in line with the item-wise analysis results that males with feminine features or females with masculine features might occasionally cause uncertainty but would not overturn people's judgment about their gender (see *Figure 5.12*).

5.4.3.2 Neural Correlates: ToM vs. Gender Conditions

In line with the behavioural findings and the literature, enhanced neural activation in the *IFG* and *STC* regions, core nodes of the social brain (Blakemore, 2008), was found in the ToM conditions compared with the Gender conditions; however, this only occurred for outgroup facial observation (e.g., Brunet et al., 2000; Katsumi & Dolcos, 2018).

If we look at the *IFG* in *Figure 5.15A* and the *ST* in *Figure 5.16A* closely, the mentalizing neural correlates were either enhanced in the ToM condition or suppressed in the non-ToM (i.e., Gender control) condition only for the outgroup. It seems likely that the intergroup modulation in the mentalizing process was "turned down" for ingroup members, which is consistent with Park and Young (2020)'s finding that people may inhibit mentalizing about ingroup members but prioritize mentalizing about outgroup members due to ingroup favouritism for maintaining social relationships. Moradi et al. (2020) also claimed that although attention naturally prefers ingroup, outgroup can be prioritized occasionally, as salience can be varied by context and goal, and the *dlPFC* might be involved in this effect. Thus, the mentalizing neural process could be modulated by intergroup bias which is in line with the literature. However, only consistent with a minority of studies, I found that mentalizing about outgroup members was prioritized in genuine and posed smile discrimination.

Unexpectedly, the results showed an interaction between ToM and intergroup bias in the *dlPFC*, despite it being absent in the behavioural performance. According to *Figure 5.14C*, the activation was only enhanced in the non-ToM condition of ingroup members. It seems likely that when ToM is not involved, like in the non-ToM (Gender) conditions, people tend to allocate more attention to ingroup than outgroup members, which is consistent with the literature on intergroup attentional bias (e.g., Amodio, 2014; Eberhardt, 2005; Zhang et al., 2023). However, attention allocation may be less biased between ingroup and outgroup members when people engage in mental state reasoning indexed by the absence of a simple Group effect in the ToM condition, which rejected the hypothesis and has been less discussed in the literature but is of particular interest.

As Rilling et al. (2008) suggested that higher cognitive efforts, like executive control, are needed to override intergroup discriminatory tendencies, mentalizing might have the same function. It is not true that ingroup members are actually superior to outgroup members, ingroup favouritism is an irrational and subjective tendency, thus when carefully evaluating both parties with sufficient information, the intergroup bias should be attenuated. In the current study, participants were explicitly asked to mentalize about ingroup and outgroup members, so participants had a chance to deliberately and consciously reason about their mental states. In this way, the same amount of attention might be allocated to ingroup or outgroup members in the ToM conditions, and therefore less concentrated on their favoured ingroup members as in the non-ToM (Gender) conditions. Thus, although ingroup information may be more salient and therefore can attract more attention than outgroup information, the recruitment of mentalizing might lead to a less biased attention distribution between ingroups and outgroup members, which may attenuate intergroup bias.

I also found enhanced activation in the *EVC* and *TS* in the non-ToM conditions, which however are not part of the hypotheses in the current study. They are presented for the sake of completeness, but they would not be further discussed.

5.4.3.3 The Role of TPJ

Unexpectedly, although the *TPJ* has been suggested to be involved in emotion recognition, mentalizing, and attentional reorientation (Baron-Cohen, Wheelwright, Hill, et al., 2001; Decety & Lamm, 2007; Frith & Frith, 2003; Gallagher et al., 2000) and be sensitive to intergroup bias (e.g., Adams Jr et al., 2010; Baumgartner et al., 2015; Bruneau et al., 2012; Cheon et al., 2011; Gamond et al., 2017; Park & Young, 2020), I did not find any intergroup bias or mentalizing reactivity in the *TPJ*. One potential explanation is that mentalizing in the smile discrimination task might be associated less with the *TPJ*. Similar to this finding, Lin et al. (2018) did not observe different activation in the *TPJ* between ingroup and outgroup emotion processing and mentalizing. Perhaps, the *TPJ* is more important in mentalizing tasks of false-belief reasoning (e.g., Saxe et al., 2004), but less so in those of emotion attribution (Zaitchik et al., 2010), such as smile discrimination, and instead the *mPFC* seems to be more responsible in the latter mentalizing processes (Ochsner et al., 2004). However, the current study did not include the *mPFC* because of practical reasons (e.g., a lack of optodes for full head coverage, interference with the Tobii eye-tracker which also uses near-infrared light as with fNIRS) and the spatial resolution of fNIRS. Future studies should explore the role of the *mPFC* in the intergroup bias on smile discrimination.

5.4.4 Mimicry – Facial Actions Effect

There was no difference in facial action or activation in the MNS between ingroup and outgroup and between ToM and Gender conditions. Accordingly, I did not find much evidence in support of mimicry in the smile discrimination task. This finding seems to support visual theories more than simulation theories. Mimicry is an important component in simulation theories but not in visual theories in understanding others' facial expressions. Participants performed significantly above the chance level without simulating the targets, which indicates mimicry was not necessary for understanding the hidden intentions of smiles at least in the smile discrimination task.

5.4.5 Advantages & Limitations of This Study

The use of a minimal group paradigm, where participants were randomly assigned to one of two mutually exclusive arbitrary groups, to induce intergroup bias in the current study allowed us to prevent multi-intergroup settings and focus on the well-controlled minimal intergroup modulation effects. Park and Young (2020) mentioned that existing group boundaries may provide plentiful information and stereotypes about the corresponding ingroup and outgroup members, which might not only represent intergroup bias per se but also indicate a Bayesian-rational reasoning through comparing and contrasting new information with the prior model of the target, especially when the two parties are in conflict (Hahn & Harris, 2014). Thus, a minimal group design becomes an elegant solution to circumvent this potential issue (Krautheim et al., 2019). One improvement over Chapter 4 is that I verified that ingroup favouritism can be induced by arbitrary labels in a laboratory rather than an online setting, which supports the validity and reliability of the minimal group induction.

Another advantage of the current study is the use of fNIRS. The superior motion tolerance for motion artefacts and a better balance between temporal and spatial resolution of fNIRS made it an ideal technique for employing a multimodal approach. A multimodal design allowed us to understand a certain cognitive function more comprehensively through a number of perspectives (e.g., behaviour, physiology, neuroscience, motion, eye movement) to explore the aims of the current study (Decety et al., 2018; Molenberghs & Louis, 2018).

On the other hand, fNIRS also has limitations. For example, although fNIRS data pre-processing and analysis pipelines have been developing rapidly, there is a lack of standardization (Pinti et al., 2019). It will take time for fNIRS to establish standard procedures and software as in other techniques (e.g., fMRI), which may lead to poor data quality and replication issues in the community at this time point. Furthermore, the current study focused on intergroup bias and ToM, so genuine and posed smile were mixed in each experimental block. As each trial was less than 6000ms with a 2500ms inter-trial interval, the Hemodynamic Response Function (HRF; peak about 4000-6000ms) measured by fNIRS is unlikely to allow me to distinguish the brain activation pattern for each of the two smile types. Indeed, previous research found that the neural activation patterns in response to genuine compared to posed facial emotional expressions are different (McLellan et al., 2012). It is possible that the distinct neural activation patterns might weaken the validity of the current study in detecting the neural correlates of intergroup bias. Future studies may look at genuine and posed smiles separately to further explore intergroup bias. Additionally, fNIRS can only detect neural activation of the cortical surface. But, as a rapidly developing technique, new methods may make it feasible to infer the subcortical and inferior cortical brain activity from the cortical activity using fNIRS (Liu et al., 2015).

I also acknowledge six additional limitations in the current study. First, the specific role of the neural activation in regions of multiple demand cortex might not be easily identified in the intergroup modulation effect (e.g., Smallwood et al., 2021). Previous studies have shown that no single brain region or network is exclusively responsible for intergroup bias (Lin et al., 2018; Molenberghs, 2013); rather intergroup bias may carefully enhance or attenuate the activation of the brain regions for the specific modality in the task. Moreover, most of the identified brain regions of interest in the current study are associated with more than one function, for example, when a ToM effect in the mentalizing, executive control and attentional salience networks (e.g., *dlPFC, IFG*) is observed, it could not be confirmed whether it indicates participants mentalized more, reallocated more attention, found the target more salient, or a combination of the three possibilities in that condition. Future research is needed to develop more advanced technologies, paradigms and analytical methods to tease apart each function that a brain area is involved in.

Second, the Gender condition might not be a “perfect” non-ToM control condition. It might not be entirely clear whether an observed effect is related to the target emotional mental state reasoning or the less relevant gender identification process (e.g., Molenberghs & Louis, 2018). To ensure sufficient power for each condition, I decided to not include additional control conditions. Future studies may compare smile discrimination with other control conditions to verify the current findings.

Third, as mentioned in Section 5.4.2.1, the Asian-dominant sample and the White-dominant actor pool might potentially interfere with the minimal group effect and bias the current findings, considering that previous studies have found evidence that intergroup bias and social mimicry vary between different racial groups and between different cultures (Molenberghs & Louis, 2018; Peng et al., 2021; Wei et al., 2013). Future research should counterbalance the race in both the recruited sample and the facial expression targets in stimuli.

Fourth, given the current study is exploratory, the neural mechanism of intergroup bias in smile discrimination is unknown in the literature. Thus, in order to explain the observed fNIRS results I made a few reverse inferences in Section 5.4, which means to infer the engagement of particular cognitive processes from specific patterns of brain activation (Poldrack, 2006, 2012). Although reverse inference has been particularly common in social cognitive neuroscience and can still provide plausible explanations (Machery, 2014; Poldrack, 2006, 2008), is not deductively valid and may lead to serious problems (Poldrack, 2006, 2012). Specifically, brain regions, like the *dlPFC* and *IFG*, are likely to be associated with multiple mental processes, so we cannot make a one-to-one mapping between a brain area and a particular cognitive function. We should always bear in mind that there are other explanations when we observe a particular pattern of brain activation. Future studies should investigate the neural mechanism of intergroup bias based on the findings in the current study to avoid reverse inference.

Fifth, based on the behavioural results, reaction time varied between conditions. This may reflect neural/cognitive processes that the current task cannot capture (e.g., the processes related to the button press and the short rest period after it), which could confound the particular processes we are interested in. Thus, the reaction time for all the answered trials was included in the fNIRS analyses as a regressor for each participant to model these processes, as mentioned in Section 5.2.6.5 and Section 5.3.1.2. Nevertheless, although I found no variance of brain activation can be explained by the variation of reaction time, there is a possibility that by removing the effect of reaction time we have also removed some brain activation relevant to our interested processes. Future studies should compare the models with and without reaction time to explore the processes relevant to it, and then decide whether to include it accordingly.

Sixth, as the current study was exploratory, there was no correction for multiple comparisons in the statistical fNIRS analyses, which may lead to erroneous inferences. Future replication studies should address this limitation with a bigger sample size.

5.5 Conclusion

To sum up, the current study makes novel contributions to the literature on the underpinning neural mechanisms of intergroup bias on mentalizing. By using a multimodal approach with fNIRS in a minimal group setting, the current study sheds light on how intergroup bias manifests at different levels, including behaviour, cognition, and neural responses. Although the behavioural findings revealed evidence of mentalizing as indicated by lower accuracy and longer reaction time, there was no indication of intergroup bias and facial mimicry in smile and gender identification. Mirroring the behavioural mentalizing effect, the fNIRS results also validate the involvement of the *IFG* and *pSTS* as part of the mentalizing network in facial emotional mental state reasoning indexed by enhanced activation in these regions in ToM conditions. However, this finding was only observed for outgroup members, which is likely to indicate an intergroup bias in ToM. Specifically, ingroup favouritism may enhance mentalizing for outgroup members in a seemingly deceptive social situation, for example, where participants were asked to identify posed smiles. Indeed, outgroup members tend to be judged to be more deceptive, sneaky and cunning (Dunham, 2018; Over, 2021). Thus, in the current context, outgroup smiles might be considered as something that is potentially socially deceptive. Accordingly, participants might work harder to process the outgroup. Importantly, intergroup bias was observed at the neural level and different brain regions seemed to respond to ingroup and outgroup information differently. Particularly, the *MFG* and *dlPFC* seem to be more sensitive to ingroup information, while the *IFG* seems to be more sensitive to outgroup information. That is, intergroup bias may possess opposing modulation effects on executive control and mentalizing during smile perception and differentiation. Additionally, neither behavioural assessments nor fNIRS results provided any social mimicry evidence, which supports visual theories rather than simulation theories of social cognition. This indicates people can understand others without mimicking them. These findings advance our understanding of the neural mechanisms underpinning the processing of intergroup bias in smile discrimination and have implications for understanding the complexity of the human brain in response to multiple higher-order cognitive modulations and how these modulations may interact with each other.

Chapter 6. General Discussion

The current thesis aims to detect mentalizing abilities in autism and investigate what factors can modulate mentalizing and how they modulate it. To achieve these aims, I implemented a multimodal approach to disentangle the complexity of measuring and modulating mentalizing and the underpinning neural mechanisms. Crucially, it establishes that social cognition difficulties in autism (difficulties in false-belief reasoning and smile discrimination) can be modulated and suggests potential factors that might mitigate such difficulties. During this process, I conducted four studies; each experiment involved processing social cues and decoding or reasoning about mental states in carefully controlled but relatively naturalistic settings compared with previous studies. Chapter 2 focused on overcoming some methodological difficulties seen in previous work and detecting mental state reasoning abilities tested with an implicit false-belief reasoning task in autistic adults. Chapter 3 investigated the modulation effect of evaluative contexts on mentalizing (false-belief reasoning) and the relationship between mentalizing abilities and individual differences in autistic traits, mental health outcomes and compensatory tendencies. Chapter 4 focused on mental state decoding ability tested with a smile discrimination task in autism and investigated the modulation effect of group membership on mentalizing. Chapter 5 used functional Near-Infrared Spectroscopy (fNIRS) to capture the underlying neural mechanisms of intergroup bias on mentalizing (smile discrimination). In this chapter, I first summarize the key results from each experimental chapter. Then, I discuss the implications of these findings in a broader context and outline the general conclusions. Finally, I point out some limitations and outstanding questions for future research in this field.

6.1 Summary of Experimental Chapters

The first step in exploring modulating mentalizing in autism is to overcome the methodological difficulties seen in previous work and probe mentalizing abilities in autism. In Chapter 2, to critically examine autistic people's ability to reason about false and true beliefs, I extended Southgate et al. (2007)'s paradigm through a multi-trial, multi-condition eye-tracking study with a more nuanced analysis of eye movements over the time course of each trial, as well as of changes in performance over the test session. Replicating the findings in Senju et al. (2009), I found that although many autistic individuals perform well in explicit mentalizing tasks, they do not engage in spontaneous false-belief reasoning in implicit tasks, consistent with their everyday social difficulties. These findings are consistent with the idea that some autistic people with average-to-high IQs may acquire the capacity to explicitly 'mentalize' about complex mental states (Frith, 2004), but still struggle to implicitly attribute simple mental states (Senju et al., 2009). I also found that despite the presence of spontaneous mentalizing difficulties, autistic adults showed a typical allocation of attentional resources to complex social stimuli, which indicates that autistic adults are capable of processing information from social cues in the same way as non-autistic adults but that this information is not then used to update alternative mental representations. Accordingly, I have been able to rule out some alternative theoretical explanations for this pattern of performance, such as, actions prediction difficulties, submentalizing processes, a true-belief bias or attentional differences, leading to a better understanding of mentalizing in both non-autistic and autistic people.

On the basis of Chapter 2, in Chapters 3 to 5, I further explored mentalizing abilities in the Broader Autism Phenotype (BAP) and the potential factors that can modulate mentalizing abilities to improve the social experience and life quality of autistic people, thereby helping them achieve their best potential. In Chapter 3, I investigated mentalizing abilities in mothers of autistic children and the modulation effect of the evaluability of the social context on implicit mentalizing. For this purpose, based on the modified implicit mentalizing paradigm in Chapter 2, I developed a more evaluative version by implementing a prompt question to assess the modulation effect of context in a non-autistic young adult sample. The results confirmed that the prompt task was better than the original non-prompted version in facilitating false-belief and true-belief reasoning, which indicates that a more evaluative context can indeed facilitate mentalizing (Woo et al., 2023), even in BAP populations. Then, I explored the relationship between implicit and explicit mentalizing abilities, autistic traits, compensatory tendencies and mental health outcomes in a bigger non-clinical sample. With the adapted prompt task, I found that both explicit mentalizing and autistic traits are associated with implicit mentalizing but not with each other. These findings support the idea of two distinct but overlapping mentalizing systems (Apperly & Butterfill, 2009) and implicit but not explicit mentalizing difficulties in autistic adults (Frith, 2004; Senju et al., 2009), consistent with what I found in Chapter 2. Given Broader Autism Phenotype (BAP) populations have been found to have similar social cognitive challenges to autistic people (Gliga et al., 2014; Green et al., 2019; Rea et al., 2019), but rarely receive any support, I also compared the aforementioned abilities and characteristics between mother of autistic (BAP) and non-autistic (non-BAP) children. Unexpectedly, BAP mothers showed better implicit mentalizing and poorer mental health than non-BAP mothers, but no other differences, which may indicate the heterogeneity within the BAP (Bora et al., 2017; Rubenstein & Chawla, 2018) as well as the need to support families with autistic members in terms of mental health and psychological resilience (Bitsika et al., 2013).

In Chapter 4, I further looked at the modulation effects of intergroup bias on emotional mental state decoding in a different task – the perception of genuine and posed smiles among autistic adults. The smile discrimination task was selected to measure mentalizing because the social-emotional ambiguity of posed expressions has been suggested to engage mentalizing to a greater degree (e.g., Cosme et al., 2021; Lavan et al., 2017; McGettigan et al., 2015; Szameitat et al., 2010). Participants were asked to watch videos of people making genuine or posed smiles and rate the authenticity of each smile. To focus on a well-controlled intergroup modulation effect, a minimal group paradigm was adopted (Hahn & Harris, 2014; Park & Young, 2020). Participants were informed (falsely) that some of the smilers were from an in-group and others were from an out-group. I found that autistic adults showed reduced sensitivity to the different smile types and were less likely to identify with ingroup members, consistent with the literature that autistic people have difficulties in mentalizing. Notably, I also found that group membership did affect authenticity judgements similarly in autistic and non-autistic adults (i.e., the main effect of group membership) but did not modulate the ability to differentiate genuine from posed smiles in either diagnostic group (i.e., no interaction between smile type and group membership). I propose that this might be due to reduced identification with, empathy for or trust in unfamiliar or diagnostic outgroup members, in combination with mentalizing or social attention differences. Accordingly, I suggest a reconsideration of past findings that might have misrepresented the social judgements of autistic people via introducing an outgroup disadvantage. As autistic people perceive ingroup members to be more authentic, this is likely to give rise to more rewarding and more comfortable interactions. This finding has implications for designing tailored support and policies that emphasize similarities and inclusion between autistic and non-autistic people to avoid intergroup conflicts (Mitchell et al., 2021), rather than focusing on how they might be different (Baron-Cohen, 2017). This might facilitate autistic people in navigating the social world more effectively and make society more inclusive.

In Chapter 5, I sought to extend the findings in Chapter 4 to further test the modulation effect of intergroup bias on social ability with fNIRS too. I primarily aimed to investigate the underlying neural mechanisms involved in this process by using a multimodal approach with fNIRS in a minimal group setting. I secondarily aimed to explore the role of mimicry in social cognition, as well as the effect of intergroup bias on mimicry during smile discrimination. I thirdly aimed to conduct a closer replication that directly measured smile discrimination. I found that intergroup bias in smile identification was not observed at the behavioural level, but was observed at the neural level. Specifically, the *MFG* and *dlPFC* seem to be more sensitive to ingroup smiles, while the *IFG* seems to be more sensitive to outgroup smiles. Accordingly, the same behaviour but different brain activity for the ingroup and outgroup contrast seems to indicate that the cognitive processes may be different for ingroup and outgroup members even though people manage to achieve the same level of behaviour. However, we do not know yet what is different about the cognition for ingroup and outgroup; future studies are needed to investigate what the differences are and what they mean. The behavioural results also revealed evidence for successful mentalizing via smile discrimination. Mirroring the behavioural mentalizing effect, the fNIRS results showed enhanced activation in the mentalizing network, covering the *IFG* and *pSTS*, during facial emotional mental state reasoning. However, this was only found for outgroup members, which is likely to imply an intergroup bias in mentalizing. Particularly, intergroup bias may enhance mentalizing for outgroup members in suspicious social situations (e.g., when identifying posed smiles). In addition, neither behavioural assessments nor fNIRS results provided any evidence of social mimicry in smile identification. This finding seems to support visual theories of social cognition indicating that mentalizing (tested via a smile discrimination task) does not require mimicry.

6.2 Implications

Taken together, the current thesis contributes to a better understanding of social cognition in autism and the importance of contextual information and individual differences in social cognition. According to the findings across the reported studies, I was able to draw three main conclusions. First, autistic adults have difficulties in mentalizing, including implicit mental state reasoning and emotional mental state decoding. Second, autism is a spectrum condition and highly heterogeneous, which is related to mental health issues. Third, mentalizing can be facilitated by contextual factors to a certain degree. Each of the three points is discussed in this section.

6.2.1 Social Cognition in Autism

6.2.1.1 Mentalizing in Autism

The current thesis firstly concludes that autistic adults indeed have difficulties in mentalizing, regarding implicit mental state reasoning and mental state decoding. In line with previous work (e.g., Senju et al., 2009), I demonstrated a dissociation between implicit and explicit mentalizing performance in the autism group in Chapter 2 – autistic adults were less accurate in implicit false belief reasoning but were indistinguishable from non-autistic adults in explicit mentalistic reasoning. Through comparing the true-belief control condition with the false-belief experimental condition and detailed analysis of gaze patterns, these results cannot be explained by submentalizing, attentional differences or a true-belief bias. These findings seem to confirm the idea that some autistic individuals with average-to-high IQs may acquire the capacity to explicitly ‘mentalize’ about complex mental states (Frith, 2004), but still struggle to implicitly attribute simple mental states (Senju et al., 2009).

Nevertheless, I might not be able to fully understand autistic people's genuine explicit mentalizing ability and even the existence of a later-developing explicit mentalizing system for three main reasons. First, explicit mentalizing has been suggested to be related to many other cognitive abilities, such as language (e.g., Happé, 1995), executive functions (e.g., Carlson et al., 2002; Jones et al., 2018) and memory (e.g., Ullman & Pullman, 2015). In Chapters 2 and 3, I used the Strange Stories Task to assess explicit mentalizing in which participants are asked to read short vignettes describing social scenarios and then explain a character's behaviour based on their mental states. Since the vignettes are presented in text and participants should answer the questions without looking at the text, this task may also tax other cognitive abilities, like language and working memory. Therefore, it is unclear what important roles are played by these abilities in explicit mentalizing tasks. Second, in Chapter 3, explicit, but not implicit, mentalizing performance was associated with camouflaging behaviours that autistic people use to compensate for their social difficulties or mask their autistic characteristics; it might be that, unlike implicit mentalizing, explicit mentalizing difficulties can be compensated by alternative cognitive strategies. At the extreme, autistic adults might be able to pass explicit mentalizing tasks by adopting other cognitive abilities and compensatory strategies, rather than using the proposed explicit mentalizing ability. Third, I did not find any correlation between autistic traits and explicit mentalizing among autistic, BAP and non-autistic populations (Chapters 2 & 3). Together with the first two points, it is hard to tell whether this means there is no relationship between explicit mentalizing and autism or explicit mentalizing cannot be accurately detected by the task. Future studies should investigate what factors can influence or actually constitute the so-called explicit mentalizing and its manifestation in autistic people.

In Chapter 4, I used a smile judgment task to assess mentalizing. Distinguishing genuine from posed smiles involves assessing the mental state of the actor and is a relatively simple task which could be implemented online during COVID-19. Using dynamic stimuli, I found that while autistic people are capable of discriminating genuine from posed smiles, performance was worse than non-autistic adults. This indicates that autistic people do have difficulties in identifying subtle facial emotional expressions that likely rely on mentalizing (Blampied et al., 2010; Boraston et al., 2008; Harms et al., 2010; Liu & Humpolíček, 2013).

In the past 10 years, especially after COVID-19, there has been rapid growth in online data collection to recruit larger and more diverse samples that would be difficult to access in laboratory studies (Anwyl-Irvine et al., 2020; Tsantani et al., 2022; Türközer & Öngür, 2020). For autism research, this approach is not only convenient, time-saving and low-cost but also advantageous for the inclusion of autistic people who might not be able to participate in laboratory experiments. Although Chapters 2 and 3 showed that the false-belief reasoning tasks we developed based on Southgate et al. (2007)'s anticipatory-looking paradigm are valid in facilitating mentalizing, it is still difficult to do them online due to the poor quality of eye movement data. Notably, Chapter 4 documented that the smile task is feasible for online testing with autistic people. I promote that the smile discrimination task has the potential to be built online as a valuable task for assessing mentalizing in future studies. As mentioned earlier, Chapter 4 also showed that autistic people performed worse than non-autistic people in the smile task, which indicates that this task is sensitive in detecting autistic people's difficulties in social cognition, here specifically in smile discrimination. Therefore, the smile task could complement the aforementioned anticipatory-looking task and traditional mentalizing tasks in autism research. Given the smile task does not contain complicated instructions and uses dynamic stimuli of smiling people, this task holds promise as being adaptable to a much wider range of populations (e.g., children, people with language or intellectual difficulties) and possesses higher ecological validity than traditional mentalizing tests.

6.2.1.2 Autistic Traits, BAP, and Mental Health

Our second conclusion is that autism is a spectrum condition and highly heterogeneous. This is not only because autism affects people in different ways, but also because autistic traits appear to varying degrees in the general population and genetically predisposed populations. I found in Chapter 3 that autistic traits were associated with implicit mentalizing ability. This indicates that higher autistic traits can be a sign of weaker social cognition, especially implicit mentalizing, in non-autistic people. This is consistent with Nijhof et al. (2017)'s finding. The absence of a relationship between autistic traits and explicit mentalizing as mentioned earlier might seem surprising. This could be because people with higher autistic traits do not possess explicit mentalizing difficulties. Alternatively, as mentioned in Section 6.2.1.1, explicit mentalizing difficulties can be compensated and passing the Strange Stories task also requires adequate other cognitive abilities, like executive function, hence it is unclear if this result truly reflects the relationship between mentalizing and autistic traits in non-autistic populations.

Parents of autistic children are believed to possess an underlying genetic liability for autism (Sasson et al., 2013), for example, the shared genetic overlap between BAP mothers and their autistic children has been observed to be associated with the mothers' autistic traits (Nayar et al., 2021). However, our BAP mothers not only performed better in the implicit and comparably in the explicit mentalizing tasks but also did not report more autistic traits compared with non-BAP mothers. This is the opposite of what Gliga et al. (2014) found with infant BAP siblings. One potential explanation is that, unlike many infant siblings, it is possible that by chance our BAP mothers are not genetically predisposed to autism themselves and hence do not contribute to their child's genetic predisposition. Although BAP is nearly ten times more prevalent in first-degree relatives than in the general population

(Green et al., 2019) and highly heritable (An et al., 2021; Freitag et al., 2010; Hill & Frith, 2003), autistic traits are tremendously heterogeneous in BAP populations. Our BAP mothers might not be representative of the wider BAP mother population and were totally unaffected at the behavioural, cognitive and neurological level. Alternatively, it might be that the interaction between protective factors and autistic advantages boosted BAP mothers' implicit mentalizing performance. Considering the fact that females require more inherited factors than males to exhibit autistic traits (Lockwood Estrin et al., 2021), our BAP mothers might possess some protective factors that made them display fewer autistic traits than their children. However, they might reserve some autism-like cognitive styles for example, a detail-focused cognitive style (Happé & Vital, 2009), that predispose them to better develop certain cognitive abilities than non-BAP mothers (Happé & Vital, 2009).

I also found that mental health issues are related to autistic traits and the BAP but maybe for different reasons. In Chapter 3, both autistic and non-autistic participants who reported more mental health problems also self-identified with higher autistic traits and compensatory tendencies. These findings are consistent with the extant literature that individuals with more socio-cognitive difficulties (Baron-Cohen, Wheelwright, Skinner, et al., 2001; Green et al., 2019; Hurley et al., 2007) need to allocate more cognitive resources to compensate for their core difficulties, which is likely to compromise their mental health (Hull et al., 2017; Lai et al., 2011; Lai et al., 2017; Livingston et al., 2019; Livingston & Happé, 2017). These results also support the idea that the mental health difficulties in BAP mothers might be more related to their chronic stress from parenting and caring for autistic children (Bishop et al., 2007; Bitsika et al., 2013; Ekas et al., 2010; Su et al., 2018) than their own autistic traits (Bolton et al., 1998; Ingersoll & Hambrick, 2011; Ingersoll, Meyer, et al., 2011; Micali et al., 2004; Pruitt et al., 2018; Sucksmith et al., 2011) or the cost of compensation (Livingston & Happé, 2017). However, the current study cannot rule out a multi-risk model of mental health outcomes in BAP mothers, as the BAP is highly heterogeneous in relatives of autistic people (Bora et al., 2017; Rubenstein & Chawla, 2018). On all accounts, support is needed to alleviate mental health issues and develop psychological resilience in people with higher autistic traits and BAP populations (Bitsika et al., 2013).

6.2.2 What Factors Can Modulate Mentalizing?

While many early studies examined mentalising as an ability that is either present or absent, here I consider how the tendency to mentalise might be modulated by other social factors. I find that the tendency to mentalize is facilitated by evaluative context (Chapter 3) and by group membership (Chapters 4 & 5).

6.2.2.1 Evaluative Context

Following Woo et al. (2023)'s theoretical idea, I developed a more evaluative anticipatory-looking paradigm in which a question was added to prompt participants to anticipate the actor's actions, which increases the interactive potential of the actor, and therefore gives participants more reasons for mentalizing. I provided the first empirical evidence showing that more socially evaluative contexts can better facilitate mentalizing than less evaluative contexts. Meanwhile, I found in Chapter 3 that both non-autistic adults and BAP mothers (i.e., mothers of autistic children) failed to show mentalizing in the original Southgate et al. (2007) anticipatory-looking paradigm, which suggests that the context of this non-prompt version might not be sufficiently evaluative to elicit implicit mentalizing (Kulke & Hinrichs, 2021; Kulke, Johannsen, et al., 2019; Schuwerk et al., 2018; Woo et al., 2023). In the original version, observers know the actor's actions have no interactive potential, thus reasoning about her mental state is unlikely to be prioritized. Therefore, the mixed results in replications in the literature using this paradigm and the inconsistent results I found between

Chapter 2 and Chapter 3 with the non-prompt mentalizing tasks may be due to their non-evaluative contexts that provide observers less reason to care about agents' mental states. Apart from prompting mentalizing by requiring action prediction, there are many ways to improve context evaluability. As I only looked at implicit false-belief reasoning ability in non-autistic populations, future studies should develop more evaluative paradigms and examine the effectiveness in facilitating mentalizing with different levels of evaluative context in other social cognitive abilities in autistic people.

6.2.2.2 Intergroup Bias

Chapter 4 showed that ingroup favouritism can modulate emotional mental state decoding to a certain degree in both autistic and non-autistic populations. To avoid confounding intergroup differences, I examined smile authenticity perception in a minimal group setting. Although autistic adults subjectively reported less identification towards their ingroup than non-autistic adults, they were equally sensitive to this ingroup favouritism – ingroup members were perceived as more authentic. I found both autistic and non-autistic people judged ingroup smiles as more genuine than outgroup smiles (Chapter 4). However, the rating difference between genuine and posed smiles did not differ between ingroup and outgroup members (Chapter 4) and ingroup smiles were not more accurately identified than outgroup smiles in non-autistic adults (Chapter 5). These findings might indicate that intergroup bias is likely to facilitate positive judgement towards ingroup smiles, but it did not facilitate the ability to discriminate between genuine and posed smiles. In other words, intergroup bias potentially makes people perceive ingroup members more positively than outgroup members, but it would not change people's abilities in social cognition. This is consistent with some previous studies in the literature (e.g., Adams Jr et al., 2010; Young, 2017). Although intergroup bias might not help improve the social abilities of autistic people;

it could potentially enhance their social experience and life quality because interaction with ingroup members can be rewarding (Shore & Heerey, 2011) and enjoyable (Krumhuber et al., 2007). This idea seems to be in line with Milton (2012)'s 'double empathy problem' which suggests that both autistic and non-autistic people are better at understanding the members of their own diagnostic group than the members of the other group, as mentioned in Chapter 1. It would be interesting in future to more directly consider how intergroup bias relates to the 'double empathy problem'. Both concepts propose that if we draw people's attention more to similarities between diverse people, we can hope it might make society more inclusive.

6.3 General Limitations & Future Directions

This thesis aims to detect mentalizing abilities in autism (Chapters 2 and 3) and to investigate how and why social context and individual differences modulate mentalizing (Chapters 3, 4, and 5). These findings contribute to advancing the current understanding of modulating mentalizing in autism, which further raises essential questions in social cognition and social neuroscience. Future studies are needed to address these outstanding questions and replicate the findings reported in the current thesis.

First of all, the study of social cognition during social observation might not generalize to social interaction. In the current thesis, experimental designs were used to study social cognition while participants observed and judged the carefully controlled and manipulated social cues, and the underlying cognitive and neural mechanisms were inferred from the detected participants' behaviours and neural activities. However, this single-person approach might lack a feeling of engaging with a social partner, and therefore measure social cognition during social observation (aka. a third-person perspective), which is fundamentally different from that during social interaction (Schilbach et al., 2013). This difference exists not only at the cognitive level but also at the behavioural and neural levels (Schilbach et al., 2013). Based on this assumption, second-person neuroscience has been proposed and suggested that to study the underlying neural correlates of social interaction it is necessary to use paradigms involving real-time social interaction and/or feel engaged with a social partner, which also apply to the corresponding behavioural and cognitive mechanisms. This could be particularly relevant to the lack of mimicry in Chapter 5. I tried to set up an experiment where participants were expected to spontaneously mimic, however, there was no evidence for mimicry at both the behavioural and neural levels. This might be because I used a single-person approach in which participants were facing a screen. If the task was done more interactively, we might find more spontaneous mimicry of smiles. Thus, future studies should prioritize the ecological validity of social cognition during social interaction assessments by introducing naturalistic social encounters and truly interactive settings. The emergence of second-person neuroscience enables theories to transcend single-brain models and encompass the reciprocal influence among diverse social agents (Konvalinka & Roepstorff, 2012; Redcay & Schilbach, 2019; Schilbach et al., 2013).

In addition, people may not be as ‘accurate’ as they think in reporting individual differences, such as BAP traits, autistic traits, empathetic concern, compensation and mental health. Self-reported inventories for assessing individual differences might measure the awareness or the perceived social expectations of these characteristics instead of genuine individual differences (Scheeren & Stauder, 2008). For example, I recruited a great number of people who self-identified as autistic. They were indeed above the cut-off line of the AQ, however, they did not receive an autism diagnosis. Perhaps, this was because of delayed diagnosis, but it is also possible that they did not meet the diagnostic threshold. This does not mean they were not being honest; that could be their genuine understanding of themselves, but might not be comparable across people. Thus, more objective measures, such as behavioural tasks and eye-tracking and neuroimaging techniques, are needed in future populational studies (Hurley et al., 2007; Livingston et al., 2019; Lord et al., 2012; Pruitt et al., 2018).

Moreover, it is important to understand the development of mentalizing and the acquisition of intergroup bias, especially in clinical populations, like autistic people showing mentalizing difficulties (Baron-Cohen et al., 1985) but moderate sensitivity to context information (Hadad et al., 2019; Kang et al., 2020; Wilson et al., 2011; Yi et al., 2016; Yi et al., 2015), which have not been fully revealed. However, I employed a cross-sectional design in all the studies, so the direction of the observed effects and relationships cannot be conclusively determined. Future research should incorporate a longitudinal design to confirm the causality of these relationships.

An obvious limitation to Chapters 2 and 3 as well as most of the autism literature in the scope of high-level social cognition is that all participants had average-to-high IQs, thus these findings cannot be generalized to autistic adults with language delay and/or intellectual disability, or young autistic children. Fortunately, the paradigms I used, like the prompted implicit mentalizing tasks and the smile discrimination tasks hold promise as being adaptable to a much wider range of individuals. Future studies should further adapt the paradigms to study mentalizing in autism and potential modulatory factors with participants possessing different levels of general abilities.

Chapters 4 and 5 provide important insights into the differences between social judgement and social ability and the experimental stimuli used to study them. I show that both autistic and non-autistic adults rated ingroup smiles as less authentic than outgroup smiles (social judgment) but they were still able to differentiate between genuine and posed smiles (social ability). This observation raised a vital notion that social ability is not necessarily equivalent to social judgment. Social judgement, but not social ability, can be modulated by contextual information in both autistic and non-autistic populations. Future studies should make clear whether social judgement and/or social ability are being assessed and/or modulated.

Also, since Chapter 5 was an exploratory study, there could be two limitations. First, to explain the fNIRS data, I made a few reverse inferences which might not be valid and may lead to problems (Poldrack, 2006, 2012). Thus, I cannot make any one-to-one mapping to conclude any particular cognitive functions were involved based on the fNIRS results. Future studies should always bear in mind that there are other explanations for neuroimaging findings and try to avoid reverse inference. Second, there was no correction for multiple comparisons in the fNIRS data analyses, which likely caused erroneous inferences. Future studies should address this problem with a bigger sample size.

Last but not least, experimental materials may bias the response from autistic people. Given autistic people showed sensitivity to intergroup bias, they could presumably also be influenced by their diagnostic group. This is consistent with the ‘double empathy problem’ suggested by Milton (2012). Accordingly, autistic adults may assume in the absence of evidence to the contrary that all the videos contained people from the non-autistic majority, thus their autism diagnostic-group identification could account for the generally lower ratings they made than those made by non-autistic people. Indeed, the idea regarding diagnostic-ingroup favouritism has been partially supported in the literature that non-autistic people rated autistic people less favourably than other non-autistic people without knowing who was autistic (Alkhaldi et al., 2019; Sasson et al., 2017). Therefore, the current thesis and many studies in the literature might have failed to fairly assess social judgements in autism which presumably used non-autistic protagonists (Gernsbacher et al., 2017). This might suggest a need to re-evaluate past findings of social perception in autism and consider whether any of those studies might have misrepresented the social judgements of autistic people by introducing an outgroup disadvantage. Future studies could test this possibility directly by including both autistic and non-autistic protagonists, as well as investigate how intergroup bias relates to the ‘double empathy problem’.

6.4 Concluding Remarks

In conclusion, the current thesis investigated mentalizing abilities in autism and the modulation effect of context information on mentalizing. It implemented a multimodal approach, involving cognitive, behavioural and neural measures, to disentangle the complexity of measuring and modulating mentalizing and the underpinning neural mechanisms. It included a comparison between implicit and explicit mentalizing and an investigation of both mental state decoding and reasoning. Findings revealed mentalizing difficulties in autism, including both mental state decoding and reasoning. For mental state reasoning specifically, autistic adults only showed difficulties in the implicit pathway, while the explicit one seemed to be intact or its difficulties were compensated by additional strategies. Additionally, the current thesis shows that social context information can facilitate and modulate mentalizing to a certain extent in some circumstances. It provides innovative insights into the intricate interplay between behaviour and neural activities in modulating mentalizing, advocating for investigating mentalizing within contexts characterized by diverse evaluability and intergroup dynamics. This has implications for designing tailored support and policies that emphasize similarities and transparency between autistic and non-autistic people, which may improve the social experience and life quality of autistic people and make society more inclusive.

References

Abell, P. (2000). Putting social theory right? *Sociological Theory*, *18*(3), 518-523.

Adams Jr, R. B., Rule, N. O., Franklin Jr, R. G., Wang, E., Stevenson, M. T., Yoshikawa, S., Nomura, M., Sato, W., Kveraga, K., & Ambady, N. (2010). Cross-cultural reading the mind in the eyes: an fMRI investigation. *Journal of Cognitive Neuroscience*, *22*(1), 97-108.

Alkhaldi, R. S., Sheppard, E., & Mitchell, P. (2019). Is there a link between autistic people being perceived unfavorably and having a mind that is difficult to read? *Journal of autism and developmental disorders*, *49*(10), 3973-3982.

Allen, V. L., & Wilder, D. A. (1975). Categorization, belief similarity, and intergroup discrimination. *Journal of personality and social psychology*, *32*(6), 971.

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief “red flags” for autism screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(2), 202-212. e207.

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in cognitive sciences*, *4*(7), 267-278.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)* (5th ed.). American Psychiatric Pub.

Ames, D. R. (2004). Strategies for social inference: a similarity contingency model of projection and stereotyping in attribute prevalence estimates. *Journal of personality and social psychology*, *87*(5), 573.

Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience*, *15*(10), 670-682.

An, K.-m., Ikeda, T., Hirosawa, T., Yaoi, K., Yoshimura, Y., Hasegawa, C., Tanaka, S., Saito, D. N., & Kikuchi, M. (2021). Decreased grey matter volumes in unaffected mothers of individuals with autism spectrum disorder reflect the broader autism endophenotype. *Scientific reports, 11*(1), 10001.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods*, *52*, 388-407.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological review*, *116*(4), 953.

Arioli, M., Cattaneo, Z., Ricciardi, E., & Canessa, N. (2021). Overlapping and specific neural correlates for empathizing, affective mentalizing, and cognitive mentalizing: A coordinate‐based meta‐analytic study. *Human brain mapping*, *42*(14), 4777-4804.

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. *Journal of psychosomatic research*, *38*(1), 23-32.

Baio, J. (2014). *Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010*. Atlanta, GA: Centers for Disease Control and Prevention (CDC) Retrieved from https://stacks.cdc.gov/view/cdc/22182#

Balliet, D., Wu, J., & De Dreu, C. K. (2014). Ingroup favoritism in cooperation: a meta-analysis. *Psychological Bulletin*, *140*(6), 1556.

Baltrušaitis, T., Mahmoud, M., & Robinson, P. (2015). *Cross-dataset learning and person-specific normalisation for automatic action unit detection* 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia.

Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2013). Constrained local neural fields for robust facial landmark detection in the wild. Proceedings of the IEEE international conference on computer vision workshops,

Baltrušaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). *Openface 2.0: Facial behavior analysis toolkit* 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018),

Bargiela, S., Steward, R., & Mandy, W. (2016). The experiences of late-diagnosed women with autism spectrum conditions: An investigation of the female autism phenotype. *Journal of autism and developmental disorders*, *46*(10), 3281-3294.

Baron-Cohen, S. (2017). Editorial Perspective: Neurodiversity–a revolutionary concept for autism and psychiatry. In (Vol. 58, pp. 744-747): Wiley Online Library.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37-46.

Baron-Cohen, S., O'riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high-functioning autism. *Journal of autism and developmental disorders*, *29*(5), 407-418.

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.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241-251.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of autism and developmental disorders*, *31*(1), 5-17.

Bartholow, B. D., & Henry, E. A. (2010). Response conflict and affective responses in the control and expression of race bias. *Social and personality psychology compass*, *4*(10), 871-888.

Bastiaansen, J. A., Thioux, M., & Keysers, C. (2009). Evidence for mirror systems in emotions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2391-2404.

Baumgartner, T., Nash, K., Hill, C., & Knoch, D. (2015). Neuroanatomy of intergroup bias: A white matter microstructure study of individual differences. *NeuroImage*, *122*, 345-354.

Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical psychology review*, *8*(1), 77-100.

Begeer, S., Mandell, D., Wijnker-Holmes, B., Venderbosch, S., Rem, D., Stekelenburg, F., & Koot, H. M. (2013). Sex differences in the timing of identification among children and adults with autism spectrum disorders. *Journal of autism and developmental disorders*, *43*(5), 1151-1156.

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science*, *18*(8), 706-712.

Bertschy, K., Skorich, D. P., & Haslam, S. A. (2020). Self-categorization and Autism: Exploring the relationship between autistic traits and ingroup favouritism in the minimal group paradigm. *Journal of autism and developmental disorders*, *50*(9), 3296-3311.

Biland, C., Py, J., Allione, J., Demarchi, S., & Abric, J.-C. (2008). The effect of lying on intentional versus unintentional facial expressions. *European Review of Applied Psychology*, *58*(2), 65-73.

Billeci, L., Calderoni, S., Conti, E., Gesi, C., Carmassi, C., Dell'Osso, L., Cioni, G., Muratori, F., & Guzzetta, A. (2016). The broad autism (endo) phenotype: neurostructural and neurofunctional correlates in parents of individuals with autism spectrum disorders. *Frontiers in neuroscience*, *10*, 346.

Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain*, *133*(5), 1515-1525.

Bird, G., & Viding, E. (2014). The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience & Biobehavioral Reviews*, *47*, 520-532.

Bishop, D. V., Maybery, M., Maley, A., Wong, D., Hill, W., & Hallmayer, J. (2004). Using self‐report to identify the broad phenotype in parents of children with autistic spectrum disorders: a study using the Autism‐Spectrum Quotient. *Journal of Child Psychology and Psychiatry*, *45*(8), 1431-1436.

Bishop, S. L., Richler, J., Cain, A. C., & Lord, C. (2007). Predictors of perceived negative impact in mothers of children with autism spectrum disorder. *American Journal on Mental Retardation*, *112*(6), 450-461.

Bitsika, V., Sharpley, C. F., & Bell, R. (2013). The buffering effect of resilience upon stress, anxiety and depression in parents of a child with an autism spectrum disorder. *Journal of Developmental and Physical Disabilities*, *25*(5), 533-543.

Blairy, S., Herrera, P., & Hess, U. (1999). Mimicry and the judgment of emotional facial expressions. *Journal of nonverbal behavior, 23*, 5-41.

Blakemore, S.-J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience, 9*(4), 267-277.

Blampied, M., Johnston, L., Miles, L., & Liberty, K. (2010). Sensitivity to differences between enjoyment and non‐enjoyment smiles in children with autism spectrum disorder. *British journal of developmental psychology*, *28*(2), 483-489.

Bolton, P. F., Pickles, A., Murphy, M., & Rutter, M. (1998). Autism, affective and other psychiatric disorders: patterns of familial aggregation. *Psychological medicine*, *28*(2), 385-395.

Bora, E., Aydın, A., Saraç, T., Kadak, M. T., & Köse, S. (2017). Heterogeneity of subclinical autistic traits among parents of children with autism spectrum disorder: Identifying the broader autism phenotype with a data‐driven method. *Autism Research*, *10*(2), 321-326.

Boraston, Z. L., Corden, B., Miles, L. K., Skuse, D. H., & Blakemore, S.-J. (2008). Brief report: Perception of genuine and posed smiles by individuals with autism. *Journal of autism and developmental disorders*, *38*(3), 574-580.

Borg, E. (2007). If mirror neurons are the answer, what was the question? *Journal of Consciousness Studies*, *14*(8), 5-19.

Borgomaneri, S., Bolloni, C., Sessa, P., & Avenanti, A. (2020). Blocking facial mimicry affects recognition of facial and body expressions. *PloS one*, *15*(2), e0229364.

Borinca, I., Van Assche, J., Gronfeldt, B., Sainz, M., Anderson, J., & Taşbaş, E. H. O. (2023). Dehumanization of outgroup members and cross-group interactions. *Current Opinion in Behavioral Sciences*, *50*, 101247.

Bourgeois, P., & Hess, U. (2008). The impact of social context on mimicry. *Biological psychology, 77*(3), 343-352.

Bowler, D. M. (1992). “Theory of Mind” in Asperger's Syndrome Dermot M. Bowler. *Journal of Child Psychology and Psychiatry, 33*(5), 877-893.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision*, *10*(4), 433-436.

Brass, M., & Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence problem? *Trends in cognitive sciences*, *9*(10), 489-495.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of social issues*, *55*(3), 429-444.

Broderick, N., Wade, J. L., Meyer, J. P., Hull, M., & Reeve, R. E. (2015). Model invariance across genders of the broad autism phenotype questionnaire. *Journal of autism and developmental disorders*, *45,* 3133-3147.

Brugha, T., McManus, S., Meltzer, H., Smith, J., Scott, F., Purdon, S., Harris, J., & Bankart, J. a. (2009). Autism spectrum disorders in adults living in households throughout England: Report from the adult psychiatric morbidity survey 2007. *Leeds: The NHS Information Centre for Health and Social Care*.

Brüne, M., & Brüne-Cohrs, U. (2006). Theory of mind—evolution, ontogeny, brain mechanisms and psychopathology. *Neuroscience & Biobehavioral Reviews, 30*(4), 437-455.

Bruneau, E. G., Dufour, N., & Saxe, R. (2012). Social cognition in members of conflict groups: behavioural and neural responses in Arabs, Israelis and South Americans to each other's misfortunes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1589), 717-730.

Brunet, E., Sarfati, Y., Hardy-Baylé, M.-C., & Decety, J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. *NeuroImage, 11*(2), 157-166.

Burnside, K., Ruel, A., Azar, N., & Poulin-Dubois, D. (2018). Implicit false belief across the lifespan: Non-replication of an anticipatory looking task. *Cognitive development, 46*, 4-11.

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development: An International Journal of Research and Practice, 11*(2), 73-92.

Carpita, B., Carmassi, C., Calderoni, S., Muti, D., Muscarella, A., Massimetti, G., Cremone, I. M., Gesi, C., Conti, E., & Muratori, F. (2020). The broad autism phenotype in real-life: clinical and functional correlates of autism spectrum symptoms and rumination among parents of patients with autism spectrum disorder. *CNS spectrums, 25*(6), 765-773.

Castelli, F. (2005). Understanding emotions from standardized facial expressions in autism and normal development. *Autism*, *9*(4), 428-449.

Chartrand, T. L., & Van Baaren, R. (2009). Human mimicry. *Advances in experimental social psychology, 41*, 219-274.

Cheon, B. K., Im, D.-m., Harada, T., Kim, J.-S., Mathur, V. A., Scimeca, J. M., Parrish, T. B., Park, H. W., & Chiao, J. Y. (2011). Cultural influences on neural basis of intergroup empathy. *NeuroImage*, *57*(2), 642-650.

Chierchia, G., Fuhrmann, D., Knoll, L. J., Pi-Sunyer, B. P., Sakhardande, A. L., & Blakemore, S.-J. (2019). The matrix reasoning item bank (MaRs-IB): novel, open-access abstract reasoning items for adolescents and adults. *Royal Society open science*, *6*(10), 190232.

Chown, N. (2014). More on the ontological status of autism and double empathy. *Disability & Society*, *29*(10), 1672-1676.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive development*, *9*(4), 377-395.

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, not autism, predicts poor recognition of emotional facial expressions. *Psychological Science*, *24*(5), 723-732.

Cosme, G., Rosa, P. J., Lima, C. F., Tavares, V., Scott, S., Chen, S., Wilcockson, T. D., Crawford, T. J., & Prata, D. (2021). Pupil dilation reflects the authenticity of received nonverbal vocalizations. *Scientific reports*, *11*(1), 3733.

Csibra, G. (2008). Action mirroring and action understanding: An alternative account. *Sensorymotor foundations of higher cognition. Attention and performance XXII*, 435-459.

Csibra, G., & Gergely, G. (2007). 'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta psychologica*, *124*(1), 60-78.

Cui, X., Bray, S., & Reiss, A. L. (2010). Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *NeuroImage*, *49*(4), 3039-3046.

Dal Monte, O., Schintu, S., Pardini, M., Berti, A., Wassermann, E. M., Grafman, J., & Krueger, F. (2014). The left inferior frontal gyrus is crucial for reading the mind in the eyes: brain lesion evidence. *cortex*, *58*, 9-17.

Dang, J., King, K. M., & Inzlicht, M. (2020). Why Are Self-Report and Behavioral Measures Weakly Correlated? *Trends in cognitive sciences*, *24*(4), 267-269.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental psychology*, *40*(2), 271.

de Bildt, A., Sytema, S., Meffert, H., & Bastiaansen, J. A. (2016). The Autism Diagnostic Observation Schedule, Module 4: Application of the revised algorithms in an independent, well-defined, Dutch sample (n= 93). *Journal of autism and developmental disorders*, *46*(1), 21-30.

De la Marche, W., Noens, I., Kuppens, S., Spilt, J. L., Boets, B., & Steyaert, J. (2015). Measuring quantitative autism traits in families: informant effect or intergenerational transmission? *European child & adolescent psychiatry*, *24*(4), 385-395.

De Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why? *Trends in cognitive sciences*, *10*(10), 435-441.

Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *The Neuroscientist*, *13*(6), 580-593.

Decety, J., Pape, R., & Workman, C. I. (2018). A multilevel social neuroscience perspective on radicalization and terrorism. *Social neuroscience*, *13*(5), 511-529.

DeMyer, M. K. (1979). *Parents and children in autism*. VH Winston.

Deschrijver, E., Bardi, L., Wiersema, J. R., & Brass, M. (2016). Behavioral measures of implicit theory of mind in adults with high functioning autism. *Cognitive Neuroscience*, *7*(1-4), 192-202.

Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of experimental social psychology*, *31*(5), 410-436.

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.

Dovidio, J. F., Johnson, J. D., Gaertner, S. L., Pearson, A. R., Saguy, T., & Ashburn-Nardo, L. (2010). Empathy and intergroup relations.

Downs, A., & Smith, T. (2004). Emotional understanding, cooperation, and social behavior in high-functioning children with autism. *Journal of autism and developmental disorders*, *34*(6), 625-635.

Dubey, I., Ropar, D., & Hamilton, A. (2018). Comparison of choose-a-movie and approach–avoidance paradigms to measure social motivation. *Motivation and emotion*, *42*(2), 190-199.

Duchenne, G.-B., & de Boulogne, G.-B. D. (1990). *The mechanism of human facial expression*. Cambridge university press.

Dunham, Y. (2018). Mere membership. *Trends in cognitive sciences*, *22*(9), 780-793.

Dworzynski, K., Ronald, A., Bolton, P., & Happé, F. (2012). How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders? *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(8), 788-797.

Dyck, M. J., Ferguson, K., & Shochet, I. M. (2001). Do autism spectrum disorders differ from each other and from non-spectrum disorders on emotion recognition tests? *European child & adolescent psychiatry*, *10*(2), 105-116.

Eberhardt, J. L. (2005). Imaging race. *American Psychologist*, *60*(2), 181.

Eisenmajer, R., & Prior, M. (1991). Cognitive linguistic correlates of ‘theory of mind’ability in autistic children. *British journal of developmental psychology*, *9*(2), 351-364.

Ekas, N. V., Lickenbrock, D. M., & Whitman, T. L. (2010). Optimism, social support, and well-being in mothers of children with autism spectrum disorder. *Journal of autism and developmental disorders*, *40*(10), 1274-1284.

Ekman, P. (2003). Darwin, deception, and facial expression. *Annals of the new York Academy of sciences*, *1000*(1), 205-221.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology: II. *Journal of personality and social psychology*, *58*(2), 342.

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior*, *1*(1), 56–75. https://doi.org/https://doi.org/10.1007/BF01115465

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of nonverbal behavior*, *6*(4), 238-252.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological Bulletin*, *128*(2), 203.

Ellemers, N., Spears, R., & Doosje, B. (2002). Self and social identity. *Annual review of psychology*, *53*(1), 161-186.

Farmer, H., Mahmood, R., Gregory, S. E., Tishina, P., & Hamilton, A. F. d. C. (2021). Dynamic emotional expressions do not modulate responses to gestures. *Acta psychologica*, *212*, 103226.

Fletcher-Watson, S., & Happé, F. (2019). *Autism: A new introduction to psychological theory and current debate*. Routledge.

Fletcher-Watson, S., Leekam, S. R., Benson, V., Frank, M., & Findlay, J. (2009). Eye-movements reveal attention to social information in autism spectrum disorder. *Neuropsychologia*, *47*(1), 248-257.

Frank, M. G., & Ekman, P. (1993). Not all smiles are created equal: The differences between enjoyment and nonenjoyment smiles. *International Journal of Humor Research*, *6*(1), 9-26. https://doi.org/https://doi.org/10.1515/humr.1993.6.1.9

Freitag, C. M., Staal, W., Klauck, S. M., Duketis, E., & Waltes, R. (2010). Genetics of autistic disorders: review and clinical implications. *European child & adolescent psychiatry*, *19*(3), 169-178.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human brain mapping*, *2*(4), 189-210.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531-534.

Frith, C. D., & Frith, U. (2008). Implicit and explicit processes in social cognition. *Neuron*, *60*(3), 503-510.

Frith, U. (2004). Emanuel Miller lecture: Confusions and controversies about Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *45*(4), 672-686.

Frith, U. (2012). Why we need cognitive explanations of autism. *The Quarterly Journal of Experimental Psychology*, *65*(11), 2073-2092.

Frith, U. (2013). Autism and dyslexia: A glance over 25 years of research. *Perspectives on Psychological Science*, *8*(6), 670-672.

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.

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind'in verbal and nonverbal tasks. *Neuropsychologia*, *38*(1), 11-21.

Gallese, V. (2007). Before and below ‘theory of mind’: embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1480), 659-669.

Gallese, V. (2009). Mirror neurons, embodied simulation, and the neural basis of social identification. *Psychoanalytic dialogues, 19*(5), 519-536.

Gallese, V., Gernsbacher, M. A., Heyes, C., Hickok, G., & Iacoboni, M. (2011). Mirror neuron forum. *Perspectives on Psychological Science*, *6*(4), 369-407.

Gamond, L., Vilarem, E., Safra, L., Conty, L., & Grèzes, J. (2017). Minimal group membership biases early neural processing of emotional expressions. *European Journal of Neuroscience*, *46*(10), 2584-2595.

Gernsbacher, M. A., Stevenson, J. L., & Dern, S. (2017). Specificity, contexts, and reference groups matter when assessing autistic traits. *PloS one*, *12*(2), e0171931.

Gillberg, C. (1998). Asperger syndrome and high-functioning autism. *The British journal of psychiatry*, *172*(3), 200-209.

Gliga, T., Senju, A., Pettinato, M., Charman, T., & Johnson, M. H. (2014). Spontaneous belief attribution in younger siblings of children on the autism spectrum. *Developmental psychology*, *50*(3), 903.

Gopnik, A., & Wellman, H. M. (1994). *The theory theory* An earlier version of this chapter was presented at the Society for Research in Child Development Meeting, 1991,

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085.

Green, C. C., Brown, N. J., Yap, V. M., Scheffer, I. E., & Wilson, S. J. (2019). Cognitive processes predicting advanced theory of mind in the broader autism phenotype. *Autism Research*.

Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2017). Implicit and explicit false belief development in preschool children. *Developmental science*, *20*(5), e12445.

Gu, X., Eilam‐Stock, T., Zhou, T., Anagnostou, E., Kolevzon, A., Soorya, L., Hof, P. R., Friston, K. J., & Fan, J. (2015). Autonomic and brain responses associated with empathy deficits in autism spectrum disorder. *Human brain mapping*, *36*(9), 3323-3338.

Hadad, B.-S., Schwartz, S., & Binur, N. (2019). Reduced perceptual specialization in autism: Evidence from the other-race face effect. *Journal of experimental psychology: general*, *148*(3), 588.

Hahn, U., & Harris, A. J. (2014). What does it mean to be biased: Motivated reasoning and rationality. In *Psychology of learning and motivation* (Vol. 61, pp. 41-102). Elsevier.

Hale, J., & Hamilton, A. F. d. C. (2016). Cognitive mechanisms for responding to mimicry from others. *Neuroscience & Biobehavioral Reviews*, *63*, 106-123.

Han, S. (2018). Neurocognitive basis of racial ingroup bias in empathy. *Trends in cognitive sciences*, *22*(5), 400-421.

Happé, F. (2018). Why are savant skills and special talents associated with autism? *World Psychiatry*, *17*(3), 280.

Happé, F., & Vital, P. (2009). What aspects of autism predispose to talent? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1522), 1369-1375.

Happé, F. G. (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.

Happé, F. G. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child development*, *66*(3), 843-855.

Happé, F. G., Cook, J. L., & Bird, G. (2017). The structure of social cognition: In (ter) dependence of sociocognitive processes. *Annual review of psychology*, *68*, 243-267.

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. *Neuropsychology review*, *20*(3), 290-322.

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, *17*(10), 847-853.

Harris, L. T., & Fiske, S. T. (2011). Dehumanized Perception: A Psychological Means to Facilitate Atrocities, Torture, and Genocide? *Zeitschrift für Psychologie*, *219*(3), 175-181. https://doi.org/10.1027/2151-2604/a000065

Hayashi, T., Akikawa, R., Kawasaki, K., Egawa, J., Minamimoto, T., Kobayashi, K., Kato, S., Hori, Y., Nagai, Y., & Iijima, A. (2020). Macaques Exhibit Implicit Gaze Bias Anticipating Others' False-Belief-Driven Actions via Medial Prefrontal Cortex. *Cell Reports*, *30*(13), 4433-4444.

Heerey, E. A. (2014). Learning from social rewards predicts individual differences in self-reported social ability. *Journal of experimental psychology: general*, *143*(1), 332.

Hennenlotter, A., Dresel, C., Castrop, F., Ceballos-Baumann, A. O., Wohlschläger, A. M., & Haslinger, B. (2009). The link between facial feedback and neural activity within central circuitries of emotion—New insights from Botulinum toxin–induced denervation of frown muscles. *Cerebral cortex*, *19*(3), 537-542.

Herold, F., Wiegel, P., Scholkmann, F., Thiers, A., Hamacher, D., & Schega, L. (2017). Functional near-infrared spectroscopy in movement science: a systematic review on cortical activity in postural and walking tasks. *Neurophotonics*, *4*(4), 041403-041403.

Herwig, U., Kaffenberger, T., Jäncke, L., & Brühl, A. B. (2010). Self-related awareness and emotion regulation. *NeuroImage*, *50*(2), 734-741.

Hess, U., Blairy, S., & Kleck, R. E. (1997). The intensity of emotional facial expressions and decoding accuracy. *Journal of nonverbal behavior*, *21*, 241-257.

Heyes, C. (2011). Automatic imitation. *Psychological Bulletin, 137*(3), 463.

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science*, *9*(2), 131-143.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, *21*(7), 1229-1243.

Hickok, G. (2013). Do mirror neurons subserve action understanding? *Neuroscience letters*, *540*, 56-58.

Hickok, G., & Hauser, M. (2010). (Mis) understanding mirror neurons. *Current Biology, 20*(14), R593-R594.

Hill, E., Berthoz, S., & Frith, U. (2004). Brief report: Cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives. *Journal of autism and developmental disorders, 34*(2), 229-235.

Hill, E. L., & Frith, U. (2003). Understanding autism: insights from mind and brain. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1430), 281-289.

Hobson, R. P. (1986). The autistic child's appraisal of expressions of emotion. *Journal of Child Psychology and Psychiatry*, *27*(3), 321-342.

Hofvander, B., Delorme, R., Chaste, P., Nydén, A., Wentz, E., Ståhlberg, O., Herbrecht, E., Stopin, A., Anckarsäter, H., & Gillberg, C. (2009). Psychiatric and psychosocial problems in adults with normal-intelligence autism spectrum disorders. *BMC psychiatry*, *9*(1), 35.

Hooker, C. I., Verosky, S. C., Germine, L. T., Knight, R. T., & D'Esposito, M. (2008). Mentalizing about emotion and its relationship to empathy. *Social Cognitive and Affective Neuroscience*, *3*(3), 204-217.

Hosokawa, T., Kennerley, S. W., Sloan, J., & Wallis, J. D. (2013). Single-neuron mechanisms underlying cost-benefit analysis in frontal cortex. *Journal of Neuroscience*, *33*(44), 17385-17397.

Hoss, R. A., Ramsey, J. L., Griffin, A. M., & Langlois, J. H. (2005). The Role of Facial Attractiveness and Facial Masculinity/Femininity in Sex Classification of Faces. *Perception (London)*, *34*(12), 1459-1474. https://doi.org/10.1068/p5154

Howard, J. W., & Rothbart, M. (1980). Social categorization and memory for in-group and out-group behavior. *Journal of personality and social psychology*, *38*(2), 301.

Hughes, B. L., Ambady, N., & Zaki, J. (2017). Trusting outgroup, but not ingroup members, requires control: neural and behavioral evidence. *Social Cognitive and Affective Neuroscience*, *12*(3), 372-381.

Hughes, B. L., Zaki, J., & Ambady, N. (2017). Motivation alters impression formation and related neural systems. *Social Cognitive and Affective Neuroscience*, *12*(1), 49-60.

Hull, L., Mandy, W., Lai, M.-C., Baron-Cohen, S., Allison, C., Smith, P., & Petrides, K. (2019). Development and validation of the camouflaging autistic traits questionnaire (CAT-Q). *Journal of autism and developmental disorders*, *49*(3), 819-833.

Hull, L., Petrides, K., Allison, C., Smith, P., Baron-Cohen, S., Lai, M.-C., & Mandy, W. (2017). "Putting on my best normal": social camouflaging in adults with autism spectrum conditions. *Journal of autism and developmental disorders*, *47*(8), 2519-2534.

Hull, L., Petrides, K., & Mandy, W. (2020). The female autism phenotype and camouflaging: A narrative review. *Review Journal of Autism and Developmental Disorders, 7,* 306-317.

Huppert, T. J., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied optics, 48*(10), D280-D298.

Hurley, R. S., Losh, M., Parlier, M., Reznick, J. S., & Piven, J. (2007). The broad autism phenotype questionnaire. *Journal of autism and developmental disorders*, *37*(9), 1679-1690.

Hyde, D. C., Aparicio Betancourt, M., & Simon, C. E. (2015). Human temporal-parietal junction spontaneously tracks others' beliefs: A functional near-infrared spectroscopy study. *Human brain mapping*, *36*(12), 4831-4846.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*(5449), 2526-2528.

Ingersoll, B., & Hambrick, D. Z. (2011). The relationship between the broader autism phenotype, child severity, and stress and depression in parents of children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *5*(1), 337-344.

Ingersoll, B., Hopwood, C. J., Wainer, A., & Donnellan, M. B. (2011). A comparison of three self-report measures of the broader autism phenotype in a non-clinical sample. *Journal of autism and developmental disorders*, *41*(12), 1646-1657.

Ingersoll, B., Meyer, K., & Becker, M. W. (2011). Increased rates of depressed mood in mothers of children with ASD associated with the presence of the broader autism phenotype. *Autism Research*, *4*(2), 143-148.

Jacob, P. (2008). What do mirror neurons contribute to human social cognition? *Mind & Language*, *23*(2), 190-223.

Johnston, L., Miles, L., & Macrae, C. N. (2010). Why are you smiling at me? Social functions of enjoyment and non‐enjoyment smiles. *British Journal of Social Psychology*, *49*(1), 107-127.

Jones, C. R., Simonoff, E., Baird, G., Pickles, A., Marsden, A. J., Tregay, J., Happé, F., & Charman, T. (2018). The association between theory of mind, executive function, and the symptoms of autism spectrum disorder. *Autism Research*, *11*(1), 95-109.

Jordan, J. J., McAuliffe, K., & Warneken, F. (2014). Development of in-group favoritism in children's third-party punishment of selfishness. *Proceedings of the National Academy of Sciences, 111*(35), 12710-12715.

Kampis, D., Karman, P., Csibra, G., Southgate, V., & Hernik, M. (2020). *A two-lab direct replication attempt of Southgate, Senju, & Csibra (2007)*.

Kampis, D., Karman, P., Csibra, G., Southgate, V., & Hernik, M. (2021). A two-lab direct replication attempt of Southgate, Senju and Csibra (2007). *Royal Society open science, 8*(8), 210190.

Kang, J., Han, X., Hu, J.-F., Feng, H., & Li, X. (2020). The study of the differences between low-functioning autistic children and typically developing children in the processing of the own-race and other-race faces by the machine learning approach. *Journal of Clinical Neuroscience*, *81,* 54-60.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature neuroscience, 3*(8), 759-763.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302-4311.

Karst, J. S., & Van Hecke, A. V. (2012). Parent and family impact of autism spectrum disorders: A review and proposed model for intervention evaluation. *Clinical child and family psychology review*, *15*(3), 247-277.

Katsumi, Y., & Dolcos, S. (2018). Neural correlates of racial ingroup bias in observing computer-animated social encounters. *Frontiers in human neuroscience*, *11*, 632.

Ketelaars, M. P., Mol, A., Swaab, H., & van Rijn, S. (2016). Emotion recognition and alexithymia in high functioning females with autism spectrum disorder. *Research in Autism Spectrum Disorders*, *21*, 51-60.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? Perception 36 ECVP Abstract Supplement.

Kliemann, D., Young, L., Scholz, J., & Saxe, R. (2008). The influence of prior record on moral judgment. *Neuropsychologia*, *46*(12), 2949-2957.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of general psychiatry*, *59*(9), 809-816.

Kocsis, L., Herman, P., & Eke, A. (2006). The modified Beer–Lambert law revisited. *Physics in Medicine & Biology*, *51*(5), N91.

Komeda, H., Kosaka, H., Fujioka, T., Jung, M., & Okazawa, H. (2019). Do individuals with autism spectrum disorders help other people with autism spectrum disorders? An investigation of empathy and helping motivation in adults with autism spectrum disorder. *Frontiers in psychiatry*, *10*, 376.

Konvalinka, I., & Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Frontiers in human neuroscience*, *6*, 215.

Korb, S., With, S., Niedenthal, P., Kaiser, S., & Grandjean, D. (2014). The perception and mimicry of facial movements predict judgments of smile authenticity. *PloS one*, *9*(6), e99194.

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: a neural prediction problem. *Neuron*, *79*(5), 836-848.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830-1834.

Krautheim, J. T., Dannlowski, U., Steines, M., Neziroğlu, G., Acosta, H., Sommer, J., Straube, B., & Kircher, T. (2019). Intergroup empathy: enhanced neural resonance for ingroup facial emotion in a shared neural production-perception network. *NeuroImage*, *194*, 182-190.

Krumhuber, E., Manstead, A. S., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, *7*(4), 730.

Kulke, L., & Hinrichs, M. A. B. (2021). Implicit Theory of Mind under realistic social circumstances measured with mobile eye-tracking. *Scientific reports*, *11*(1), 1-13.

Kulke, L., Johannsen, J., & Rakoczy, H. (2019). Why can some implicit Theory of Mind tasks be replicated and others cannot? A test of mentalizing versus submentalizing accounts. *PloS one*, *14*(3), e0213772.

Kulke, L., & Rakoczy, H. (2019). Testing the role of verbal narration in implicit theory of mind tasks. *Journal of Cognition and Development*, *20*(1), 1-14.

Kulke, L., Reiß, M., Krist, H., & Rakoczy, H. (2018). How robust are anticipatory looking measures of Theory of Mind? Replication attempts across the life span. *Cognitive development*, *46*, 97-111.

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.

Kulke, L., Wübker, M., & Rakoczy, H. (2019). Is implicit Theory of Mind real but hard to detect? Testing adults with different stimulus materials. *Royal Society open science*, *6*(7), 190068.

Lai, M.-C., & Baron-Cohen, S. (2015). Identifying the lost generation of adults with autism spectrum conditions. *The Lancet Psychiatry*, *2*(11), 1013-1027.

Lai, M.-C., Lombardo, M. V., Pasco, G., Ruigrok, A. N., Wheelwright, S. J., Sadek, S. A., Chakrabarti, B., Baron-Cohen, S., & Consortium, M. A. (2011). A behavioral comparison of male and female adults with high functioning autism spectrum conditions. *PloS one*, *6*(6), e20835.

Lai, M.-C., Lombardo, M. V., Ruigrok, A. N., Chakrabarti, B., Auyeung, B., Szatmari, P., Happé, F., Baron-Cohen, S., & Consortium, M. A. (2017). Quantifying and exploring camouflaging in men and women with autism. *Autism*, *21*(6), 690-702.

Lavan, N., Rankin, G., Lorking, N., Scott, S., & McGettigan, C. (2017). Neural correlates of the affective properties of spontaneous and volitional laughter types. *Neuropsychologia*, *95*, 30-39.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

Lee, M., Lori, A., Langford, N. A., & Rilling, J. K. (2023). The neural basis of smile authenticity judgments and the potential modulatory role of the oxytocin receptor gene (OXTR). *Behavioural Brain Research*, *437*, 114144.

Leedham, A., Thompson, A. R., Smith, R., & Freeth, M. (2020). 'I was exhausted trying to figure it out': The experiences of females receiving an autism diagnosis in middle to late adulthood. *Autism*, *24*(1), 135-146.

Lehnhardt, F.-G., Gawronski, A., Pfeiffer, K., Kockler, H., Schilbach, L., & Vogeley, K. (2013). The investigation and differential diagnosis of Asperger syndrome in adults. *Deutsches Ärzteblatt International*, *110*(45), 755.

Leslie, A. M. (1987). Pretense and representation: The origins of" theory of mind.". *Psychological review*, *94*(4), 412.

Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false belief task: Two conjectures. *Developmental science*, *1*(2), 247-253. https://doi.org/https://doi.org/10.1111/1467-7687.00038

Lewis, L. F. (2016). Exploring the experience of self-diagnosis of autism spectrum disorder in adults. *Archives of psychiatric nursing*, *30*(5), 575-580.

Leyens, J. P., Demoulin, S., Vaes, J., Gaunt, R., & Paladino, M. P. (2007). Infra-humanization: The wall of group differences. *Social Issues and Policy Review*, *1*(1), 139-172.

Li, Q., Li, Y., Liu, B., Chen, Q., Xing, X., Xu, G., & Yang, W. (2022). Prevalence of Autism Spectrum Disorder Among Children and Adolescents in the United States from 2019 to 2020. *JAMA pediatrics*, *176*(9), 943-945.

Lin, L. C., Qu, Y., & Telzer, E. H. (2018). Intergroup social influence on emotion processing in the brain. *Proceedings of the National Academy of Sciences*, *115*(42), 10630-10635.

Liu, L., & Humpolíček, P. (2013). The comparative study of facial emotion recognition ability in ADHD individuals and individuals with high functioning autism/Asperger syndrome. *Klinická psychologie a osobnost*, *2*(1), 15-25.

Liu, N., Cui, X., Bryant, D. M., Glover, G. H., & Reiss, A. L. (2015). Inferring deep-brain activity from cortical activity using functional near-infrared spectroscopy. *Biomedical optics express*, *6*(3), 1074-1089.

Livingston, L. A., Colvert, E., Social Relationships Study Team, Bolton, P., & Happé, F. (2019). Good social skills despite poor theory of mind: exploring compensation in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 60*(1), 102-110.

Livingston, L. A., & Happé, F. (2017). Conceptualising compensation in neurodevelopmental disorders: Reflections from autism spectrum disorder. *Neuroscience & Biobehavioral Reviews, 80,* 729-742.

Lloyd-Fox, S., Papademetriou, M., Darboe, M. K., Everdell, N. L., Wegmuller, R., Prentice, A. M., Moore, S. E., & Elwell, C. E. (2014). Functional near infrared spectroscopy (fNIRS) to assess cognitive function in infants in rural Africa. *Scientific reports, 4*(1), 4740.

Lockwood Estrin, G., Milner, V., Spain, D., Happé, F., & Colvert, E. (2021). Barriers to autism spectrum disorder diagnosis for young women and girls: A systematic review. *Review Journal of Autism and Developmental Disorders, 8*(4), 454-470.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders*, *30*(3), 205-223.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule (ADOS-2)* (Second ed.). Western Psychological Services.

Machery, E. (2014). In defense of reverse inference. *The British Journal for the Philosophy of Science*.

MacInnis, C. C., & Hodson, G. (2012). Intergroup bias toward “Group X”: Evidence of prejudice, dehumanization, avoidance, and discrimination against asexuals. *Group processes & intergroup relations*, *15*(6), 725-743.

MacLachlan, M. (2020). Commentary: Challenges and opportunities in autism assessment–a commentary on Kannes and Bishop (2020). *Journal of Child Psychology and Psychiatry*.

Mandy, W., & Tchanturia, K. (2015). Do women with eating disorders who have social and flexibility difficulties really have autism? A case series. *Molecular autism*, *6*(1), 6.

Manera, V., Del Giudice, M., Grandi, E., & Colle, L. (2011). Individual differences in the recognition of enjoyment smiles: No role for perceptual–attentional factors and autistic-like traits. *Frontiers in psychology*, *2*, 143.

Mathur, V. A., Harada, T., Lipke, T., & Chiao, J. Y. (2010). Neural basis of extraordinary empathy and altruistic motivation. *NeuroImage*, *51*(4), 1468-1475.

Matsumoto, D., & Hwang, H. C. (2018). Microexpressions differentiate truths from lies about future malicious intent. *Frontiers in psychology*, *9*, 2545.

Mazza, M., Pino, M. C., Vagnetti, R., Peretti, S., Valenti, M., Marchetti, A., & Di Dio, C. (2020). Discrepancies between explicit and implicit evaluation of aesthetic perception ability in individuals with autism: a potential way to improve social functioning. *BMC psychology*, *8*(1), 1-15.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., & Pike, B. (2001a). A four-dimensional probabilistic atlas of the human brain. *Journal of the American Medical Informatics Association*, *8*(5), 401-430.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., & Pike, B. (2001b). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical*

Transactions of the Royal Society of London. Series B: Biological Sciences, *356*(1412), 1293-1322.

McGettigan, C., Walsh, E., Jessop, R., Agnew, Z., Sauter, D., Warren, J., & Scott, S. (2015). Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. *Cerebral cortex*, *25*(1), 246-257.

McLellan, T., Wilcke, J., Johnston, L., Watts, R., & Miles, L. (2012). Sensitivity to posed and genuine displays of happiness and sadness: a fMRI study. *Neuroscience letters*, *531*(2), 149-154.

McLoughlin, N., & Over, H. (2017). Young children are more likely to spontaneously attribute mental states to members of their own group. *Psychological Science*, *28*(10), 1503-1509.

McQuaid, G. A., Lee, N. R., & Wallace, G. L. (2022). Camouflaging in autism spectrum disorder: Examining the roles of sex, gender identity, and diagnostic timing. *Autism*, *26*(2), 552-559.

Mehu, M., Grammer, K., & Dunbar, R. I. (2007). Smiles when sharing. *Evolution and Human Behavior*, *28*(6), 415-422.

Micali, N., Chakrabarti, S., & Fombonne, E. (2004). The broad autism phenotype: findings from an epidemiological survey. *Autism*, *8*(1), 21-37.

Milton, D. E. (2012). On the ontological status of autism: the ‘double empathy problem’. *Disability & Society*, *27*(6), 883-887.

Mitchell, P., Cassidy, S., & Sheppard, E. (2019). The double empathy problem, camouflage, and the value of expertise from experience. *Behavioral and brain sciences, 42*.

Mitchell, P., Sheppard, E., & Cassidy, S. (2021). Autism and the double empathy problem: Implications for development and mental health. *British journal of developmental psychology*, *39*(1), 1-18.

Mitteroecker, P., Windhager, S., Müller, G. B., & Schaefer, K. (2015). The morphometrics of "masculinity" in human faces. *PloS one*, *10*(2), e0118374-e0118374. https://doi.org/10.1371/journal.pone.0118374

Molavi, B., & Dumont, G. A. (2012). Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiological measurement*, *33*(2), 259.

Molenberghs, P. (2013). The neuroscience of in-group bias. *Neuroscience & Biobehavioral Reviews*, *37*(8), 1530-1536.

Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2009). Is the mirror neuron system involved in imitation? A short review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *33*(7), 975-980.

Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, *65*, 276-291.

Molenberghs, P., & Louis, W. R. (2018). Insights from fMRI studies into ingroup bias. *Frontiers in psychology*, *9*, 1868.

Mondillon, L., Niedenthal, P. M., Gil, S., & Droit-Volet, S. (2007). Imitation of in-group versus out-group members' facial expressions of anger: A test with a time perception task. *Social neuroscience*, *2*(3-4), 223-237.

Montagu, A. (1997). *Man's most dangerous myth: The fallacy of race*. Rowman & Littlefield.

Moradi, Z., Najlerahim, A., Macrae, C. N., & Humphreys, G. W. (2020). Attentional saliency and ingroup biases: From society to the brain. *Social neuroscience*, *15*(3), 324-333.

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European journal of social psychology, 22*(2), 103-122.

Naughtin, C. K., Horne, K., Schneider, D., Venini, D., York, A., & Dux, P. E. (2017). Do implicit and explicit belief processing share neural substrates? *Human brain mapping, 38*(9), 4760-4772.

Nayar, K., Sealock, J. M., Maltman, N., Bush, L., Cook, E. H., Davis, L. K., & Losh, M. (2021). Elevated polygenic burden for autism spectrum disorder is associated with the broad autism phenotype in mothers of individuals with autism spectrum disorder. *Biological psychiatry*, *89*(5), 476-485.

Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and brain sciences, 33*(6), 417-433.

Nijhof, A. D., Bardi, L., Brass, M., & Wiersema, J. R. (2018). Brain activity for spontaneous and explicit mentalizing in adults with autism spectrum disorder: An fMRI study. *NeuroImage: Clinical, 18,* 475-484.

Nijhof, A. D., Brass, M., Bardi, L., & Wiersema, J. R. (2016). Measuring mentalizing ability: A within-subject comparison between an explicit and implicit version of a ball detection task. *PloS one, 11*(10), e0164373.

Nijhof, A. D., Brass, M., & Wiersema, J. R. (2017). Spontaneous mentalizing in neurotypicals scoring high versus low on symptomatology of autism spectrum disorder. *Psychiatry research, 258*, 15-20.

Obrig, H. (2014). NIRS in clinical neurology—a ‘promising’tool? *NeuroImage, 85,* 535-546.

Ochsner, K. N., Knierim, K., Ludlow, D. H., Hanelin, J., Ramachandran, T., Glover, G., & Mackey, S. C. (2004). Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other. *Journal of Cognitive Neuroscience*, *16*(10), 1746-1772.

Olsson, A., & Ochsner, K. N. (2008). The role of social cognition in emotion. *Trends in cognitive sciences*, *12*(2), 65-71.

Orlowska, A., Rychlowska, M., Szarota, P., & Krumhuber, E. G. (2023). Facial mimicry and social context affect smile interpretation. *Journal of nonverbal behavior*, 1-18.

Orlowska, A. B., Krumhuber, E. G., Rychlowska, M., & Szarota, P. (2018). Dynamics matter: recognition of reward, affiliative, and dominance smiles from dynamic vs. static displays. *Frontiers in psychology*, *9*, 938.

Over, H. (2021). Seven challenges for the dehumanization hypothesis. *Perspectives on Psychological Science*, *16*(1), 3-13.

Palmen, S. J., Pol, H. E. H., Kemner, C., Schnack, H. G., Sitskoorn, M. M., Appels, M. C., Kahn, R. S., & Van Engeland, H. (2005). Brain anatomy in non-affected parents of autistic probands: a MRI study. *Psychological medicine*, *35*(10), 1411-1420.

Pan, Y., Borragán, G., & Peigneux, P. (2019). Applications of functional near-infrared spectroscopy in fatigue, sleep deprivation, and social cognition. *Brain Topography*, *32*, 998-1012.

Park, B., & Young, L. (2020). An association between biased impression updating and relationship facilitation: A behavioral and fMRI investigation. *Journal of experimental social psychology*, *87*, 103916.

Pelli, D. G., & Vision, S. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, *10*, 437-442.

Peng, S., Kuang, B., & Hu, P. (2020). Right temporoparietal junction modulates in-group bias in facial emotional mimicry: A tDCS study. *Frontiers in Behavioral Neuroscience*, *14*, 143.

Peng, S., Zhang, L., & Hu, P. (2021). Relating self–other overlap to ingroup bias in emotional mimicry. *Social neuroscience*, *16*(4), 439-447.

Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., & Tachtsidis, I. (2019). Current status and issues regarding pre-processing of fNIRS neuroimaging data: an investigation of diverse signal filtering methods within a general linear model framework. *Frontiers in human neuroscience*, *12*, 505.

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the new York Academy of sciences*, *1464*(1), 5-29.

Piven, J., Palmer, P., Jacobi, D., Childress, D., & Arndt, S. (1997). Broader autism phenotype: evidence from a family history study of multiple-incidence autism families. *American Journal of Psychiatry*, *154*(2), 185-190.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in cognitive sciences*, *10*(2), 59-63.

Poldrack, R. A. (2008). The role of fMRI in cognitive neuroscience: where do we stand? *Current opinion in neurobiology*, *18*(2), 223-227.

Poldrack, R. A. (2012). The future of fMRI in cognitive neuroscience. *NeuroImage*, *62*(2), 1216-1220.

Portway, S. M., & Johnson, B. (2005). Do you know I have Asperger's syndrome? Risks of a non-obvious disability. *Health, Risk & Society*, *7*(1), 73-83.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, *1*(4), 515-526.

Pruitt, M. M., Rhoden, M., & Ekas, N. V. (2018). Relationship between the broad autism phenotype, social relationships and mental health for mothers of children with autism spectrum disorder. *Autism*, *22*(2), 171-180.

Pu, S., Matsumura, H., Yamada, T., Ikezawa, S., Mitani, H., Adachi, A., & Nakagome, K. (2008). Reduced frontopolar activation during verbal fluency task associated with poor social functioning in late‐onset major depression: Multi‐channel near‐infrared spectroscopy study. *Psychiatry and Clinical Neurosciences*, *62*(6), 728-737.

Qian, C., Tei, S., Itahashi, T., Aoki, Y. Y., Ohta, H., Hashimoto, R.-i., Nakamura, M., Takahashi, H., Kato, N., & Fujino, J. (2022). Intergroup bias in punishing behaviors of adults with autism spectrum disorder. *Frontiers in psychiatry*, *13*, 884529-884529.

Qu, C., Ligneul, R., Van der Henst, J.-B., & Dreher, J.-C. (2017). An integrative interdisciplinary perspective on social dominance hierarchies. *Trends in cognitive sciences*, *21*(11), 893-908.

Quaresima, V., & Ferrari, M. (2019). Functional near-infrared spectroscopy (fNIRS) for assessing cerebral cortex function during human behavior in natural/social situations: a concise review. *Organizational Research Methods*, *22*(1), 46-68.

Rea, H. M., Factor, R. S., Swain, D. M., & Scarpa, A. (2019). The Association of the Broader Autism Phenotype with Emotion-Related Behaviors in Mothers of Children With and Without Autism Spectrum Traits. *Journal of autism and developmental disorders*, *49*(3), 950-959.

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, *20*(8), 495-505.

Rilling, J. K., Dagenais, J. E., Goldsmith, D. R., Glenn, A. L., & Pagnoni, G. (2008). Social cognitive neural networks during in-group and out-group interactions. *NeuroImage*, *41*(4), 1447-1461.

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and embryology*, *210*(5-6), 419-421.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.*, *27*, 169-192.

Rizzolatti, G., & Sinigaglia, C. (2016). The mirror mechanism: a basic principle of brain function. *Nature Reviews Neuroscience*, *17*(12), 757-765.

Rodd, J. M. (2023). Moving Experimental Psychology Online: How to Maintain Data Quality When We Can't See Our Participants.

Rosenberg, E. L., & Ekman, P. (2020). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press.

Rubenstein, E., & Chawla, D. (2018). Broader autism phenotype in parents of children with autism: a systematic review of percentage estimates. *Journal of child and family studies*, *27*(6), 1705-1720.

Rubenstein, E., Pretzel, R. E., Windham, G. C., Schieve, L. A., Wiggins, L. D., DiGuiseppi, C., Olshan, A. F., Howard, A. G., Pence, B. W., & Young, L. (2017). The broader autism phenotype in mothers is associated with increased discordance between maternal-reported and clinician-observed instruments that measure child autism spectrum disorder. *Journal of autism and developmental disorders*, *47*(10), 3253-3266.

Rudman, L. A., & Goodwin, S. A. (2004). Gender differences in automatic in-group bias: Why do women like women more than men like men? *Journal of personality and social psychology*, *87*(4), 494.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The ‘windows task’as a measure of strategic deception in preschoolers and autistic subjects. *British journal of developmental psychology*, *9*(2), 331-349.

Rutherford, M., McKenzie, K., Johnson, T., Catchpole, C., O’Hare, A., McClure, I., Forsyth, K., McCartney, D., & Murray, A. (2016). Gender ratio in a clinical population sample, age of diagnosis and duration of assessment in children and adults with autism spectrum disorder. *Autism*, *20*(5), 628-634.

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*. Western Psychological Services.

Rychlowska, M., Cañadas, E., Wood, A., Krumhuber, E. G., Fischer, A., & Niedenthal, P. M. (2014). Blocking mimicry makes true and false smiles look the same. *PloS one*, *9*(3), e90876.

Salminen, J. K., Saarijärvi, S., Äärelä, E., Toikka, T., & Kauhanen, J. (1999). Prevalence of alexithymia and its association with sociodemographic variables in the general population of Finland. *Journal of psychosomatic research*, *46*(1), 75-82.

Sasson, N. J., Faso, D. J., Nugent, J., Lovell, S., Kennedy, D. P., & Grossman, R. B. (2017). Neurotypical peers are less willing to interact with those with autism based on thin slice judgments. *Scientific reports*, *7*(1), 1-10.

Sasson, N. J., Lam, K. S., Parlier, M., Daniels, J. L., & Piven, J. (2013). Autism and the broad autism phenotype: familial patterns and intergenerational transmission. *Journal of Neurodevelopmental Disorders*, *5*(1), 11.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: linking developmental psychology and functional neuroimaging. *Annu. Rev. Psychol.*, *55*, 87-124.

Scheepers, D., & Derks, B. (2016). Revisiting social identity theory from a neuroscience perspective. *Current Opinion in Psychology*, *11*, 74-78.

Scheeren, A. M., & Stauder, J. E. (2008). Broader autism phenotype in parents of autistic children: reality or myth? *Journal of autism and developmental disorders*, *38*(2), 276.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience1. *Behavioral and brain sciences, 36*(4), 393-414.

Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others' mental states. *Journal of experimental psychology: general*, *141*(3), 433.

Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts implicit theory-of-mind processing. *Psychological Science*, *23*(8), 842-847.

Schneider, D., 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.

Schneider, D., Slaughter, V. P., Becker, S. I., & Dux, P. E. (2014). Implicit false-belief processing in the human brain. *NeuroImage*, *101*, 268-275.

Schneider, D., Slaughter, V. P., & Dux, P. E. (2017). Current evidence for automatic Theory of Mind processing in adults. *Cognition, 162,* 27-31.

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Pavia, J. M., Wolf, U., & Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *NeuroImage, 85,* 6-27.

Schug, J., Matsumoto, D., Horita, Y., Yamagishi, T., & Bonnet, K. (2010). Emotional expressivity as a signal of cooperation. *Evolution and Human Behavior*, *31*(2), 87-94.

Schupp, H. T., Junghöfer, M., Weike, A. I., & Hamm, A. O. (2003). Attention and emotion: an ERP analysis of facilitated emotional stimulus processing. *Neuroreport*, *14*(8), 1107-1110.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews, 42,* 9-34.

Schuwerk, T., Jarvers, I., Vuori, M., & Sodian, B. (2016). Implicit mentalizing persists beyond early childhood and is profoundly impaired in children with autism spectrum condition. *Frontiers in psychology, 7,* 1696.

Schuwerk, T., Priewasser, B., Sodian, B., & Perner, J. (2018). The robustness and generalizability of findings on spontaneous false belief sensitivity: A replication attempt. *Royal Society open science, 5*(5), 172273.

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.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., Reiss, A. L., & Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, *27*(9), 2349-2356.

Senju, A., Southgate, V., Miura, Y., Matsui, T., Hasegawa, T., Tojo, Y., Osanai, H., & Csibra, G. (2010). Absence of spontaneous action anticipation by false belief attribution in children with autism spectrum disorder. *Development and psychopathology*, *22*(2), 353-360.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science*, *325*(5942), 883-885.

Shackman, A. J., McMenamin, B. W., Maxwell, J. S., Greischar, L. L., & Davidson, R. J. (2009). Right dorsolateral prefrontal cortical activity and behavioral inhibition. *Psychological Science*, *20*(12), 1500-1506.

Shah, P., Hall, R., Catmur, C., & Bird, G. (2016). Alexithymia, not autism, is associated with impaired interoception. *cortex*, *81*, 215-220.

Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *The Neuroscientist*, *17*(1), 18-24.

Sheppard, E., Pillai, D., Wong, G. T.-L., Ropar, D., & Mitchell, P. (2016). How easy is it to read the minds of people with autism spectrum disorder? *Journal of autism and developmental disorders*, *46*(4), 1247-1254.

Shore, D. M., & Heerey, E. A. (2011). The value of genuine and polite smiles. *Emotion*, *11*(1), 169.

Singer, T., Seymour, B., O'doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, *303*(5661), 1157-1162.

Sinzig, J., Morsch, D., & Lehmkuhl, G. (2008). Do hyperactivity, impulsivity and inattention have an impact on the ability of facial affect recognition in children with autism and ADHD? *European child & adolescent psychiatry*, *17*(2), 63-72.

Smallwood, J., Bernhardt, B. C., Leech, R., Bzdok, D., Jefferies, E., & Margulies, D. S. (2021). The default mode network in cognition: a topographical perspective. *Nature Reviews Neuroscience*, *22*(8), 503-513.

Smith, A. (2009). The empathy imbalance hypothesis of autism: a theoretical approach to cognitive and emotional empathy in autistic development. *the Psychological record*, *59*(3), 489-510.

Smith, E. H., Horga, G., Yates, M. J., Mikell, C. B., Banks, G. P., Pathak, Y. J., Schevon, C. A., McKhann, G. M., Hayden, B. Y., & Botvinick, M. M. (2019). Widespread temporal coding of cognitive control in the human prefrontal cortex. *Nature neuroscience, 22*(11), 1883-1891.

Sodian, B., Schuwerk, T., & Kristen, S. (2015). Implicit and spontaneous theory of mind reasoning in autism spectrum disorders. *Autism spectrum disorder—Recent advances*, 113-135.

Song, R., Over, H., & Carpenter, M. (2016). Young children discriminate genuine from fake smiles and expect people displaying genuine smiles to be more prosocial. *Evolution and Human Behavior*, *37*(6), 490-501.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*(7), 587-592.

Spilberger, C. (1983). Manual for the State-Trait Anxiety Inventory: STAI (Form Y). Palo Alto. In: CA: Consulting Psychologists Press.

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron, 56*(1), 185-196.

Spunt, R. P., & Lieberman, M. D. (2012). An integrative model of the neural systems supporting the comprehension of observed emotional behavior. *NeuroImage, 59*(3), 3050-3059.

Stagg, S. D., & Belcher, H. (2019). Living with autism without knowing: receiving a diagnosis in later life. *Health Psychology and Behavioral Medicine*, *7*(1), 348-361.

Stallen, M., Rossi, F., Heijne, A., Smidts, A., De Dreu, C. K., & Sanfey, A. G. (2018). Neurobiological mechanisms of responding to injustice. *Journal of Neuroscience, 38*(12), 2944-2954.

Steele, S., Joseph, R. M., & Tager-Flusberg, H. (2003). Brief report: Developmental change in theory of mind abilities in children with autism. *Journal of autism and developmental disorders*, *33*(4), 461-467.

Steinbeis, N., Bernhardt, B. C., & Singer, T. (2012). Impulse control and underlying functions of the left DLPFC mediate age-related and age-independent individual differences in strategic social behavior. *Neuron*, *73*(5), 1040-1051.

Stel, M., & Van Knippenberg, A. (2008). The role of facial mimicry in the recognition of affect. *Psychological Science*, *19*(10), 984-985.

Sterzing, P. R., Shattuck, P. T., Narendorf, S. C., Wagner, M., & Cooper, B. P. (2012). Bullying involvement and autism spectrum disorders: Prevalence and correlates of bullying involvement among adolescents with an autism spectrum disorder. *Archives of pediatrics & adolescent medicine*, *166*(11), 1058-1064.

Stewart, G. R., Wallace, G. L., Cottam, M., & Charlton, R. A. (2020). Theory of mind performance in younger and older adults with elevated autistic traits. *Autism Research*, *13*(5), 751-762.

Su, X., Cai, R. Y., & Uljarević, M. (2018). Predictors of mental health in chinese parents of children with autism Spectrum disorder (ASD). *Journal of autism and developmental disorders*, *48*(4), 1159-1168.

Sucksmith, E., Allison, C., Baron-Cohen, S., Chakrabarti, B., & Hoekstra, R. A. (2013). Empathy and emotion recognition in people with autism, first-degree relatives, and controls. *Neuropsychologia*, *51*(1), 98-105.

Sucksmith, E., Roth, I., & Hoekstra, R. (2011). Autistic traits below the clinical threshold: re-examining the broader autism phenotype in the 21st century. *Neuropsychology review*, *21*(4), 360-389.

Surian, L., & Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *British journal of developmental psychology, 30*(1), 30-44.

Suzuki, S., Jensen, E. L., Bossaerts, P., & O'Doherty, J. P. (2016). Behavioral contagion during learning about another agent's risk-preferences acts on the neural representation of decision-risk. *Proceedings of the National Academy of Sciences, 113*(14), 3755-3760.

Szameitat, D. P., Kreifelts, B., Alter, K., Szameitat, A. J., Sterr, A., Grodd, W., & Wildgruber, D. (2010). It is not always tickling: distinct cerebral responses during perception of different laughter types. *NeuroImage, 53*(4), 1264-1271.

Tachtsidis, I., & Scholkmann, F. (2016). False positives and false negatives in functional near-infrared spectroscopy: issues, challenges, and the way forward. *Neurophotonics*, *3*(3), 031405-031405.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific american, 223*(5), 96-103.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual review of psychology, 33*(1), 1-39.

Tak, S., Uga, M., Flandin, G., Dan, I., & Penny, W. (2016). Sensor space group analysis for fNIRS data. *Journal of neuroscience methods, 264*, 103-112.

Thompson, E. L., Bird, G., & Catmur, C. (2022). Mirror neuron brain regions contribute to identifying actions, but not intentions. *Human brain mapping*, *43*(16), 4901-4913.

Thornton, M. A., & Mitchell, J. P. (2018). Theories of person perception predict patterns of neural activity during mentalizing. *Cerebral cortex*, *28*(10), 3505-3520.

Tsantani, M., Gray, K. L., & Cook, R. (2022). New evidence of impaired expression recognition in developmental prosopagnosia. *cortex*.

Türközer, H. B., & Öngür, D. (2020). A projection for psychiatry in the post-COVID-19 era: potential trends, challenges, and directions. *Molecular Psychiatry, 25*(10), 2214-2219.

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: a formal meta-analysis. *Journal of autism and developmental disorders*, *43*(7), 1517-1526.

Ullman, M. T., & Pullman, M. Y. (2015). A compensatory role for declarative memory in neurodevelopmental disorders. *Neuroscience & Biobehavioral Reviews*, *51*, 205-222.

Uono, S., Yoshimura, S., & Toichi, M. (2021). Eye contact perception in high-functioning adults with autism spectrum disorder. *Autism*, *25*(1), 137-147.

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.

Van der Meer, L., Groenewold, N. A., Nolen, W. A., Pijnenborg, M., & Aleman, A. (2011). Inhibit yourself and understand the other: neural basis of distinct processes underlying Theory of Mind. *NeuroImage*, *56*(4), 2364-2374.

Van Overwalle, F., & Vandekerckhove, M. (2013). Implicit and explicit social mentalizing: dual processes driven by a shared neural network. *Frontiers in human neuroscience, 7*, 560.

Vanman, E. J. (2016). The role of empathy in intergroup relations. *Current Opinion in Psychology*, *11*, 59-63.

Vaucheret Paz, E., Martino, M., Hyland, M., Corletto, M., Puga, C., Peralta, M., Deltetto, N., Kuhlmann, T., Cavalié, D., & Leist, M. (2020). Sentiment analysis in children with neurodevelopmental disorders in an ingroup/outgroup setting. *Journal of autism and developmental disorders*, *50*(1), 162-170.

Wainer, A. L., Ingersoll, B. R., & Hopwood, C. J. (2011). The structure and nature of the broader autism phenotype in a non-clinical sample. *Journal of Psychopathology and Behavioral Assessment, 33*, 459-469.

Wang, L., & Leslie, A. M. (2016). Is implicit theory of mind the 'Real Deal'? The own-belief/true-belief default in adults and young preschoolers. *Mind & Language*, *31*(2), 147-176.

Wang, Y., & Hamilton, A. F. d. C. (2012). Social top-down response modulation (STORM): a model of the control of mimicry in social interaction. *Frontiers in human neuroscience*, *6*, 153.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II).* NCS Pearson.

Wei, M., Su, J. C., Carrera, S., Lin, S.-P., & Yi, F. (2013). Suppression and interpersonal harmony: a cross-cultural comparison between Chinese and European Americans. *Journal of counseling psychology*, *60*(4), 625.

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.

Wheelwright, S., Auyeung, B., Allison, C., & Baron-Cohen, S. (2010). Defining the broader, medium and narrow autism phenotype among parents using the Autism Spectrum Quotient (AQ). *Molecular autism*, *1*(1), 10.

White, S. J., Coniston, D., Rogers, R., & Frith, U. (2011). Developing the Frith-Happé animations: A quick and objective test of Theory of Mind for adults with autism. *Autism Research*, *4*(2), 149-154.

White, S. J., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the strange stories: Revealing mentalizing impairments in autism. *Child development*, *80*(4), 1097-1117.

Willey, L. H. (2014). *Pretending to be normal: Living with asperger's syndrome (autism spectrum disorder) expanded edition*. Jessica Kingsley Publishers.

Wilson, C. E., Palermo, R., Burton, A. M., & Brock, J. (2011). Recognition of own-and other-race faces in autism spectrum disorders. *The Quarterly Journal of Experimental Psychology*, *64*(10), 1939-1954.

Woo, B. M., Tan, E., Yuen, F. L., & Hamlin, J. K. (2023). Socially evaluative contexts facilitate mentalizing. *Trends in cognitive sciences*, *27*(1), 17-29.

Wood-Downie, H., Wong, B., Kovshoff, H., Mandy, W., Hull, L., & Hadwin, J. A. (2021). Sex/gender differences in camouflaging in children and adolescents with autism. *Journal of autism and developmental disorders*, *51*, 1353-1364.

World Health Organization. (2018). *International classification of diseases for mortality and morbidity statistics (11th Revision)*. https://icd.who.int/

Xie, Y., Zhong, C., Zhang, F., & Wu, Q. (2019). *The ingroup disadvantage in the recognition of micro-expressions* 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France.

Ye, J. C., Tak, S., Jang, K. E., Jung, J., & Jang, J. (2009). NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy. *NeuroImage*, *44*(2), 428-447.

Yi, L., Quinn, P. C., Fan, Y., Huang, D., Feng, C., Joseph, L., Li, J., & Lee, K. (2016). Children with Autism Spectrum Disorder scan own-race faces differently from other-race faces. *Journal of experimental child psychology*, *141*, 177-186.

Yi, L., Quinn, P. C., Feng, C., Li, J., Ding, H., & Lee, K. (2015). Do individuals with autism spectrum disorder process own-and other-race faces differently? *Vision research*, *107*, 124-132.

Young, S. G. (2017). An outgroup advantage in discriminating between genuine and posed smiles. *Self and Identity*, *16*(3), 298-312.

Young, S. G., & Hugenberg, K. (2010). Mere social categorization modulates identification of facial expressions of emotion. *Journal of personality and social psychology*, *99*(6), 964.

Young, S. G., Slepian, M. L., & Sacco, D. F. (2015). Sensitivity to perceived facial trustworthiness is increased by activating self-protection motives. *Social Psychological and Personality Science*, *6*(6), 607-613.

Yücel, M. A., Selb, J., Aasted, C. M., Lin, P.-Y., Borsook, D., Becerra, L., & Boas, D. A. (2016). Mayer waves reduce the accuracy of estimated hemodynamic response functions in functional near-infrared spectroscopy. *Biomedical optics express*, *7*(8), 3078-3088.

Zadeh, A., Chong Lim, Y., Baltrušaitis, T., & Morency, L.-P. (2017). Convolutional experts constrained local model for 3d facial landmark detection. [IEEE Xplore All Conference Series]. Proceedings of the IEEE International Conference on Computer Vision Workshops,

Zaitchik, D., Walker, C., Miller, S., LaViolette, P., Feczko, E., & Dickerson, B. C. (2010). Mental state attribution and the temporoparietal junction: an fMRI study comparing belief, emotion, and perception. *Neuropsychologia*, *48*(9), 2528-2536.

Zhang, F., & Roeyers, H. (2019). Exploring brain functions in autism spectrum disorder: A systematic review on functional near-infrared spectroscopy (fNIRS) studies. *International Journal of Psychophysiology*, *137*, 41-53.

Zhang, H., Yang, J., Ni, J., De Dreu, C. K., & Ma, Y. (2023). Leader–follower behavioural coordination and neural synchronization during intergroup conflict. *Nature Human Behaviour*, 1-13.

Zimeo Morais, G. A., Balardin, J. B., & Sato, J. R. (2018). fNIRS Optodes' Location Decider (fOLD): a toolbox for probe arrangement guided by brain regions-of-interest. *Scientific reports*, *8*(1), 3341.