THE IMPACT OF PREMATURITY ON SOCIAL UNDERSTANDING

By

Emma Elizabeth Covernton

A THESIS SUBMITTED TO THE UNIVERSITY OF BIRMINGHAM FOR THE DEGREE OF DOCTOR OF CLINICAL PSYCHOLOGY

Department of Clinical Psychology

School of Psychology

The University of Birmingham

June 2023

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Thesis overview

This thesis comprises one volume with two chapters and is submitted by Emma Elizabeth Covernton to fulfil the requirements of the Clinical Psychology Doctorate programme. The first chapter is a systematic review and meta-analysis of studies measuring differences in social cognition of children born prematurely compared to their full-term peers. The key outcome of analysis was that children born at earlier gestational ages present with differences in affect recognition and theory of mind abilities, with children born prematurely performing worse in these social cognitive domains.

Chapter two presents an empirical pilot study with two parts. Part One investigates the association between implicit and explicit measures of social cognition, using eye tracking methodology (implicit measurement) and behavioural observational methods (explicit measurement). The aim of part one was to estimate the effect size of the correlation between implicit and explicit measures of social cognition. Findings suggest large correlations between measures are unlikely when measured cross-sectionally.

Part Two investigates differences in social cognition between a small group of children born moderately to late preterm compared to age matched full-term children. The aim of part two was to estimate the effect size of differences between groups on measures of social preference and skills considered precursors of theory of mind abilities. Findings suggest important differences may exist in social preference between children born moderate to late preterm compared to their full-term peers.

Acknowledgements

There are many to thank for the product of this thesis. Special gratitude to all of the families and children who spent their time participating in this research and to all of the students who supported with data collection. I am incredibly thankful to both of my research supervisors – it was so great to have the opportunity of your expertise and guidance throughout my three years. Thank you for your compassionate support and endless re-reading of my drafts! Finally, a special thank you to Catherine – your endless support and guidance through the research process has been invaluable, and I am forever grateful.

Dedication

To Luke – I could not have got through this without you. Thank you for everything you have done to support me through these difficult three years and thank for you all you continue to do.

Table of contents

CHAPTER ONE:

Abstract……1

Introduction……2

Social Development……..……..2

Outcomes of Preterm Birth……..5

Existing Reviews: Prematurity and Social Development…….……5

Methodology……12

Registration of Systematic Review……..……..12

Search Strategy……...……12

Study Selection……...…….……13

Data Extraction……...……17

Quality Assessment ……...…….…18

Statistical Analysis ……...……..…25

Structure of Results……...…….……27

Results……...29

Meta-Analysis Results ……..…….....48

***Social Preference*……50**

***Responding to Joint Attention (Gaze Following*) …….…….51**

***Responding to Joint Attention (Gaze and Point*) ……..….52**

***Initiating Joint Attention*……..……...54**

***Affect Recognition*……...……56**

***Theory of Mind*……58**

Narrative Review……...…….60

Social Attention and Perception ……..…….……...…...61

***Face Recognition*……...……..61**

***Imitation*……....61**

***Face and Eye Preference*……...……..61**

***Biological Motion Perception*……...……...62**

Joint Attention……...……63

Social Reasoning and Empathy……..……...63

Discussion……68

Early Social Attention……..…….….68

Joint Attention……...…….69

Affect Recognition. ……70

Theory of Mind……...……71

Social Reasoning and Empathy……..……...72

Methodological Limitations……..……72

***Limitations of Included Studies*……...……72**

***Limitations of Current Meta-Analysis*…….……....74**

Clinical Implications……...75

Future Directions ……...……76

Conclusions ……77

References……78

CHAPTER TWO:

Abstract……...……104

Introduction……..……105

Implicit and Explicit Social Cognition…….……...106

Prematurity……...107

Moderate to Late Prematurity ……..……..107

Social Development and Moderate to Late Prematurity……..108

Problems with Current Research ……..109

The Early Social Cognition Battery……....111

Rationale……...111

The Current Study……...……112

Objectives and Hypotheses……...……...112

Methodology ……114

Ethical Approval ……...……...114

Recruitment Strategy……...……114

Participants ……..114

Materials ……...120

Procedure ……128

Data Processing and Statistical Analysis ……132

Results……136

Preliminary Analysis ……...……136

Part One Results ……...……139

Part Two Results ……..…...144

Discussion……...148

Preliminary Analysis ……...……....148

Part One ……...……149

Part Two ……...……...150

Tolerability and Feasibility of the Protocol ……152

Strengths and Limitations ……...……153

Clinical Implications and Considerations ……154

Recommendations for Future Studies ……...….155

Conclusions……...156

References……...……157

PRESS RELEASE……...172

Chapter One……..……172

Chapter Two……..…174

Appendices

CHAPTER ONE

Appendix 1: Appendix showing completed PRISMA checklist.

Appendix 2: Appendix showing QQ plots for the distribution of standardised mean differences for both random and fixed effect models for studies measuring responding to joint attention (gaze following), responding to joint attention (point + gaze), initiating joint attention, theory of mind (after removal of Marleau et al., 2020) and affect recognition.

A. Social preference
B. Responding to joint attention (gaze following)
C. Responding to joint attention (gaze + point)
D. Initiating joint attention
E. Affect recognition.
F. Theory of mind

Appendix 3: Appendix showing figure of funnel plot of the standardised mean difference in responding to joint attention (gaze following), responding to joint attention (gaze + point), initiating joint attention, affect recognition and theory of mind (post removal of outlier) between preterm and full-term children. The 95% confidence interval of the expected distribution of effect is shown as an inverted "funnel".

A. Social preference
B. Responding to joint attention (gaze following)
C. Responding to joint attention (gaze + point)
D. Initiating joint attention
E. Affect recognition.
F. Theory of mind (prior to trim and fill procedure)

Appendix 4: Appendix showing figures of Baujat scatter plots for studies measuring responding to joint attention (gaze following), responding to joint attention (point + gaze) and initiating joint attention.

A. Responding to joint attention (gaze following)
B. Responding to joint attention (gaze + point)
C. Initiating joint attention

Appendix 5: Appendix showing figures of leave one out analysis for studies measuring responding to joint attention (gaze following), responding to joint attention (gaze + point) and initiating joint attention.

A. Responding to joint attention (gaze following)
D. Responding to joint attention (gaze + point)
B. Initiating joint attention

Appendix 6: Appendix showing theory of mind analyses prior to removal of outlier and correction.

A. Figure showing QQ plots for both fixed and random effects model for theory of mind meta-analysis (prior to removal of Marleau et al., (2020).
B. Figure showing Baujat scatterplot for theory of mind meta-analysis (prior to removal of Marleau et al., 2020).
C. Figure showing outlier and sensitivity analysis via leave one out method for theory of mind meta-analysis (prior to removal of Marleau et al., 2020)
D. Figure showing funnel plot of the standardised mean differences in theory of mind between preterm and full-term children prior to trim and fill procedure. The 95% confidence interval of the expected distribution of differences in theory of mind is shown as an inverted “funnel”.

Appendix 7*:* Appendix of tables showing the impact of study level effects for studies measuring performance on responding to joint attention (gaze following), responding to joint attention (gaze + point), initiating joint attention, affect recognition and theory of mind between preterm and full-term children.

A. Social preference
B. Responding to joint attention (gaze following)
C. Responding to joint attention (point + gaze)
D. Initiating joint attention
E. Affect recognition.
F. Theory of mind

CHPATER TWO

List of appendices

Appendix 1: Appendix showing ethical approval for project.

Appendix 2: Appendix showing Health and background questionnaire.

Appendix 3: Appendix showing Modified Checklist for Autism in Toddlers-Revised.

Appendix 4: Appendix showing table of the Early Social Cognition Scale – full description of administrative discrepancies.

Appendix 5: Appendix showing contractual agreement for use of the Geometric Preference task.

Appendix 6: Appendix showing contractual agreement for use of the biological motion task.

Appendix 7: Appendix showing the Early Social Cognition Scale – description of task counterbalancing.

Appendix 8: Appendix of table of group scaling and participant characteristics according to the Early Social Cognition Scale compared to Mullen composite score and age for both moderate to late preterm children and full-term children.

A. Moderate to late preterm children.

B. Full-term children.

Appendix 9: Appendix of table showing group scaling and participant characteristics according to the Early Social Cognition Scale compared to Mullen composite score and for both moderate to late preterm children and full-term children demonstrating **other** patterns of performance on the Early Social Cognition Scale.

A. Moderate to late preterm children.

B. Full-term children.

List of Illustrations

CHAPTER ONE

Figure 1: Figure depicting a developmental perspective to the acquisition of early social cognitive skills - from Soto-Icaza et al., 2015 – open access journal with permission to use with citation.

Figure 2: Figure showing PRSIMA flowchart.

Figure 3: Figure showing forest plots for social preference studies.

Figure 4: Figure showing forest plots for responding to joint attention (gaze following) studies.

Figure 5: Figure showing forest plots for responding to joint attention (gaze + point) studies.

Figure 6: Figure showing forest plots for initiating joint attention studies.

Figure 7: Figure showing forest plots for affect recognition studies.

Figure 8: Figure showing forest plots for theory of mind studies.

Figure 9: Figure showing funnel plot of the standardised mean differences in theory of mind between preterm and full-term children *after* trim and fill procedure. The 95% confidence interval of the expected distribution of differences in theory of mind is shown as inverted.

CHAPTER TWO

Figure 1: Figure showing the four-stage scale of early social cognitive skills -from Ellis et al., 2020 - open access journal with permission to use with citation.

Figure 2: Figure depicting Geometric Preference task stimuli from Pierce et al., (2011) – permissions to use as per contractual agreement.

Figure 3: Figure depicting stimuli biological motion task (condition one and condition two) - from Annaz et al., (2012) – permissions to use as per contractual agreement.

(a) Biological motion walker.

(b) Scrambled.

(c) Spinning top.

Figure 4: Figure depicting infant calibration image taken from SR Research LTD - Image open access through SR Research LTD.

Figure 5: Figure depicting order of presentation of eye tracking stimuli – permissions to as per contractual agreement.

Figure 6: Figure showing scatterplot of social preference measures correlated with the Early Social Cognition Scale.

A. Geometric Preference task and Early Social Cognition Scale.
B. Biological motion task (condition one) and Early Social Cognition Scale.
C. Biological motion task (condition two) and Early Social Cognition Scale.

Figure 7: Figure showing scatterplot of social preference tasks and the Early Social Cognition. Scale scores with line of best fit for moderate to late preterm children and full-term children.

A. Geometric Preference task and the Early Social Cognition Scale.
B. Biological motion task (condition one) and the Early Social Cognition Scale.

List of Tables

CHAPTER ONE

Table 1: Table summarising existing systematic, narrative, and meta-analytic reviews of social development and prematurity and definitions of social outcomes employed. Reviews have been organised by date of review (oldest to most recent).

Table 2: Table describing search terms used to search MEDLINE, Embase, PsycINFO, PubMed, Networked Digital Library of Thesis and Dissertations, open access thesis and dissertations and ProQuest, from database creation to February 2022.

Table 3: Table describing stage one inclusion/exclusion criteria for title and abstract screening.

Table 4: Table describing stage two Inclusion/exclusion criteria for full text screening.

Table 5: Table describing quality assessment framework and scoring.

Table 6: Table describing studies grouped by way of social cognitive domain and number of studies progressing to meta-analysis and narrative review.

Table 7: Table describing individual study characteristics, key demographics of preterm and full-term children and individual study quality ratings. Table organised by social cognitive domain.

Table 8: Table summarising participant numbers and important demographics of preterm and full-term children within each meta-analysis.

Table 9: Table showing the impact of study level quality assessment for studies measuring responding to joint attention (gaze following).

Table 10: Table showing the impact of study level quality assessment for studies measuring responding to joint attention (point +gaze)

Table 11: Table showing the impact of study level quality assessment for studies measuring initiating joint attention.

Table 12: Table showing the impact of study level quality assessment for studies measuring affect recognition.

Table 13: Table summarising outcomes across all social cognitive domains included in systematic review and meta-analysis.

CHAPTER TWO

Table 1: Table describing important participant demographics, health, birth, and family characteristics according to analyses.

Table 2: Table summarising the Early Social Cognition Scale tasks, original authors of tasks, and description of individual task procedures, as described in Ellis et al., (2020).

Table 3: Table showing individual pass rates on the Early Social Cognition Scale for all participants in comparison to Ellis et al., (2020) typically developing sample.

Table 4: Table showing group scaling and participant characteristics according to the Early Social Cognition Scale compared to Mullen composite score and age.

Table 5: Table showing group scaling and participant characteristics according to the Early Social Cognition Scale compared to Mullen composite score and age, for participants demonstrating **other** patterns of performance on the Early Social Cognition Scale.

Table 6: Table showing mean dwell time and percentage dwell time to social images across the three eye tracking tasks for all participants (Geometric preference task, biological motion task – condition one and biological motion task – condition two).

Table 7: Table showing individual participant characteristics across measures, including sex, chronological age, corrected age, Mullen composite scores, Modified Checklist for Autism in Toddlers-Revised scores, Early Social Cognitive Scale scores, percentage time viewing social area of interest and number of trials viewed for each eye tracking paradigm.

Table 8: Table showing Early Social Cognitive Scale individual task pass rates for moderate to late preterm children and full-term children.

Other Preliminary Listings

CHAPTER ONE: abbreviations

GA: gestational age

PT: preterm (children)

FT: full-term (children)

RJA: Responding to joint attention.

IJA: Initiating joint attention

ToM: Theory of mind

CHAPTER TWO: abbreviations

GA: gestational age

MPT: moderate to late preterm (children)

FT: full-term (children)

ToM: Theory of mind

ESCogS: Early Social Cognition Scale

GeoPref: Geometric preference task

BioMotion (C1): Biological motion task – condition one

BioMotion (C2): Biological motion task – condition two

CHAPTER ONE:

DO CHILDREN BORN PRETERM PRESENT WITH DIFFERENCES IN SOCIAL COGNITION AS COMPARED TO THEIR FULL-TERM PEERS?

Abstract

OBJECTIVES: This systematic review and meta- analysis aimed to investigate whether children born prematurely present with differences in social cognition as compared to their full-term peers.

METHODS: All literature from start of database creation to February 2022 was searched for studies measuring social cognition in preterm children and adolescence (age 0-18 years) with a full-term comparison group. A bespoke quality assessment tool was developed to assess quality of included studies. Results were synthesised by way of social cognitive domain.

RESULTS: Six separate meta-analyses were possible for early social preference, joint attention behaviours (gaze following, responding to joint attention – point +gaze behaviours and initiating joint attention), affect recognition and theory of mind. Studies not possible to meta-analyse were retained for narrative review and categorised into the social cognitive domains of early social attention/perception and, social reasoning and empathy. Results suggest preterm children born at earlier gestational ages present with differences in affect recognition and theory of mind abilities, demonstrating fewer of these skills when compared to full-term children. The results of earlier developing social cognitive domains remain inconclusive.

DISCUSSION: This paper is the first to quantify differences in social cognition using meta-analytic methods beyond joint attention skills. The trajectory of social cognitive development in children born prematurely is currently unclear. Identification of prematurity as a risk factor for social cognitive differences suggests preterm children may benefit from increased professional attention.

Introduction

Preterm birth, defined as birth occurring at a gestational age (GA from here onwards) of less than 37 weeks (Howson et al., 2013) affects approximately 11% of births worldwide (Vogel et al., 2018). Within the broader category of prematurity, subcategories are recognized, such as extremely preterm (<28 weeks GA), very preterm (28-<32 weeks GA) and moderate preterm (32-<37 weeks GA), with the latter also further sub-divided into late preterm birth (34-<37 weeks GA; Blencowe et al., 2013). Rates of survival from preterm birth are increasing, particularly in higher income countries (Blencowe et al., 2012; Vogel et al., 2018). With advancements in neonatal medical practices, mortality continues to decline (UN Children's Fund, 2017), with many infants born prematurely surviving into adulthood. The growing number of children, young people and adults born prematurely highlights the importance of understanding the potential ongoing consequences of preterm birth. An area of particular interest relates to the interactions between social development and premature birth.

Social Development

Research on social development employs a diverse range of definitions and measurement approaches (Rose-Krasnor, 1997; Rubin et al., 1995; Yeates et al., 2007). The transactional nature (contingent on individual child characteristics *and* interactions with others in their social world) and the impact of time and context (developmental nature) of social development has perhaps contributed to the diversity of definitional constructs (Rubin & Krasnor, 1986; Rubin & Rose-Krasnor, 1992; Yeates et al., 2007). Taylor et al. (2022) summarises a hierarchical three-component model of *social competence* which draws on theoretical understandings reported by Yeates et al. (2007). This model suggests that the construct of social competence includes social cognition, social interaction, *and* social adjustment which represent distinct, yet related components relevant to understanding social development. Each part of the model is a building block representing different *levels* of social

complexity (the hierarchical nature of the model), and all parts influence and are influenced by one another across development. The theory underpinning this model has support from other theories whereby social information processing has been conceptualised as a critical determinant of social competence (Crick & Dodge, 1994; Rubin & Krasnor, 1986; Yeates et al., 2007), but also research suggesting social information processing, social interactions and social adjustment are closely associated (Crick & Dodge, 1994; Parker et al., 2006; Rubin et al., 2006; Wacaldo & Rieger, 2006; Yeates et al., 2007). Explicit definitions are important to be able to synthesise findings in a research area which has employed varied definitions of related constructs (as shown in recent reviews such as Dean et al., 2021). The three- component model of social competence (Yeates et al., 2007) has been helpful in the study of other divergent developmental trajectories (Yeates, 2000; Yeates et al., 2007) and distinguishing among different levels of social complexity can allow for a better understanding of pathways to pathology (Crick & Dodge, 1994). This paper focuses on *social cognition* to summarise current knowledge on this level of social complexity in preterm infants as compared to their full-term peers.

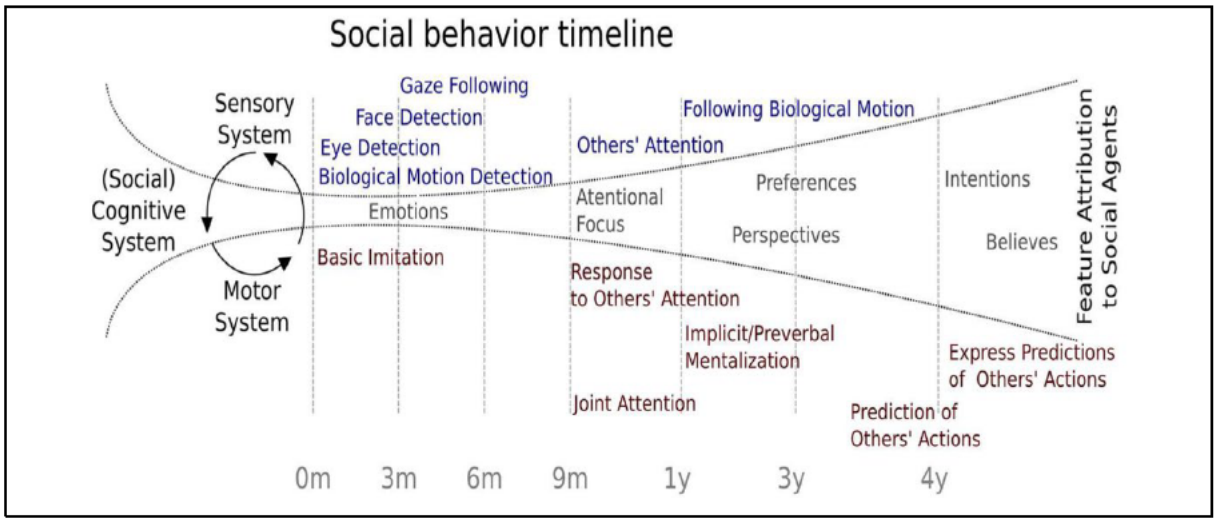
Social cognition is a broad term which encompasses many different processes by which individuals understand and make sense of the world (Frith, 2008). This includes the mental processes involved in social interactions (Taylor et al., 2022). Such processes may therefore include social attention (Pavlova, 2011), affect perception, emotion processing and theory of mind (Yeates et al., 2007; Taylor et al., 2022). Social cognition develops throughout childhood and adolescence, therefore adopting a developmental perspective can be helpful to understand how social understanding evolves (Alaerts et al., 2011; Rao et al., 2008; Soto-Icaza et al., 2015). Figure 1 describes a developmental perspective of the acquisition of earlier social cognitive skills. This figure highlights the social behaviour milestones during the first five years of childhood (Soto-Icaza et al., 2015) and the inter-related nature of social-cognitive skill

development. Additionally, it demonstrates how the development of social cognitive skills is closely tied to the development of other processes such as motor and sensory skills.

Joint attention is a crucial element of *social interaction* and is suggested to be related to empathy, as the sharing of representational worlds may benefit and promote its development (Frith & Frith, 2007). Joint attention is sometimes considered as a precursor to social cognition and not necessarily an *outcome* of the development of social cognition (Mundy & Newell, 2007). These skills have also been suggested to underpin other higher order competencies such as theory of mind (Nelson et al., 2008). Additionally, the ability to imitate the actions of another has been suggested as a foundational skill for the development of theory of mind and empathy for others (Meltzoff & Decety, 2003). Under the three-component model as described, imitation could arguably fall within social interaction, as to imitate, would be a response to a social partner, however as with joint attention, it may also underpin other social cognitive skills (Meltzoff & Decety, 2003). Therefore, it may be helpful to include both joint attention and imitation under the umbrella of social cognition.

Figure 1

Figure depicting a developmental perspective to the acquisition of early social cognition - from Soto-Icaza et al., 2015 - open access journal with permission to use with citation.



Outcomes of Preterm Birth

Studies exploring sequalae of preterm birth more broadly have suggested increased adverse risks across many areas of child development (Taylor et al., 2022). Risks include cognitive deficits, such as lower intellectual functioning, impairments in motor development, early language development, processing speed and executive functioning (Aarnoudse-Moens et al., 2009; Bhutta et al., 2014; Bos et al., 2013; Capobianco & Cerniglia., 2017; Hirschberger, 2018; Ream & Lehwald, 2018; Serenius et al., 2016). Additionally, a tendency toward internalising traits such as anxiety and depression (Johnson & Marlow, 2011), and an increased risk of psychiatric disorders such as anxiety and neurodevelopmental difference such as attention deficit hyperactivity disorder – ADHD are suggested (Bhutta et al., 2002; Johnson et al, 2010; Norsarti et al., 2012; Treyvaud et al., 2013). Heightened rates of autism are also observed in preterm populations, as compared to full-term peers (Agarwal et al., 2018; Fenoglio et al., 2018; Guo et al., 2022; Johnson et al., 2010; Laverty et al., 2021; Singh et al., 2013). Furthermore, indidivuals born prematurely may be more likely to have difficulties *consistent* with an autism diagnosis without meeting diagnostic cut-offs (Laverty et al., 2021). Differences in social communication and functioning represent a key feature for a diagnosis of autism and as such increased research exploring the impact of prematurity on different social outcomes is unsurprising.

Existing Reviews: Prematurity and Social Development

Several existing reviews explore different aspects of social development in children born prematurely, a summary of which is presented in Table 1. Focusing on children born *very preterm* (<32 weeks GA), Montagna and Nosarti's (2016) systematic review included broader social-emotional skills, as well as social cognition and described "sub-optimal functioning" in early social attention, joint attention, emotion recognition and theory of mind in children born preterm compared to those born at term (Montagna & Nosarti, 2016). However, this review

also highlighted the very few studies available at the time. Ritchie et al.'s, (2015) systematic review focused on social development in children born *very preterm,* (<33 weeks GA), with studies suggesting increased risk of social adjustment difficulties such as social withdrawal and peer relationship problems. Increased risk was suggested to be across development from infancy to early adolescence.

Systematic review including children across *all* GAs of children born prematurely suggest poor social/interactive skills as a definitive part of the behavioural profile (Arpi & Ferrari, 2013). The review also demonstrated that these differences in social and interactive skills are present across preterm GA and not limited to those children born at earlier gestations. Differences between preterm and term-born children have also been described in social interaction and joint attention (Fenoglio et al., 2017; Zmyj et al., 2017), in addition to atypical social-emotional development, increased internalizing behaviours, dysregulation and impairments in gaze following and early social attention (Fenoglio et al., 2017). Johnson et al., (2018) more recently suggested a "preterm phenotype" which encompasses increased risk for cognitive, attention and social-emotional differences which may expand beyond earlier GAs to include children born moderately to late preterm (32 weeks$^{+0/6 days}$ – 36 weeks$^{+6/7 days}$ GA).

Most recently Dean et al. (2021) synthesised a large body of literature assessing the nature of social cognition following preterm birth between 1995 and 2017 across *all* GAs (Dean et al., 2021). This review showed great irregularity in published results due to considerable variability in outcome assessed, choice of assessment method and varying definitions of prematurity. This systematic review was important in identifying and summarising research in this area, by including a very broad definition of social cognition, in addition to the inclusion of attachment and diagnostic measures of autism. However, a lack of distinction among different levels of social complexity (as suggested by Yeates et al., 2007 as important when studying social competence), in addition to limits placed on searches (year of study and age

range 0-5 years) resulted in a systematic review that could not quantify described differences through meta-analysis.

Despite the number of systematic and narrative reviews, only one study has progressed to meta-analysis to quantify social cognitive differences in preterm children. Mateus et al. (2019) explored differences in a defined subdomain of social cognition, joint attention, between pre-term and full-term children and found no significant differences in responding to joint attention, initiating joint attention and joint attention episodes. The authors highlighted the poor quality of the studies available at the time (searches conducted in 2016), in addition to the need for an updated review once more research was available. This study is important in highlighting the emerging possibility of meta-analytic methods in this area of research to quantify the size and nature of differences described in narrative reviews. However, the focus on one subset of social cognition in isolation limits broader interpretation of the findings. A more comprehensive approach to meta-analysis, considering the breadth of social cognitive skills, is important due to the inter-related nature of social cognitive development.

Across all reviews described, major limitations in the literature were highlighted. These included definitional issues, with social competence used interchangeably with broader socio-emotional and behavioural adjustment domains (Ritchie et al., 2015). Methodological limitations were also identified, including the large variation in outcomes assessed and tools used (Ritchie et al., 2015), There was a predominance of studies using caregiver report for assessment of social development (Montagna & Nosarti, 2016; Ritchie et al., 2015) which may over-estimate differences in comparison to experimental methods. Finally, many studies lacked comparisons to full-term peers making interpretation as to true differences between groups more difficult to capture and a clear profile of social strengths and weaknesses difficult to delineate (Ritchie et al., 2015).

Fenoglio et al. (2017) and Zmyj et al. (2017) further identified an important problem in the research in that preterm children's *social cognitive abilities* compared to their full-term peers have received much less attention and should be investigated more intensely. Additionally, the focus of the majority of reviews on only earlier gestations of prematurity limits the generalisability of interpretation to later gestational ages, which represent a larger proportion of this cohort (Boyle et al., 2015; Moser et al., 2007). Existing reviews focus on 0-5 age range, with only one extending beyond this (refer to Table 1 for additional detail). The acquisition of social cognitive skills spans development beyond early childhood to adolescence where environmental and biological changes typically result in new social encounters which rely on earlier foundational social cognitive skills in the evaluation of social stimuli/situations (Choudhury et al., 2006). Earlier divergent trajectories may influence later developing skills (Masten & Cicchetti, 2010) and so a more holistic understanding of the longer-term trajectory in those born prematurely is important.

Rationale

The complexity of social development research means there are many ways of conceptualising, defining, and measuring constructs of interest. The number of studies exploring aspects of social development in individuals born prematurely evidence this is an area of interest. A fundamental limitation of existing reviews is lack of a clear definition of terms, which is essential when attempting to synthesise data in a way in which makes sense theoretically and may be suitable for meta-analysis. The vast number of studies with varying methodologies and definitional constructs identified by Dean et al.'s (2021) recent synthesis perhaps suggests a narrower definition of social cognition might be helpful to synthesise data in a way in which makes sense theoretically and may be suitable for meta- analysis.

Following identification of a deficit in research in social cognition and prematurity and subsequent call for further studies in this area (Fenoglio et al., 2018; Zymj et al., 2017) there is

a clear rationale for a further review, synthesising, and meta-analysing emerging studies in the area of *social cognition*. Existing reviews (Dean et al., 2021; Fenoglio et al., 2018; Mateus et al., 2019) include studies from 1994-2017 and since this time there has been an increased focus on further exploration of outcomes related to social development, re-examining this area and assessing suitability for the application of meta-analytic methods for synthesis is therefore of great value.

This systematic review and meta- analysis aimed to investigate whether children born prematurely present with differences in social cognition as compared to their full-term peers in the following way:

1. Synthesise *all* existing literature using a clearly defined construct of social cognition, by extending date restriction parameters to studies pre 1994 to present day and extending the age range of participants from 0-18 years.
2. Combine studies measuring broad social cognitive abilities across domains, extending beyond joint attention. This will include meta-analysis where possible and providing a narrative review for papers meeting inclusion criteria which are not appropriate for meta-analytic methods.
3. Identify participant characteristics that may be associated with differences in social cognition in preterm samples compared to full-term peers.

Table 1

Table summarising existing systematic, narrative, and meta-analytic reviews of social development and prematurity and definitions of social outcomes employed. Reviews have been organised by date of review (oldest to most recent).

Author[1]	Time span of review	Term Comparison ? Y/N	Category of PT[2]	Age range	Types of studies included	Definition of social development
Arpi & Ferrari (2013)	2000-2012	Y	All	0-5	Observational & informant report	Evaluated behavioural outcomes which included social/interactive skills but did not explore this exclusively.
Ritchie et al. (2015)	1991-2014	Y	VPT	0-17	Observational & informant report	Used the three-component model (Yeates et al., 2007) to organize a review of those born at earlier gestations and included all three areas (social cognition, social interaction, and social adjustment). Focus of review was broad and included *all* three aspects of social development as listed above.
Mateus et al. (2019)	1970-2016	Y	All	9-24 months	Observational	This meta-analysis focused exclusively on responding to joint attention, initiating joint attention and joint attention episodes.
Dean et al. (2021)	1995-2017	Both	All	0-5	Observational	Employed a broad definition of social cognition which included attachment paradigms if they reported specific social cognitive behaviours (e.g., gaze, vocalisation or motor engagement behaviours), autism diagnostic tools, early social attention, joint attention, emotion and belief recognition.
Narrative Reviews without clearly specified criteria						
Montagna and Nosarti (2016)	Not stated	Not specified	VPT	0-18	Not specified	Summarised boarder socio-emotional behavioural problems, of which 'problems with peers', is relevant to social competence. The paper goes on to summarise studies in the areas of joint attention, emotion recognition, emotion regulation and social

[1] Authors in italics and bold indicate meta-analysis was conducted. All other authors in normal font are systematic reviews.
[2] PT = prematurity.

Author[1]	Time span of review	Term Comparison ? Y/N	Category of PT[2]	Age range	Types of studies included	Definition of social development
						cognition (to include early social attention, biological motion perception and theory of mind).
Zymj et al. (2017)	Not specified	Not specified	All	Not specified	Not specified	This review broadly summarises studies measuring joint attention, theory of mind & interaction difficulties with others.
Fenoglio et al. (2017)	Not specified	Both	All	Not specified	Not specified	Summarised social cognition and brain circuits involved in the processing social information. Included joint attention, reciprocal social behaviours and gaze following.
Taylor (2020)	Not specified	Not specified	VPT	Not specified	Not specified	Summarised studies organised by three-component model by Yeates et al. (2007) and included all three areas of social cognition, social interaction, and social adjustment.

Please note *VPT = very preterm; <32 weeks GA.*

Method[3]

Registration of Systematic Review

Before a search was undertaken, the present study was preregistered on the international prospective register of systematic reviews (PROPSERO). Available at: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=229456.

Search Strategy

The identification, screening and data extraction of eligible studies was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses – PRISMA (Liberati et al., 2009). See Appendix 1 for PRISMA checklist. A comprehensive electronic search strategy was used to identify all studies indexed in MEDLINE, Embase, PsycINFO, PubMed, Networked Digital Library of Thesis and Dissertations, open access thesis and dissertations and ProQuest, from the beginning of database creation to February 2022. Search terms were generated by examining terms used in existing reviews in the research area (Dean et al., 2021; Mateus et al., 2019; Ritchie et al., 2015). From this, a comprehensive electronic search strategy was devised with the aim of identifying all studies involving assessment of social cognition in children born prematurely. Search terms were intentionally broad to allow for identification of all relevant studies. This was considered of particular importance due to the heterogeneity of assessment methods and terms used in the literature on social cognition (Dean et al., 2012). Prior to undertaking the search, the term "eye tracking" was added as during preliminary searching, further relevant studies were located that deployed eye tracking methods. See Table 2 for full list of search terms employed.

Three recent systemic reviews were hand searched to identify any further publications (Dean et al., 2021; Fenoglio et al., 2018; Mateus et al., 2019). Additionally, reference lists of

[3] From here onwards acronyms have been used to describe the groups of interest in this review. To support the reader these are as follows; PT = preterm; FT = full-term.

studies meeting the inclusion criteria were hand searched to identify further eligible studies. Three additional studies were identified through hand searching.

Table 2

Table describing search terms used to search MEDLINE, Embase, PsycINFO, PubMed, Networked Digital Library of Thesis and Dissertations, open access thesis and dissertations and ProQuest, from database creation to February 2022.

	Search terms
#1Social cognition	"Social cognition"; "social development"; "social learning"; "social interaction"; "social understanding"; "eye tracking"*; "theory of mind"; ("mentalising" OR "mentalizing"); "empathy"; ("mind reading" OR "mindreading");"joint attention"; "shared attention"; "shared intentionality"
	"AND"
#2Prematurity	"preterm"; "premat*"; "low birth weight"; "LBW".

Study Selection

Included studies reported on directly observed measures of *social cognition* in PT children born across all gestational ages, in both published and unpublished studies. Publication bias can be a concern when conducting a systematic review and meta-analysis (Field & Gillett, 2010; Greenland, 1994; Rosenthal, 1979), with studies reporting statistically significant results more likely to be accepted for publication (Coursol & Wagner, 1986). Therefore, inclusion of non-published work was considered important. Two recent reviews in this area reviewed studies after 1995, due to changes in maternal-fetal medicine and neonatal care practices and changes in diagnostic criteria for diagnosis of autism (Dean et al., 2021; Mateus et al., 2017). The current review chose to expand this time frame and explore any differences related to year of publication empirically if data permitted.

Study selection included three stages. One reviewer (the author) undertook electronic database searches. Relevant articles were initially screened by title and abstract and were sorted in accordance with a predetermined inclusion/exclusion criteria detailed in Table 3.

Table 3.

Table describing stage one inclusion/exclusion criteria for title and abstract screening.

Inclusion criteria	Exclusion criteria
Empirical studies	Conference abstracts, conference papers, literature reviews or book chapters
Full text published or available in English	Full text not available in English
Human participants	Studies with animal participants
Title/abstract has social cognition reported/references categories which may be associated with social cognition such as social interaction/competence for example.	Title/abstract does not refer to social cognition or reports categories which may be associated with social cognition such as social interaction/competence for example.
Title or abstract has prematurity reported/references.	Title/abstract does not report prematurity.
Title or abstract refers to sample as low birth weight	Title/abstract does not refer to sample as low birth weight (LBW)
Title/abstract reporting attachment and prematurity kept to carefully examine whether aspects of infant behaviour relevant to social cognition were measured	Title/abstract reports only participants who are the *caregivers* of children born prematurely and not the preterm child/adolescent.

At full text screening, studies were carefully examined and selected in accordance with a second stage of inclusion/exclusion criteria, full details provided in Table 4 below.

Table 4

Table describing stage two Inclusion/exclusion criteria for full text screening.

Inclusion criteria	Exclusion criteria
Participants aged between 0-18 years at time of assessment.	Participants over the age of 18 years at time of assessment.
Studies including samples with comorbidities so as not to flatten the likely variability in preterm cohorts.	Studies where children have received a diagnosis of developmental problems that could obscure findings in relation to prematurity (for examples, intellectual disability) *and* Studies where preterm sample was recruited *because* of comorbidities
Children born at any stage of prematurity (<37 weeks gestational age).	Preterm sample described only by birthweight without description of gestational age.
Direct experimentally measured assessments of social cognition (examples include eye tracking, play based measures, objective coding of parent/child play[4]).	Social cognition measured by informant report.
Healthy full term comparison group included.	Studies without a healthy full term control group (>37 weeks GA).
Study measures social cognition defined as social information processing, and so includes any study measuring the *mental processes* involved in social interactions. This includes imitation, social attention, responding to joint attention, initiating joint attention, affect recognition, theory of mind, biological motion perception, mentalizing, social reasoning and empathy. Definition from Taylor et al. (2022) was used to guide inclusion of studies falling outside those listed above. This definition is as follows:	Studies examining *social interaction*– this includes studies exploring how children interact with social partners, such as prosocial and antisocial, initiating and sustaining of interactions, co-operating, turn-taking, joint problems solving (as defined by Yeates et al., 2007; Taylor et al., 2022).

[4] Studies using objective coding of parent/child play excluded if *separate* coding of infant behaviour is not measured. Examples include coding behaviours of the dyad/coding which captures both behaviour of both social partners and not the solely child behaviour. This exclusion would therefore include joint attention episodes as coding of this is usually based on the amount of time a dyad is in a state of joint attention, irrespective of who initiated or responded to the bid for joint attention.

Inclusion criteria	Exclusion criteria
Social cognition is referred to as "social information processing [which] includes the mental processes involved in social interactions. An interaction begins when a child encounters a social cue, such as another child's verbalization, tone of voice, or facial expression. The child draws from a store of knowledge and past emotional experience in interpreting the meaning of the cue and in considering different options for responding to it. Following the interpretation of the cue and consideration of social goals and the consequences of alternative response options, the child decides how to respond and enacts the decision in a behavioural response" (Taylor et al., 2022).	
	Studies examining *social adjustment* – includes studies exploring peer acceptance, peer rejection/victimization and the extent and quality of peer relationships (as defined by Taylor et al., 2022; Yeates et al., 2007). Studies reporting diagnostic cut-offs for autism using the Autism Diagnostic Observation Scale (ADOS), *unless* aspects of social cognition were measured/coded independently.

During review, a third stage of screening was required which involved closer examination of studies in which the definition of social cognition was unclear. This involved review of studies reporting to measure joint attention and studies measuring aspects of social cognition outside of the categories established a priori to stage two screening. Responding to joint attention was defined as the ability to follow the direction of gaze and/or gestures of others (Mundy et al., 2007). Initiating joint attention was defined as the ability to use direction of gaze and/or gestures to direct the attention of others to spontaneously share experiences (Mundy et al., 2007). When information provided in a study was not sufficient to establish measurement in line with these definitions, studies were excluded. The rationale for this was that the aim of the current study was to attempt to group studies together which were measuring the same construct to allow for meta-analysis. Decisions for inclusion/exclusion were made collaboratively in supervision. Authors were contacted when studies did not report adequate statistical information. Studies were excluded from analysis when this information could not be accessed but retained for narrative review.

Domains of social cognition were not established a priori, and as such final categorisation of studies into sub-domains of social cognition were established based on how studies defined the sub-domain of social cognition they were measuring. The broad and varied terms used to describe the same, and different, aspects of social cognition meant this was important to meet the aims of this review and to be able to work towards assembling studies in a way which made quantification of possible differences between PT and FT children possible.

Data Extraction

Data from final studies meeting the inclusion criteria were extracted by a single reviewer (the author) and statistical information checked multiple times for accuracy. The final list of papers was allocated a number and a random number generator was used to select papers

at random. A second researcher extracted data from 20% of papers generated at random which were visually checked for accuracy. Only one discrepancy was detected, which was resolved by re-checking the original paper by both researchers. Each paper included in the meta-analysis was coded for study descriptors (such as sample size, gestational age, age of assessment) and methodology (such as assessment task, type of social cognitive skill measured).

All data were initially extracted as reported in the original paper (for example, all time points and all tasks used to measure social cognition). Every effort was made to ensure studies included were distinct samples from one another, verified through contacting study authors for confirmation where required. Specific decisions were made across studies once all data had been extracted. Where there was replication, or a study assessed participants longitudinally, the largest sample size was included. Where different time points contained the same sample size, the time point which most closely matched the age at assessment of other studies measuring the same social cognitive domain was chosen. Where a sample was used in both a published study and unpublished dissertation/thesis, the larger sample was chosen. Where both reported the same sample size, the published study was chosen for inclusion. Further decisions relating to studies using multiple measures of the same construct and those reporting outcomes for subgroups are detailed under the statistical analysis sub-heading.

Quality Assessment

The quality of all studies that progressed to the final stage of the review process were assessed using a bespoke quality assessment framework (Table 5) developed for the current review. This was developed to control for threats to validity which included aspects of quality considered relevant to the research question.

Factors of quality considered important fitted within six broad criteria. The sample (PT and FT sample identification and PT sample description) where random or total population sampling *and* a detailed description of gestational age, age at assessment, co-morbidities and

health status were considered excellent. Secondly, comparability of PT and FT groups, where group equivalence on both age and cognitive ability were considered excellent. The measure of social cognition, where the use of well-validated tasks/behavioural coding schemas were considered excellent. Blinding to group membership, where studies were considered excellent if both experimenters and coder were blinded. Inter-rater reliability, where excellent studies reported a moderate to excellent reliability. These benchmarks for inter-rater reliability were set based on literature (Cohen, 1960) for the purposes of establishing quality. Quality criteria relating to blinding to group membership and inter-rater reliability were considered relevant only to studies using behavioural measures, where such practices are considered gold standard. Studies employing eye tracking technology, in which behavioural coding is not required were rated as a three as studies employing these methods would not be subject to the same threats to validity.

To establish quality related to statistical bias, sample size was considered. A G*power analysis revealed that to detect an effect size of 0.2 (at $p < 0.05$ with 80% power) a sample of 620 (310 per group) would be required and therefore classified as excellent as per quality framework. To detect an effect size of 0.5 (at $p < 0.05$ with 80% power) a sample of 102 (51 per group) would be required. To detect and effect size of 0.8 (at $p < 0.05$ with 80% power) a sample of 42 (21 per group) would be required. The calculation of power was based on a one-tailed test which was considered appropriate as the difference between groups is reasonably expected in a specific direction.

To eliminate risk of bias, an additional researcher completed quality ratings against the quality weighting framework for a subset of included papers (20%). Studies were numbered and a random number generator was used to randomly select papers. An excellent level of reliability for the independent quality weightings was obtained (weighted Cohen's Kappa 0.87)

(Cohen. 1960). Where discrepancies were identified, agreement was made between the two raters.

Table 5

Table describing quality assessment framework and scoring.

		0 – Poor	1 – Fair	2 – Good	3 – Excellent
Sample	**Preterm Sample identification**	Unspecified	Single restricted or non-random sample (e.g., from a specialist clinic/single hospital site or previous research study)	Multiple restricted or non-random samples (multi-region specialist clinics/hospital sites)	Random or total population sample
	Preterm sample description - ***Co-morbidities/health***[5] - ***Adjusted/chronological age***	No description of preterm sample co-morbidities & health complications No statement of use of adjusted age or chronological age (< 2 years)[6]	Limited/broad description of the following: - Gestational age - Age at assessment - *Limited/broad* description of health status (for example, described only as *healthy*). *And* Clear description of use of adjusted age or	*Adequate* description of the following[7]: - Gestational age - Birth weight - Age at assessment - Co-morbidities & health complications (*some* physical & neurological comorbidities) *And* Clear description of use of adjusted age or chronological age (< 2 years)	*Detailed* description of the following: - Gestational age (mean. SD, range) - Birth weight (mean, SD, range) - Age at assessment (mean, SD, Range) - Co-morbidities & health complications (for example

[5] Studies reporting health status/comorbidities as part of inclusion/exclusion criteria to be checked and included within quality assessment (e.g., if a paper states that children with certain comorbidities were excluded etc, this should be taken into account when allocating a rating under this domain).
[6] A study should be scored as poor in the absence of a statement of use of adjusted/chronological age prior to the age of two (consistent with NICE guidance (NICE, 2017), even if co-morbidities and health status of sample are described.
[7] Adequate to include reporting of some statistical information such as mean, SD, range but not all need to be reported to score as good.

		0 – Poor	1 – Fair	2 – Good	3 – Excellent
			chronological age (< 2 years)		Apgar scores, physical, neurological comorbidities etc) *And* Clear description of use of adjusted age or chronological age (< 2 years)
	Full term sample identification	Unspecified	Single restricted or non-random sample (e.g. from a specialist clinic/single hospital site or previous research study/singular nursery)	Multiple restricted or non-random samples (multi-region specialist clinics/hospital sites/multiple nurseries)	Random or total population sample
Comparability of groups		Unspecified *OR* Groups not equivalent in age	Groups equivalent on adjusted age /chronological age ONLY[8]	Groups equivalent on adjusted age /chronological age ONLY *And* Attempted to control for cognitive ability however meaningful differences exist between groups despite this. *Or*	Groups equivalent on adjusted age/chronological age *And* Controlled for cognitive ability ensuring there were no meaningful differences between groups.

[8] Equivalent on gestational age before the age of two consistent with NICE guidance (NICE, 2017)

		0 – Poor	1 – Fair	2 – Good	3 – Excellent
				Controlling for cognitive ability was not completed in a robust way (e.g., the use of only one subtest)	
Measure of social cognition	**Task**	Novel task developed for the study *poorly* reported/insufficient information to allow for replication	Novel task behavioural coding scheme developed for study with clear description, sufficiently described for replication	Task *adapted* from well validated task/ well validated behavioural coding scheme, with clear description, sufficiently described for replication	Use of a well validated task/well validated behavioural coding scheme *And* Task used with the appropriate age for which the task was developed.
	Blinding	Unspecified	Attempted blinding of coders & experimenters to risk group but not possible due to child characteristics	Coders *only* blinded to group membership[9]	Both experimenters *and* coders both blind to group membership[10]
	Inter-rater reliability	Unspecified	Inter-rater reliability discussed but no statistical information or qualitative descriptions provided	Fair inter-rater reliability reported using either statistical information OR qualitative descriptions in	Moderate to excellent inter-rater reliability reported in line with general rule of thumb[12]

[9] Scored as two if study did not explicitly state experimenters were blind; scored as two if coder was blind to risk group/purpose of the study.
[10] Scored as three when eye tracking methodology was utilised. The rationale was that blinding to group was considered an unnecessary aspect of quality with eye tracking methods. Scoring as three allowed to not disadvantage eye tracking studies unfairly when comparing to other methods where blinding to group is an important aspect of quality related to the outcome of interest.
[12] Additionally scored as three when eye tracking methodology was utilised. The rationale was that inter-rater reliability was considered an unnecessary aspect of quality with eye tracking methods. Scoring as three allowed to not disadvantage eye tracking studies unfairly when comparing to other methods where inter-rater reliability is an important aspect of quality related to the outcome of interest.

		0 – Poor	1 – Fair	2 – Good	3 – Excellent
				line with general rule of thumb[11]	
Statistical Bias	**Size of sample[13]**	*n < 21*	*n=22-51*	*n=52-310*	*n=311 +*

[11] The general rule of thumb suggested by Cohen (1960) as follows: (0.21-0.40 – fair; 0.41-0.60 – moderate; 0.61-0.80 – good/substantial; 0.81-0.80 - excellent).

[13] Where PT and FT groups had different sample sizes, this was scored based on the smallest *N*. Where subgroups were combined for analysis, this was scored based on the combined *N*.

Statistical Analysis

Statistical analysis was performed in R, version 3.6.0 with RStudio, using the 'meta' and 'metafor' statistical software packages. Standard deviations and means for each social cognitive domain across PT and FT comparison groups were inputted into the spreadsheet. When this information was not available, effect sizes were calculated on the information made available in the paper, when sufficient information was provided to do so. When papers reported multiple measures of the same construct or reported outcomes for subgroups within PT and FT groups (for example gender, sub-categories of prematurity) these were combined to create a single pair-wise comparison in line with Cochrane guidance (Cochrane, 2022).

Separate meta-analyses were conducted to produce overall effect size estimates for early social preference, responding to joint attention (gaze following), responding to joint attention (gaze + point), initiating joint attention, affect recognition and theory of mind. Each pooled effect size was expressed as a standardised mean difference (*SMD,* Hedges' *g*) with corresponding 95% confidence intervals.

Across analyses, the Restricted Maximum Likelihood estimator (REML) was chosen as the appropriate method for the calculation of the variation of the difference between studies, as this estimator has been shown to be more robust to deviations from normality (Banks et al., 1985). QQ plots were then inspected for evidence of non-linearity in the distribution of *SMD* in both random and fixed effects model. If QQ plots suggested non-linearity and possible evidence of outliers, this was further explored prior to examining between study variation. Higgins et al. 2003 provide benchmarks for between study variation based on the value of Higgins I^2. When between study variation was low, defined by Higgins et al. 2003 as a Higgins $I^2 < 50\%$) *and* QQ plots indicated a normal distribution, the fixed effects model was chosen to describe the weighted average *SMD*. When between study variation was substantial (defined

as a Higgins I^2 > 50%), the random effects model was chosen to calculate the meta-analytic synthesis.

When substantial heterogeneity was evident (Higgins I^2 >50%), this was investigated further by inspecting Baujat scatterplots and the leave-one-out analysis for possible outliers impacting the magnitude of the weighted average effect. When the leave one out analysis suggested omitting a study would result in a significant reduction in heterogeneity (defined as I^2 >20%), studies exerting an influence were explored further via forest plot. The cut-off of I^2 > 20% here was a decision made by the research team as a change in variation of 20% by the removal of one study was interesting and therefore warranted further exploration. Individual study quality was also inspected to explore whether certain aspects of quality may account for discrepant studies. If substantial methodological variation was observed *or* the study was substantially different from the rest of included studies, studies were omitted from meta-analysis but included in narrative review.

The impact of study quality on effect size estimates was explored when Higgins I^2 >10% and hence indicated variability. The cut-off of I^2 > 10% here was a decision made by the research team as a change in variation of 10% due to aspects of study quality was interesting and therefore warranted further exploration. Individual This was achieved using a series of meta-regressions for the eight areas of study quality scored within the quality assessment framework. Given the predicted effects of PT children performing worse on social cognitive skills compared to FT children, negative estimates from meta-regression analysis would be consistent with better quality studies predicting greater differences between groups. Positive estimates from meta-regression analysis would be consistent with better quality studies predicting fewer differences between groups.

Despite the inclusion of unpublished studies in this meta-analysis, publication bias was still explored, particularly as the majority of included studies were from published literature.

Publication bias was identified by via inspection of funnel plots. Tests for funnel plot asymmetry should only be used when there are > 10 studies as when there are fewer than this, the power of the tests is too low to distinguish chance from real asymmetry (Cochrane, 2022). Therefore, visual inspection of funnel plots was supplemented with a statistical test of publication bias (Egger et al., 1997) only when there were >10 studies. When there was evidence of publication bias, the effect was simulated using a trim and fill procedure (Duval & Tweedle, 2000), and Rosenthal's failsafe number was used to determine the number of non-significant studies required to reduce this effect to non-significance.

Subgroup analysis and meta-regression were performed where possible to further explore the impact of study level covariates upon social cognitive outcomes. Meta-regression should generally not be considered when there are fewer than ten studies in a meta-analysis (Cochrane, 2022) and so this was performed based on whether the required number of studies was located and the availability of consistently measured covariates across studies. Due to inconsistencies across studies in definition of GA, GA was not explored as a covariate. As is described below, GA for PT children was categorised based on the reported mean to be able to offer a summary/indication of GA within each meta-analysis. This however was not an accurate reflection of GA, where studies often included a broad range of GAs and therefore analysing this statistically was unlikely to offer anything meaningful due to this inconsistency.

Structure of Results

The results are presented by way of approximate developmental age of which the skills typically emerge, in accordance with wider literature.

Due to variation in the reporting of participant characteristics across studies, demographics were grouped into categories to be able to describe and summarise PT and FT children within each meta-analyses in a meaningful way. The GA for children born prematurely were grouped into one of four categories using the reported means. These categories were <28

weeks (extremely PT), 28-<32 weeks (very PT), 32-<34 weeks (moderate PT) and 34-<37 weeks (late PT), which are widely report classifications of GAs of prematurity. For FT children, GA ranges were not reported for each meta-analysis due to inconsistences of reporting means and standard deviations for GA for this group across studies.

Age at assessment was reported for both PT and FT children within each meta-analysis as approximate age range only, again due to inconsistencies in methods of reporting. For example, some studies reported age in days instead of months or years, *and/or* means and standard deviations were not consistently reported across studies.

The measurement and reporting of cognitive ability between groups was varied and not possible to combine meaningfully. The number of studies reporting this information were summarised and tabulated to be able to describe characteristics of PT and FT samples in relation to differences in cognitive ability.

Results

As shown in the PRISMA flowchart in Figure 2, 19,038 records were returned from searches performed across all electronic databases listed. Subsequently 2,069 records were removed due to duplication. The remaining 16,969 records were reviewed in accordance with stage one criteria, and 701 were reviewed according to stage two criteria. The remaining studies *(n=56)* were identified as relevant to the current review. A visual matrix of quality weighting for *all* included studies is presented in Table 6, alongside study characteristics and outcome data.

Studies were organised into five overarching social cognitive domains as shown in table 7[14]. This table details how papers were categorised into these domains, reasons for exclusions and number of studies included in meta-analysis and review. Some studies measured multiple social cognitive domains and so are reported in multiple domains. Six separate meta-analyses were possible. Those papers not eligible for inclusion in analysis were retained for narrative review. Papers excluded from meta-analysis due to duplicate samples were not included in the narrative review as the sample was already represented within the reported meta-analyses. Results from individual meta-analyses are presented first, followed by narrative review of studies. A negative effect size demonstrated PT children performed worse than FT children across domains.

[14] Several terms have been referred to using acronyms where possible to avoid repetition of the same term throughout the text. To support the reader, they are as follows: Social cognitive domains have been referred to using the following acronyms: RJA = responding to joint attention, IJA = initiating joint attention and ToM = theory of mind, The groups of interest in this review have continued to be referred to as follows: PT = preterm and FT = full-term.

Figure 2

Figure showing Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flowchart.

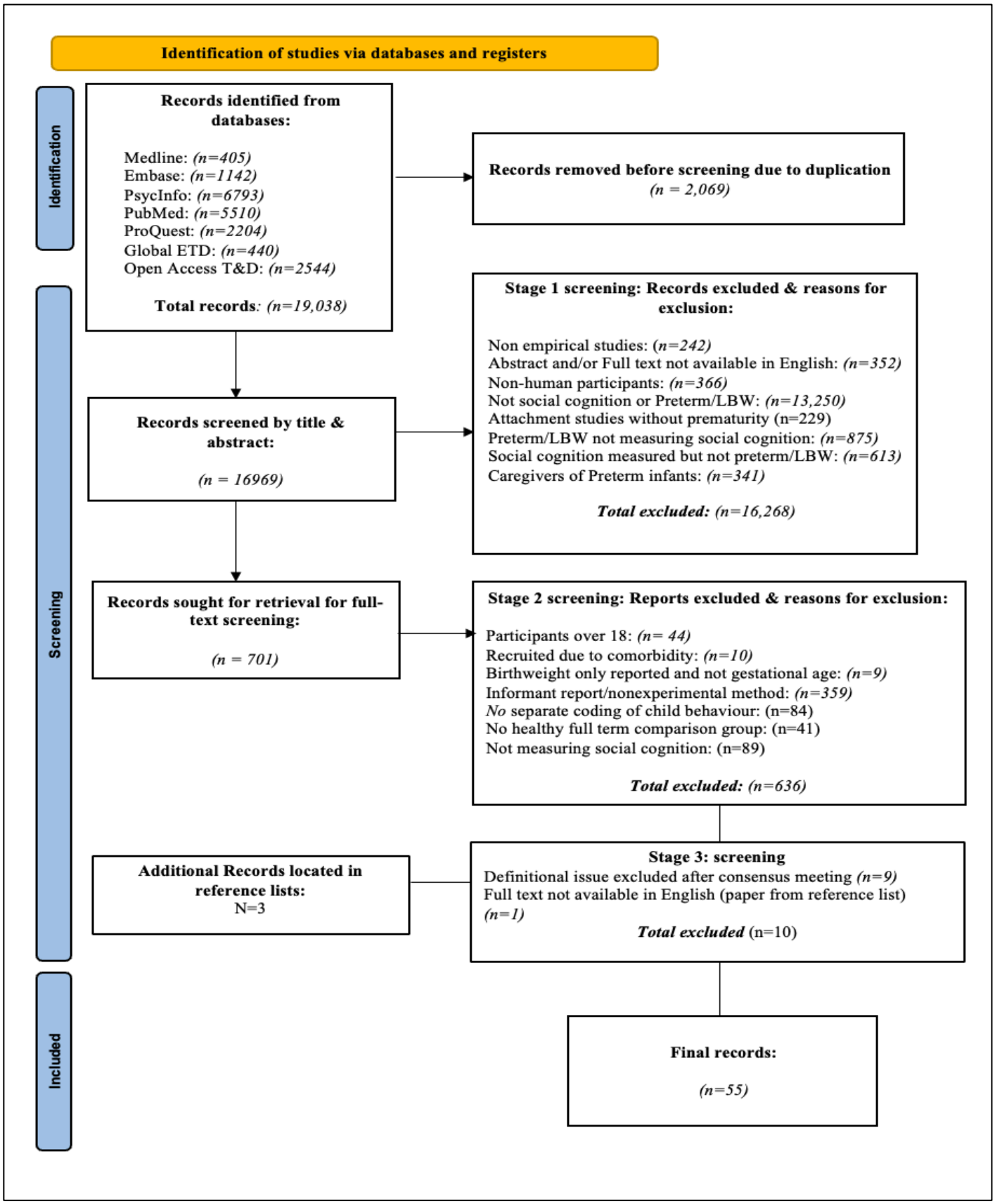


Table 6

Table describing individual study characteristics, key demographics of preterm and full-term children and individual study quality ratings.

Table organised by social cognitive domain.

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author ***Country***	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Social attention & preference																					
Early social attention																					
Dean et al., (2021)[18] *UK*	49.2	29 (NR) *23.2-33.0*	7.9 (NR) *6.8-8.7*	56.1	40 (NR) *37.0-42.1*	8.0 (NR) 7.4-9.1	3 social preferential looking tasks[19]	0.3	0.2	81	0.39	0.19	66	1	3	1	2	3	3	3	2
Imafuku et al., (2021)[20] *Japan*	34.2	31.9 (3.4) *24.4-36.9*	376.2 (14.4) *350-427*	30.8	39.6 (1.2) *37.1-41.9*	281.1 (10.8) *362-402*	Human/ geometric preference task	79.7	16.9	49	85.6	7.78	29	0	3	0	2	3	3	3	1

[15] 1=PT sample identification; 2=PT sample description; 3=FT sample identification; 4=Comparability of groups; 5=Measure of social cognition; 6=Blinding; 7=Inter-rater reliability; 8=Statistical bias.
[16] Gestational age reported in weeks. Mean: (SD); *range.*
[17] Age at assessment reported in its original format reported in the respective study. Age reported in months unless otherwise stated. Studies reporting chronological age only are marked with an *. Age at assessment is reported as mean: (SD); *range.*
[18] Time point 1 data extracted. Median and interquartile range reported for demographics in original paper and so reported as such here.
[19] Three social viewing tasks (Gillespie-Smith et l., 2016). Face task, pop-out task (Gilga et al., 2009); social preferential looking task (Fletcher-Watson et al., 2008; Fletcher-Watson et al., 2009; Telford et al., 2019).
[20] Time point 2 data extracted. Age reported by authors in days.

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author ***Country***	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Imafuku et al., (2017)[21] *Japan*	40	32.4 (2.9) *26-36*	376.4 (21.6) *334-110*	43.5	39.4 (1.1) *37-41*	379.3 (16.8) *336-410*	Biological motion task	52.6	12.2	18	51.2	15.9	23	0	2	0	1	3	3	3	0
							Human geometric task	80.4	15.8	16	83.4	12.5	21								
Face and eye preference																					
Sekigawa-Hosozawa et al., (2017)[22] *Japan*	45	28 (2) *24-32*	49 (25) *15-20*	37.84	*States full term only*	37 (23) *14-93)*	Dynamic social scenes depicting multiple social agents[23]	-	-	47	-	-	25	1	2	0	3	1	3	3	1
Yamamoto et al., (2021) *Japan*	50	30.4 (3.5) *23.4-36.9*	NR (NR) *9-10*	45	39.2 (1.6) *37.0-41.6*	NR (NR) *9-10*	Static stimuli of faces[24]	62.7		20	84.88		20	1	1	2	1	1	3	3	0
								79.8	-	20	74.49	-	20								
Pereira et al., (2017) *Brazil*[25]	50	35 (1.11) *33-36*	30* (11.9) *10--48*	6.5	21 (12.12) *2-47*	39 (1.13) *37-4*	Static stimuli with distorted & natural face[26]	1.8	1.1	28	3.2	0.8	31	1	1	1	0	1	3	0	1
								2.0	1.2	28	2.5	0.9	31								

[21] Time point 2 data extracted due to crossover in 6-month sample with Imafuku et al. (2021) sample. Both tasks reported here, however groups were combined for analysis as per Cochrane guidance. Age reported by authors in days.

[22] Statistical analysis not reported. Paper provides average proportion of face viewing for subgroups of preterm group. MDS does not provide for mean and SD in this type of analysis.

[23] Images for task measuring social cognition taken from previous work by the authors Nakano et al., (2010).

[24] Two facial stimuli presented simultaneously with familiarisation and novel stimulus (a new face) presented side by side. Study reports the median eye gaze rate (%) for familiarisation and test phases. The former is reported here first, and test phase statistics are reported second.

[25] Age reported by authors as hours of life.

[26] Methodology adapted by Goren and Sarty (1975) and the original study reports statistical information for natural face and distorted face. The former is reported here first and the latter second.

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author ***Country***	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Face recognition																					
Perez-Roche et al., (2017)[27] ***Spain***	41.7	31 (2.6) *NR*	8.4 (2.5) *NR*	51.2	39 (1.4) *NR*	8.9 (2.1) *NR*	TOMAL[28]	20	5.7	36	23	5.5	41	0	1	0	1	3	0	0	1
Frie et al., (2016) ***Sweden***	70.4	25.6 (NR) *23.6-27.7*	NR (NR) *6-10*	77.8	39.7 (NR) *35.7-42.3*	NR (NR) *6-10*	Task developed for study[29]	-	-	27	-	-	26	1	1	1	1	1	0	0	1
Imitation																					
Haley et al., (2019)[30] ***USA*** *Suppressors*	47.4	30.36 (2.31) 22.86-32.86	6 (NR) NR	36.8	39.85 (1.03) 37.29-41.71	6 (NR) NR	Delayed imitation puppet task[31]	-	-	38	-	-	38	1	1	1	1	1	2	3	2
Non-suppressors	57.8	28.77 (2.57) 23.71-32.86	6 (NR) NR	40.0	40.25 (1.19) 37.29-42.14	6 (NR) NR		-	-	45	-	-	30								

[27] Age reported by authors in years.
[28] The facial memory subtest of the test of memory and learning (TOMAL) developed by Reynolds (1994). Authors provide TOMAL recognition and percentage scores. Recognition scores only reported here.
[29] Methodology used in previous work by the authors (Carlsson et al., 2008). Four stimuli presented including a photo of mother's face and an unknown female. No statistical information provided by authors and therefore not reported here.
[30] Demographics reported in separate groups by the authors (suppressors and suppressors).
[31] Task developed by Barr et al. (1996). For statistical information please refer to original paper.

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author *Country*	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Linares (1986)[32] ***USA***	28.6	12 (NR) NR	32.0 (NR) NR	28.6	40.0 (NR) NR	12 (NR) NR	Imitation battery developed for study[33]	3.50	1.73	7	5.85	0.89	7	1	1	1	1	1	0	3	0
Biological motion perception																					
Taylor et al., (2009)[27] ***Canada***	47.8	29 (2) 26-31	7.3 (15.0) 5:2-9:1	50.0	NR (NR) 38-42	7.3 (12.7) 5:5-8:8	3 Biological perception tasks[34]	-	-	23	-	-	20	2	1	2	3	3	0	0	0
Williamson et al., (2015)[35] ***Canada***	43.8	28.3 (2.7) 25-33	8:10 (7.8) 8-9	46.7	NR (NR) 38-42	8:10 (6.9) 8-9	Biological motion perception test battery[36]	-	-	33	-	-	35	2	1	2	2	3	0	0	1
Joint attention behaviours																					
Yamoaka & Takada (2016) ***Japan***	51.6	28.4 (3.0) NR	19.1 (1.3) NR	60	39.1 (1.3) NR	19.1 (1.1) NR	JAOS[37]	-	-	31	-	-	45	1	2	1	1	1	0	3	1

[32] Time point 1 extracted.
[33] Task developed based on previous research by Mazur and Ritz, 1984; McCall et al. (1975).
[34] Global motion coherence task (Newsome & Pare, 1988); global from coherence task (Heywood et al., 1992); biological motion task (Grossman & Blake, 2001). Means and SD not reported in study.
[35] Time point 1 data extracted. Age reported by authors in years.
[36] Tasks taken from Saunders (2011). Means and SDs not reported by authors.
[37] Joint attention Observation Scale (JAOS) developed for the study devised to conform to criteria for the ESCS. Paper reports statistics individually for each item and so not reported here. please refer to original paper for this information.

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author ***Country***	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Responding to joint attention (RJA) - Gaze following																					
De Schuymer et al., (2011)[38] ***Belgium***	76	29.44 (1.80) 26-32	270.62 (4.66) NR	63	NR (NR) 38-42	273.62 (4.66) NR	Gaze following paradigm[39]	6.14	11.9	25	15.7	12.5	35	1	2	2	2	3	0	3	1
Imafuku et al., (2017)[40] ***Japan***	40	32.4 (2.9) *26-36*	367.4 (21.6) *334-110*	56.5	39.4 (1.1) *37-41*	375.2 (14.9) *340-408*	Gaze following paradigm (eye tracking)	0.56	0.27	20	0.67	0.26	23	0	2	0	1	2	3	3	0
Imafuku et al., (2021)[41] ***Japan***	34.2	31.9 (3.4) *24.4-36.9*	376.2 (14.4) *350-427*	30.8	39.6 (1.2) *37.1-41.9*	281.1 (10.8) *362-402*	Gaze following paradigm (eye tracking)	57.2	26.6	38	73.9	22.9	26	0	3	0	2	2	3	3	1
Landry (1986)[42] ***US*** *Low risk group*	24	30.3 (1.6) *NR*	6 (NR) *NR*	40	40.1 (2.8) *NR*	6 (NR) *NR*	Coding schema developed for study	6.17	3.65	25	5.75	2.79	25	1	1	2	0	1	0	3	1

[38] Age reported by authors in days.
[39] Task developed by Striano and Stahl (1999).
[40] 12-month time point extracted due to some cross over with 6-month time point in Imafuku et al., (2021) sample. Age reported by authors in days.
[41] 12-month time point extracted. Age reported by authors in days.
[42] Study included two groups (low risk and high risk) and report demographic and statistical information separately for groups. Subgroups were combined for analysis as per Cochrane guidance.

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author ***Country***	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
High risk group	47.6	30.2 (2) *NR*	6 (NR) *NR*	40	40.1 (2.8) *NR*	6 (NR) *NR*	Coding schema developed for study	6.52	3.14	25	5.75	2.79	25								
Pena et al., (2014)[43] a ***France***	63.2	29.2 (1.2) *29-31*	7* (NR) NR	50.0	29 (1.2) *38-42*	4 (NR) NR	Gaze following paradigm (head & gaze) eye tracking	0.615	0.17	19	0.50	0.13	20	0	2	0	1	2	3	3	0
Pena et al., (2014) b ***France***	52.6	NR (NR) 29-31	10* (NR) NR	43.5	NR (NR) 38-42	7 (NR) NR	Gaze following paradigm (head & gaze) eye tracking	0.655	0.15	19	0.61	0.13	23	0	2	0	1	2	3	3	0
Pena et al., (2014) c ***France***	50	NR (NR) 29-31	7* (NR) NR	42.9	NR (NR) 38-42	4 (NR) NR	Gaze following paradigm (head & gaze) eye tracking	0.63	0.19	18	0.49	0.16	21	0	2	0	1	2	3	3	0
Pena et al., (2014) d ***France***	45	NR (NR) 29-31	10* (NR) NR	50	NR (NR) 38-42	7 (NR) NR	Gaze following paradigm (gaze) - eye tracking	0.635	0.09	20	0.58	0.17	22	0	2	0	1	2	3	3	0
Rowell (2015)[44] ***USA*** *Gaze behaviour*	54	<32 weeks (NR) *NR*	20 (1.3) *18-23*	58	NR (NR) *37-42*	19 (1.7) *18-23*	Tasker's (2012) Coding Manual for Mother-Child Joint Attention	3.00	2.03	26	1.00	0.74	26	1	1	2	1	2	2	3	1

[43] This study reports 4 independent samples across 2 experiments.
[44] Rowell (2015) reported gaze behaviour separately and so was possible to be included in gaze following meta-analysis.

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author ***Country***	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Responding to joint attention (gaze + point behaviour combined) - RJA point+ gaze																					
De Schuymer et al., (2011)[45] ***Belgium***	76	29.44 (1.8) *26-33*	430.09 (10.9) *NR*	63	NR (NR) *38-42*	420.91 (3.93) NR	ESCS[46] *RJA side*	78.40	21.2	25	78.8	25.3	35	1	2	2	2	3	0	3	1
							ESCS[46] *RJA behind*	22.1	25.3	25	33.9	34.2	35	1	2	2	2	3	0	3	1
Fenoglio (2019)[47] ***USA***	47.4	29.31 (3.3) *NR*	12.78 (1.09) *NR*	47.7	39.78 (1.15) *NR*	11.98 (1.77) *NR*	DJAA[48]	2.10	0.93	57	2.28	1.03	149	1	2	0	1	2	0	0	2
PT clinic																					
PT registry	48.0	34.28 (1.18) *NR*	13.32 (1.26) *NR*	47.7	39.78 (1.15) *NR*	11.98 (1.77) *NR*	DJAA[48]	2.49	0.97	48	2.28	1.03	149	1	2	0	1	2	0	0	2
Gueron-Sela et al., (2015)[49] ***Israel***	NR	32 (1.7) *NR*	11.8 (0.6) *NR*	NR	39 (1.2) *NR*	11.7 (0.5) *NR*	ESCS[46]	-	-	83	-	-	67	1	1	1	2	3	2	3	2

[45] 14-month time point extracted. Authors report statistical information for multiple measures of the same construct (responding to joint attention -RJA *side* and responding to joint attention -RJA *behind*). These were combined for analysis as per Cochrane guidance. Age at assessment reported by authors in days.
[46] The Early Social Communication Scales developed by Mundy et al. (2003).
[47]Authors report statistical information for subgroups (PT clinic and PT registry) which were combined for analysis as per Cochrane guidance.
[48] The Dimensional Joint Attention Assessment developed by authors of the study.
[49] Statistical information taken from Mateus et al. (2019). The mean and SD were calculated from the hedges G and 95% confidence intervals of the effect size reported in this meta-analysis. Please refer to Mateus et al. (2019) for statistical information.

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author ***Country***	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Mateus et al., (2020)[50] ***Brazil***	76.5	32.16 (2.73) *26-36*	12.26 (0.23) *11.83-12.67)*	56.3	39.04 (1.34) *37-42*	12.15 (0.46) *11.06-12.87*	ESCS[46]	39.29	20.2	17	58.5	24.1	16	1	3	1	3	3	2	3	0
Olafsen et al., (2006)[51] ***Norway***	50.0	30 (3.5)	12* (NR)	54.3	39.3 (1.3)	12 (NR)	ESCS[46]	5.89	2.37	35	7.55	2.58	38	1	2	1	1	3	2	3	2
Boys		*NR*	*NR*		*NR*	*NR*															
Girls								7.03	2.30	31	9.56	2.23	32								
Rowell (2015)[52] ***USA***	54	<32 weeks (NR)	20 (1.3)	58	NR (NR)	19 (1.7)	Tasker's (2012) Coding Manual for Mother-Child Joint Attention	3.00	2.03	26	1.00	0.74	26	1	1	2	1	2	2	3	1
Gaze behaviour		*NR*	*18-23*		*37-42*	*18-23*															
Point behaviour								0.50	0.76	26	9.00	4.23	26								
Speretto (2015)[53] ***UK***	NR	34.27 (1.8) *30.0-36.6*	13* (NR) *NR*	NR	39.95 (1.45) 37.0-42.2	19 (NR) *NR*	ESCS[46]	1.03	1.07	35	3	1.89	53	2	2	2	3	2	2	3	1

[50] 12-month time point extracted.
[51] Authors report statistical information for subgroups (boys and girls) which were combined for analysis as per Cochrane guidance. Demographic information reported for entire preterm and full-term samples in paper.
[52] Authors report responding to joint attention behaviours separately for gaze and point. Other papers included within the meta-analysis reported these combined behaviours. Therefore, were combined for analysis as per Cochrane guidance.
[53] 13-month time point extracted.

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author ***Country***	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Initiating joint attention (IJA)																					
De Groote et al., (2006) ***Belgium***	40	30.8 (2.8) *NR*	26.4 (1.3) *NR*	22.9	NR (NR) *39-41*	26.4 (2.2) *NR*	ADOS[54]	2.24	2.09	23	1.11	0.99	19	1	2	0	3	2	0	3	0
De Schuymer at el., (2011)[55] ***Belgium***	76	29.44 (1.8) *26-33*	430.09 (10.9) *NR*	63	NR (NR) *38-42*	420.91 (3.93) NR	ESCS[46]	1.02	1.07	35	3	1.89	53	2	2	2	3	2	2	3	1
Gueron-Sela et al., (2015)[56] ***Israel***	NR	32 (1.7) *NR*	11.8 (0.6) *NR*	NR	39 (1.2) *NR*	11.7 (0.5) *NR*	ESCS[46]	-	-	83	-	-	67	1	1	1	2	3	2	3	2
Mateus et al., (2020) ***Brazil***	76.5	32.16 (2.73) *26-36*	12.26 (0.23) *11.83-12.67*	56.3	39.04 (1.34) *37-42*	12.15 (0.46) *11.06-12.87*	ESCS[46]	27.12	13.5	17	25.4	11.3	16	1	3	1	3	3	2	3	0
Olafsen et al., (2006)[57] ***Norway***	50.72	30 (3.5)	12* (NR)	54	39.3 (1.3)	12 (NR)	ESCS[46]	15.89	7.74	35	23.74	15.8	38	1	0	1	0	3	2	3	1

[54] Coding schema developed for study for responding to joint attention behaviours.
[55] Age reported by authors in days.
[56] Statistical information taken from Mateus et al. (2019). The mean and SD were calculated from the hedges G and 95% confidence intervals of the effect size reported in this meta-analysis. Please refer to Mateus et al. (2019) for statistical information.
[57] Authors report statistical information for subgroups (boys and girls) which were combined for analysis as per Cochrane guidance. Demographic information reported for entire preterm and full-term samples in by authors.

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author ***Country***	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Boys		*NR*	*NR*		*NR*	*NR*															
Girls								22.23	12.7	31	24.75	10.2	32								
Rowell (2015)[58]	54	<32	20	58	NR	19		5.00	2.76	26	1.00	1.12	26								
USA		weeks	(1.3)		(NR)	(1.7)	Tasker's (2012)														
Gaze behaviour		(NR)	*18-23*		*37-42*	*18-23*	Coding Manual for Mother-Child							1	1	2	1	2	2	3	1
Point behaviour		*NR*					Joint Attention	1.00	1.12	26	1.50	1.83	26								
Speretto (2015)[59]	NR	34.27	13*	NR	39.95	19	ESCS[46]														
UK		(1.8)	(NR)		(1.45)	(NR)		10.65	6.04	40	9.46	6.58	65	2	2	2	3	2	2	3	1
		30.0-36.6	*NR*		37.0-42.2	*NR*															
Theory of mind (ToM)																					
Ford et al., (2017)[60]	48.5	27	*n=10	51.4	Born	N=9 age	False Belief														
Australia		(2.18)	(7);		after 37	7; n=16	Task[61];	2.86	1.03		2.89	0.81									
		24-32	*n*=19		weeks	age 8;				35			37	1	2	2	2	3	0	0	1
			(8);		GA	n=12	Faux Pas	10.52	2.57		11.76	2.03									
						age 9	Task[62]														

[58] Authors report responding to joint attention behaviours separately for gaze and point. Other papers included within the meta-analysis reported these combined behaviours. Therefore, were combined for analysis as per Cochrane guidance.
[59] 13-month time point extracted.
[60] Age reported by authors in years. Authors report age at assessment information as number of participants within three different age groups.
[61] (Stories 1 and 5 from Liddle & Nettle, 2006) and two higher-order false-belief tests (the ice- cream-van story from Perner & Wimmer, 1985) partly devised by authors for the study. Participants completed multiple measures of TOM tasks which were combined for analysis as per Cochrane guidance.
[62] child version of the Faux Pas Test (Baron-Cohen et al., 1999).

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author *Country*	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
			n=6 (9)																		
Fu et al., (2022)[63] ***USA***	50.0	25.96 (1.3)	*14.06 (1.8) *NR*	48	39.08 (1.12) *NR*	14 (1.61) *NR*	EEFT[64]	27.77	6.25	22	33.24	6.11	25	0	2	0	1	3	0	0	1
Jones et al., (2013)[27] ***New Zealand***	49.5	27.84 (2.35) *NR*	4 NR *NR*	54.3	39.47 (1.18) *NR*	4 *NR* *NR*	Sally Anne task[65]; Unexpected Contents task[65]; Fishing story task[66]	0.76	0.84	103	0.92	0.89	105	1	2	1	2	3	0	0	2
Marleau et al., (2020) ***Canada***	43.3	25 (3) *22.1-28.7*	*66.3 (2.3) *NR*	50	39 (1) *NR*	60.6 (4.5) *NR*	Wellman & Liu Battery[67]	1.87	1.50	30	5.47	1.11	30	1	2	2	2	3	0	0	1

[63] Age reported by authors in years. Authors provide a composite score for all three which was extracted for analysis and reported here.
[64] The Emotional & Emotive Faces Task (EEFT; Dennis et al., 1998; 2013a). The 'look on face' trials are described by authors as the measure of affective TOM as it taps into the capabilities of understanding the actual emotion of the protagonist feels might be different form the emotion that he/she displays (Dennis et al., 2013b; Robinson et al., 2014; Ryan et al., 2021). Such scores are reported to correlated with scores on other aspects of theory of mind (Fu et al., 2022; Ryan et al., 2021) and as such this measure was extracted here as a measure of TOM (Fu et al., 2022).
[65] Task developed by Premack and Woodruff (1978).
[66] Task developed by Woolfe et al. (2002). Authors provide a composite score for all three tasks which was extracted for analysis and reported here.
[67] Wellman and Lui (2004) assessment battery.

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author ***Country***	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Mossad et al., (2017)[27] ***Canada***	50	28.4 (2.1) *NR*	*10.25 (1.7) *NR*	50	>40 *NR* *NR*	10.3 (1.8) *NR*	TOM subtest NEPSY-II[68]	22.80	2.70	24	24.80	2.60	24	2	1	2	2	3	0	0	1
Mossad et al., (2021)[27] ***Canada***	60	28.9 (1.9) *NR*	*8.7 (0.51) *NR*	51.2	39 (1.35) *NR*	8.6 (0.46) *NR*	Social Attribution Task[69]	82.6	19.4	40	90.4	15.6	42	2	1	2	2	3	0	0	1
Roldan-Tapia et al., (2017) ***Spain***	31.4	29 (2) *NR*	*78 (15.8) *NR*	68.6	39 (0) *NR*	77.9 (15.9) *NR*	Sally Anne Task[65] Unexpected contents task[65]	0.51 0.48	0.51 0.50	35	0.74 0.77	0.44 0.43	35	1	2	2	1	3	0	0	1
Williamson & Jakobson (2014a)[27] ***Canada***	47.0	28.3 (2.3) 25-33	*10.0 (15.0) *8.0-11.9*	54.3	*NR* (NR) 38-42	10.2 (14.7) *8.1-11.11*	Social Attribution Task[70]	2.74	0.78	34	3.25	0.77	36	1	3	2	3	3	0	0	1

[68] The Comprehensive Neuropsychological Battery - NEPSY-11 (Korkman et al., 2007). This study also used the Jack and Gill Tom task developed by Dennis et al. (2012). The author was only able to provide mean accuracy scores for this task due to a corrupt data file. Therefore, only the NEPSY-II theory of mind task data was extracted for analysis and reported here.

[69] Task designed by Klin (2000) and Schultz et al. (2003). Mean and SD not reported in original paper but kindly provided by the author.

[70] Animated triangles task developed by Abell et al. (2000) and Castelli et al. (2000). The mean and SD of the intentionality scores for the TOM displays in this task were not reported in the original paper but kindly provided by the author.

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author ***Country***	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Witt et al., (2014) ***Switzerland***	43.9	26.7 (1.13) *NR*	43;22 (2:18) *NR*	42.6	>37 *(NR)* *NR*	42:26 (2:18) *NR*	TOM subtest NEPSY-II[71]	44.31	19.9	41	51.09	22.0	47	1	2	1	2	3	0	0	1
Witt et al., (2018) [27] ***Germany***	NR	28.3 (1.92) *NR*	3:2 8: 51) *NR*	*NR*	39.6 (1.28) *NR*	3: 2 (2:12) *NR*	Extended Theory of Mind Scale[72]	1.23	0.65	26	1.58	0.76	26	1	2	0	3	2	0	0	1
Wolfe et al., (2015)[27] ***USA***	55.0	28.05 (2.7) 22-32	*5.66 (0.7) *NR*	55.6	38.89 (1.1) 37-41	6.26 (0.6) *NR*	Wellman & Liu assessment battery[67]	3.30	0.90	20	3.94	1.10	18	2	2	2	2	3	0	0	0
Ye et al., (2015)[27] ***Canada***	52.9	28 (2) *NR*	*10.2 (2) *NR*	52.5	>37 *NR* *NR*	10.2 (1.9) *NR*	TOM subtest NEPSY-II[71]	22.50	2.60	17	24.20	3.10	12	0	2	0	3	3	0	0	0
Affect recognition																					
Longa et al., (2022)	55.9	30 (2.65) *NR*	*104.0 (15.4) *NR*	42.9	<37 *NR* *NR*	105.10 (15.94) *NR*	Emotion Recognition Task (ERT)[73]	69.80	7.41	34	73.55	5.10	21	2	2	0	3	2	0	0	0

[71] The Comprehensive Neuropsychological Battery - NEPSY-11 (Korkman et al., 2007).
[72] This task was developed by Peterson et al. (2012) and the German version of this task was used in this study which was developed by Henning et al. (2012).
[73] The task here was built on Ekman and Friesen's (1976) "pictures of facial affect" (Ekman, 1976), adapted, and validated by (Kanel et al., 2021) with stimuli of children's faces taken from the validated Dartmouth database of children's faces (Dalrymple et al., 2013) as reported in Longa et al. (2022).

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author *Country*	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Italy																					
Marleau et al., (2020) ***Canada***	43.3	25 (3) *22.1-28.7*	*66.3 (2.3) *NR*	50	39 (1) *NR*	60.6 (4.5) *NR*	Affect Recognition subtest of NEPSY-II[71]	8.97	2.34	30	11.3	1.42	30	1	2	2	2	3	0	0	1
Marotta et al., (2020)[27] ***Italy***	39.6	35.34 (0.83) *34-36*	*8.30 (1.62) *5-11*	39.6	39.11 (0.47) *39-42*	8.34 (1.56) *5-11*	Emotion Recognition test[74]	9.42	2.7	53	10.63	3.08	53	2	2	3	3	3	0	0	2
Mossad et al., (2021)[75] ***Canada***	62.5	29 (1.6) *NR*	*6.6 (0.3) *NR*	42.1	39 (1.6) *NR*	6.5 (0.5) *NR*	Affect Recognition subtest of NEPSY-II[71]	9.4	2.9	40	10.8	3.9	38	1	1	0	2	3	0	0	1
Sato et al., (2021)[27] ***Canada***	53.6	28 (2) *NR*	*5.8 (0.2) *NR*	50.0	40 (1.3) *NR*	5.6 (0.4) *NR*	Affect Recognition subtest of NEPSY-II[71]	103.9	13.6	28	111.9	11.8	24	3	2	2	2	3	0	0	1
Twilhaar et al., (2019)[76] ***Netherlands***	50.8	29 (1.57) *NR*	*13.35 (0.36) *NR*	44.0	>37 *NR* *NR*	13.27 (0.53) *NR*	Morphed Facial Emotion Recognition task[77]	70.93	0.84 (SE)	61	73.83	0.77 (SE	61	1	2	1	2	2	0	0	2

[74] The emotion recognition subtest of the (IDS-SEC) developed by Grob et al. (2009) and Meyer et al. (2009).
[75] Age reported by authors in years; Time point 1 data extracted.
[76] Age reported by authors in years. For task statistical information provided in this paper, authors report SE which is reported here. This was converted to SD prior to analysis.
[77] A modified version of the Morphed Facial Emotion Recognition Task developed by Bergwerff et al. (2016).

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author ***Country***	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Williamson & Jakobson (2014)b[78] ***Canada***	47.1	28 (2.3) 25-33	*10.0 (15.0) *8.0-11.9*	52.8	2 weeks of due date (NR) *NR*	10.2 (14.7) 8.1-*11.11*	The Child & Adolescent Social Perception measure (CASP)[79]	30.2	1.5 (SEM)	34	36.8	1.5 (SEM)	36	2	3	2	3	3	0	0	1
Witt et al., (2014) ***Switzerland***	43.9	27 (1.13) *NR*	*43:22 (2:18) *NR*	42.6	>37 (NR) *NR*	Age matched (NR) *NR*	Affect Recognition subtest of NEPSY-II[71]	37.5	25.4	41	48.9	15.3	47	1	2	1	2	3	0	0	1
Wolfe et al., (2015)[27] ***USA***	55.0	28 (2.7) *22-32*	5.66 (0.7) *NR*	55.6	39 (1.1) *37-41*	6.26 (0.6) *NR*	The Emotion Understanding Assessment (EUA)[80]	17.25	2.4	20	18.44	1.8	18	2	2	2	2	3	0	0	0
Ye et al., (2015)[27] Canada	52.9	28 (2) *NR*	10.2 (2) *NR*	52.9	>37 (NR) *NR*	10.2 (1.9) *NR*	Affect Recognition subtest of NEPSY-II[71]	11.4	1.3	17	11.5	2.1	17	0	2	0	3	3	0	0	0
Social reasoning																					

[78] Age reported by authors in years. For task statistical information provided in this paper, authors report SEM which is reported here. This was covered to SD prior to analysis.
[79] Task developed by Magill-Evans et al. (1995). The emotion score of this measure provides a measure of a child's ability to identify and label the emotions displayed by each character as reported by Williamson and Jakobson (2014) b. This emotion score was extracted for analysis.
[80] Task developed by Howlin et al. (1999) and Downs et al. (2007).

	Preterm (PT)			Full-Term (FT)				Statistical Information						Quality Assessment[15]							
Author *Country*	**% male**	**GA.[16]**	**Age At Ass.[17]**	**% male**	**GA**	**Age At Ass.**	**Measure**	**PT mean**	**PT SD**	**PT N**	**FT mean**	**FT SD**	**FT N**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Lejeune et al., (2016)[27] ***Switzerland***	41	28.4 (2.7) *NR*	6.06 (0.38) *NR*	38	>38 (NR) *NR*	6.06 (0.43) *NR*	Social Resolution Task (SRT)[81]	27	12	102	32	14	88	2	2	2	3	1	0	0	2
Marotta et al., (2020)[27] ***Italy***	38.6	35.34 (0.83) 34-36	8.30 (1.62) *5-11*	39.6	39.11 (0.47) *39-42*	8.34 (1.56) *5-11*	Understanding of social situations test[82]	10.74	3.36	53	10.81	2.88	53	2	2	3	3	3	0	0	2
Williamson & Jakobson (2014) b[83] ***Canada***	47.1	28 (2.3) 25-33	*10.0 (15.0) *8.0-11.9*	52.8	2 weeks of due date (NR) *NR*	10.2 (14.7) 8.1-*11.11*	The Child & Adolescent Social Perception measure (CASP)	-	-		-	-		2	3	2	3	3	0	0	1
Campbell et al., (2015)[84] ***Australia***	NR	26 (NR) 24-28	6.83 (0.16) NR	NR	NR (NR) *NR*	7.09 (0.42) NR	KEDS[85]	58.13	8.75	32	79.26	11.1	38	1	1	1	2	3	0	0	1

[81] Social resolution task developed by Barisnikov et al. (2005). For task statistical information, the social reasoning score is reported here. Global social reasoning score and individual subscale scores for judgement and identification are also reported in the original paper.
[82] The understanding of social situations test is a task taken from the IDS-SEC developed by Grob et al. (2009) and Meyer et al. (2009).
[83] Age reported by authors in years. Means and SD not reported for this measure in the paper. The non-verbal cue score of this measure provides a measure of a child ability to identify specific face, body, situational and voice cues that are typically used to determine a characters emotional state as reported by Williamson and Jakobson (2014)b. This verbal cue score measure was retained for narrative review related the social reasoning domain of social cognition.
[84] Age reported by authors in years. Median instead if the mean is reported by authors as measure of central tendency in demographic information only.
[85] The Kids Empathy Development Scale (KEDS) developed by Reid et al. (2013).

Table 7

Table describing studies grouped by social cognitive domain and number of studies progressing to meta-analysis and narrative review.

Domain of social cognition	Studies (Total K)	Studies removed & reason for removal[86]	Meta-analysis studies (K)	Narrative review studies (K)
1. Social attention/perception	15	-	-	-
Social preference	6	*K=3* duplicate (Telford et al., 2016, Telford, 2017; Davis et al., 2022)	3	-
Face recognition	2	-	-	2
Imitation	3	*K=1* No statistical information reported (Speretto, 2015)	-	2
Biological motion perception	3	*K=1* Duplicate (Williamson, 2014)	-	2
Miscellaneous	3	-	-	3
2.Joint attention	15	*K=2* Duplicate (Olafsen et al., 2012; Egotubov et al., 2020) *K=1* Retained for review (Yamaoka & Takada, 2016) due to differences in measurement	-	1
Gaze following	6	-	6[87]	-
RJA (gaze + point)	7	-	7	-
RJA (point)	1	-	-	1
IJA (gaze + point/show)	7	-	7	-
IJA (gaze)	2	-	-	2
IJA (point + show)	2	-	-	2
3.Affect recognition	14	*K=1* Duplicate (Mossad, 2015) *K=1* No statistical information reported *K=2* Retained for review (Mossad et al., 2020; Wocaldo & Rieger, 2006)	10	2
4.Theory of mind (ToM)	15	*K=3* Duplicate (Hood, 2009; Mossad, 105; Williamson, 2014) *K=1* Outlier removed from analysis	11	-
5.Social reasoning & empathy	5	-	-	5

[86] Duplicate refers to sample was used in another study included in meta-analysis/narrative review.
[87] Pena et al. (2014) included 4 independent samples as so were entered into meta-analysis as separate studies.

Meta-Analysis Results

For each meta-analyses, QQ plots for both fixed effects and random effects models are presented in Appendix 2. Important descriptors of included studies within each meta-analyses are summarised in Table 8. For *all* meta-analyses further exploration of the impact of study level covariates upon differences between PT and FT children across the different social cognitive domains was not possible due to insufficient data for subgroup analysis. The data reported across studies for variables such as age at assessment, gestational age, birthweight, and socio-economic status were reported inconsistently. For example, many studies did not report important information such as mean, standard deviation and range of these variables. This made the analysis of important participant characteristics impossible to explore statistically.

Table 8

Table showing participant numbers and relevant demographics within each meta-analysis.

		Total (n)		Age[88]		GA[89]		
Domain of social cognition	**Number of studies**	**Preterm**	**Full-term**	**Preterm**	**Full-term**	**Preterm**	**Full-term**	**Cognitive Ability** *(Cog. Ability)*
Social preference	3	148	118	7-12 (months)	8-9 (months)	*K*=3 (28-<32)	37-42.1 range	*K*=3 no measure of cog. Ability
Gaze following	9	235	246	6-12 (months)	4-12 (months)	*K*=9 (28-<32)	37-42 range	*K*=8 no measure of cog. Ability *K*=1 significant differences between groups
RJA (gaze + point)	7	357	416	12-20 (months)	11-19 (months)	*K*=5 (28-<32) *K*=2 (32-<34)	37-42 range	*K*=3 no measure of cog. Ability *K*=2 significant differences between groups *K*=2 no differences between groups
IJA	7	207	249	12-26.4 (months)	12-26.4 (months)	*K*=6 (28-<32) *K*=1 23-<34	Range 37-42.2	*K*=3 no measure of cog. Ability *K*=2 significant differences between groups *K*=2 no differences between groups
Affect recognition	10	358	345	4-13 (years)	4-13 (years)	*K*=2 (<28) *K*=7 (28-<32) *K*=1 (34-<37)	37-40 range	*K*=1 no measure of cog. Ability *K*=2 no differences between groups *K*=5 FT outperformed PT but PT within normal range *K*=2 significant differences between groups
TOM	11	397	407	4-14 (years)	4-14 (years)	*K*=4 (<28) *K*=7 (28-<32)	37-41 range	*K*=2 no measure of cog. Ability *K*=4 no differences between groups *K*=4 FT outperformed PT but PT within normal range *K*=1 significant differences

Note *– Preterm GA categorised based on mean GA reported in studies to be able to offer approximate summary statistics. It is important to note however that range of GA varied between studies and although categorised as such here, preterm infants outside these GA's were included within each study.*

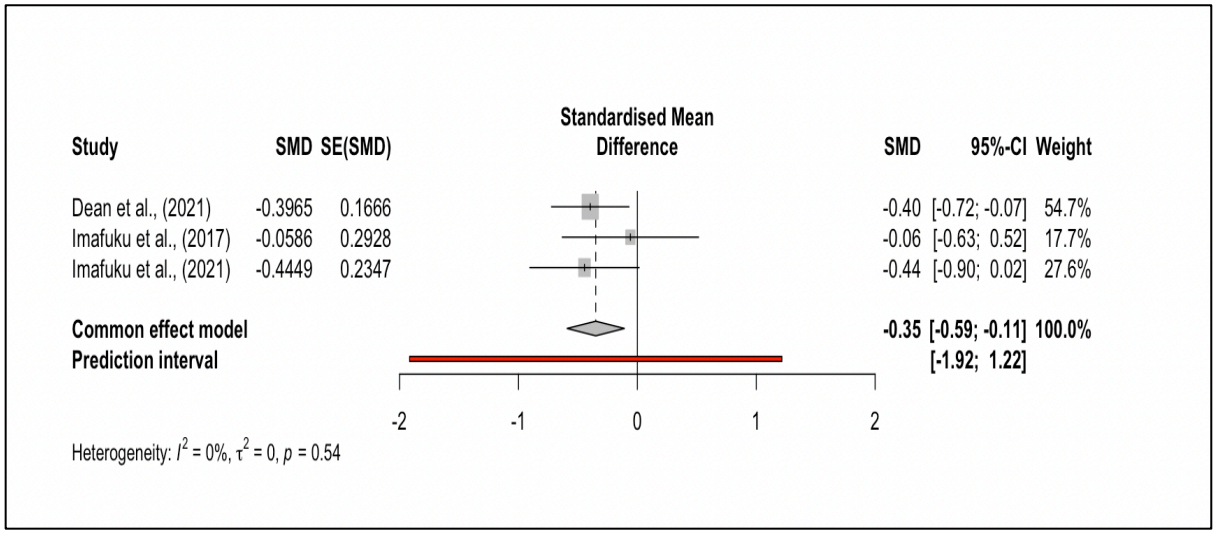
[88] Age at assessment reported as range only due to differences in reporting across studies.
[89] Gestational age.

Social Preference

A total of three studies measuring social preference were included in the meta-analysis. This included 148 PT children and 118 FT children. The fixed effects model was used due to effects being consistent within a normal distribution on the QQ plot (see Appendix 2) and a low level of heterogeneity (Higgins I^2=0%, tau^2= 0.0, p= 0.54). The model suggested a weighted average *SMD* = -0.35 z= -2.84, p = <.01) and a 95% confidence interval of between -0.69 and -0.11. See Figure 3 for forest plot. Leave-one-out analysis and exploration impact of study quality were not conducted (as heterogeneity was <50% and <10% respectively). The funnel plot (see Appendix 3A) was examined for evidence of publication and small study bias. Visual inspection did not show any evidence of bias; however, conclusions are limited due to the small number of studies included. Egger's test (Egger et al., 1997) was not appropriate due to <10 studies.

Figure 3

Figure showing forest plot for social preference studies.



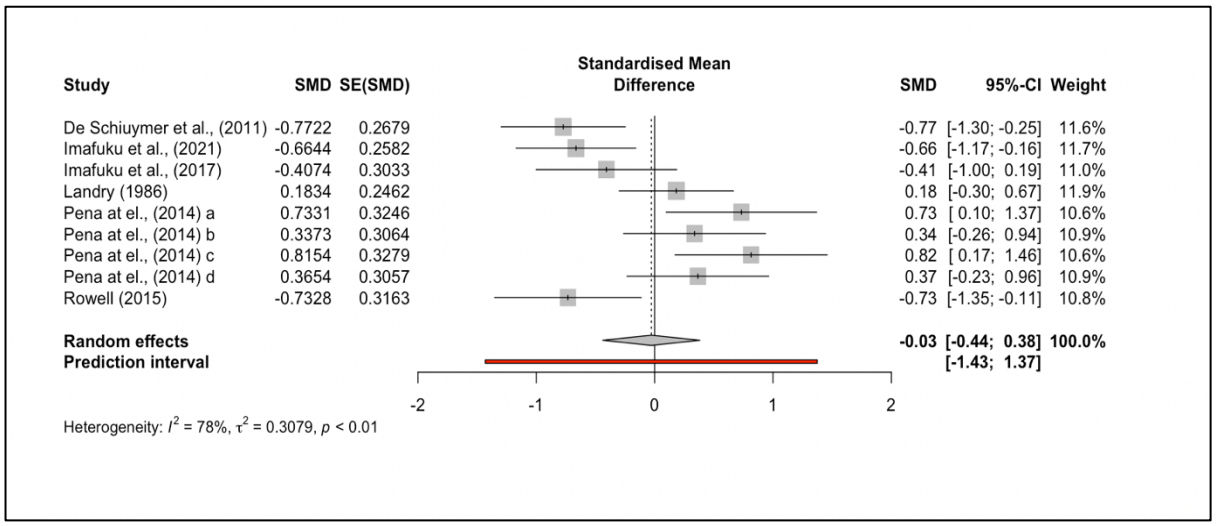
Note *– A negative value indicates preterm children demonstrated reduced social preference compared to full-term children.*

Responding to Joint Attention - Gaze Following[90]

A total of nine studies measuring gaze following were included in the meta-analysis. This included 235 PT children and 246 FT children. The random effects model was chosen as substantial heterogeneity was indicated (Higgins I^2=77.80%, tau^2=0.30, p= <.01). The model suggested a weighted average *SMD* = -0.01, z= 0.13, p = 0.89 and a 95% confidence interval of between -0.44 and 0.38 See Figure 4 for forest plot.

Figure 4

Figure showing forest plot for responding to joint attention (gaze following) studies.



Note *– A negative value indicates preterm children demonstrated reduced gaze following abilities compared to full-term children.*

Baujat scatterplot and the leave-one-out procedure were inspected (see Appendix 4A and Appendix 5A respectively). Imafuku et al. (2021) and Rowell (2015) appeared at the far-right corner of the scatterplot indicating these studies *may* be exerting a substantial influence on the overall analysis and might be discrepant from the rest of the literature. The leave-one-out analysis did not result in a significant reduction in overall heterogeneity when omitted (Higgins I^2=76.60%. when omitting Imafuku et al., 2021; Higgins I^2=77.60% when omitting

[90]Throughout the results section only, responding to joint attention (gaze following) will be referred to as gaze following.

Rowell, 2015). These studies were examined on the forest plot presented in Figure 4 and did not appear to be notably inconsistent with the rest of the literature.

As Higgins I^2>10% a series of meta regression analysis were conducted for each of the eight areas of study quality shown in Table 9. Statistical bias reached statistical significance suggesting this quality domain was associated with changes in the effect size estimate of differences between PT and FT children's gaze following behaviours. Higher quality studies in these domains predicted lower effect size estimates.

Table 9

Table showing the impact of study level risk of bias for studies measuring gaze following.

					Confidence interval	
Quality domain	**Estimate**	**SE**	**Z**	**P**	**Lower**	**Upper**
PT sample identification	-0.43	0.50	-0.86	0.39	-1.41	0.55
PT sample description	-0.04	0.38	-0.10	0.92	-0.78	0.69
FT sample identification	-0.21	0.25	-0.86	0.39	-0.71	0.28
Comparability of groups	-0.42	0.41	-1.03	0.30	-1.22	0.38
Measure of social cognition	-0.13	0.67	-0.19	0.85	-1.45	1.19
Blinding	0.05	0.23	0.22	0.83	-0.39	0.49
IR Reliability	0.07	0.21	0.32	0.75	-0.34	0.48
Statistical Bias	**-0.75**	0.37	-2.04	**0.04**	-1.46	-0.03

Note – *bold indicates statistical significance.*

The funnel plot of included studies is presented in Appendix 3B. Visual inspection suggested approximate symmetry. Further inspection demonstrated approximately 50% of studies were in the expected direction (to the left of the funnel plot and 50% in the unexpected direction (to the right of the funnel plot), suggesting *improved* gaze-following skills in PT children. Egger's test (Egger et al., 1997) was not appropriate due to <10 studies.

Responding to Joint Attention (Gaze + Point)[91] Results

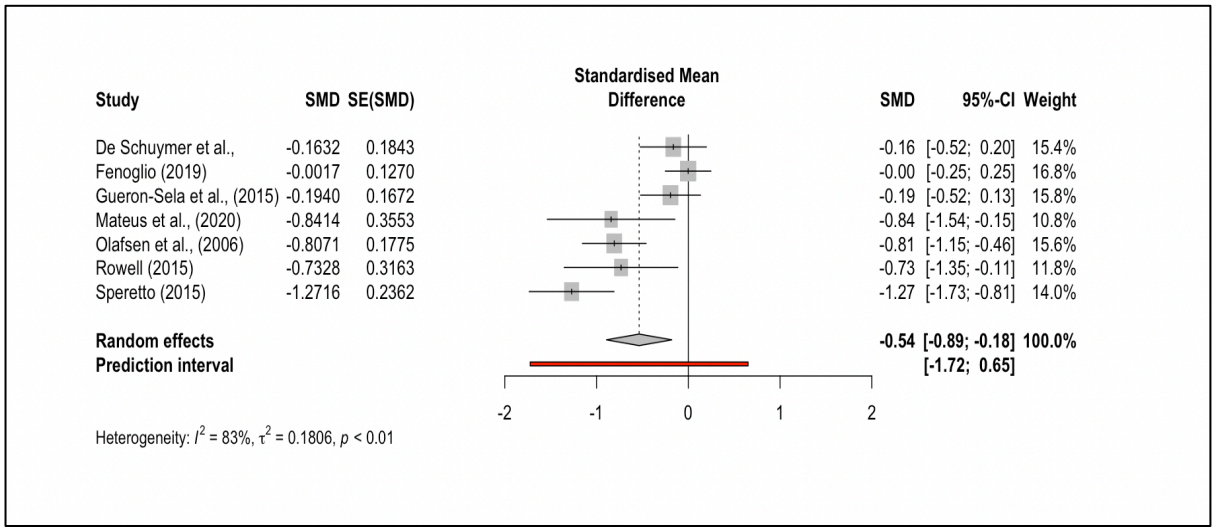
A total of seven studies measuring RJA (gaze + point) were included in the meta-analysis. This included 357 PT children and 416 FT children. The random effects model was chosen as substantial heterogeneity was indicated (Higgins I^2=82.60%, tau^2=0.42, p=<.01). The

[91]Responding to joint attention (point + gaze) will be referred to using the acronym RJA (point + gaze) from here onwards.

model suggested a weighted average *SMD* = -0.05, *z*= -2.95, *p* = <.01) and a 95% confidence interval of between -0.89 and -0.18**.** See Figure 5 for forest plot. Baujat scatterplot and the leave-one-out procedure were inspected (see Appendix 4B and 5B respectively), whereby Speretto (2015) appeared at the far-right corner of the scatterplot indicating this study *may* be exerting a substantial influence on the overall analysis and might be discrepant from the rest of the literature. The leave-one-out analysis did not result in a significant reduction in overall heterogeneity when omitting this study (a reduction in Higgins to I^2=73.4%). Consistent with visual inspection of the funnel plot in Figure 5, Speretto (2015) did not appear to be notably inconsistent with the rest of the literature.

Figure 5

Figure showing forest plot for studies measuring responding to joint attention (gaze and point).



Note – *A negative value indicates preterm children demonstrated reduced responding to joint attention (gaze and point) behaviours compared to full-term children.*

As Higgins I^2 >10% a series of meta regression analysis were conducted for each of the eight areas of study quality shown in Table 10. PT sample identification and blinding to group reached statistical significance suggesting these two quality domains were associated with changes in the effect size estimate of differences between PT and FT children's RJA (gaze

+ point) behaviours. Higher quality studies in these domains predicted lower effect size estimates.

Table 10

Table showing the impact of study level risk of bias for studies measuring responding to joint attention (gaze + point)

					Confidence interval	
Quality domain	**Estimate**	**SE**	**Z**	**P**	**Lower**	**Upper**
PT sample identification	**-0.87**	0.42	-2.06	**0.04**	-1.70	-0.04
PT sample description	-0.18	0.32	-0.56	0.57	-0.80	0.45
FT sample identification	-0.30	0.24	-1.28	0.20	-0.77	0.16
Comparability of groups	-0.25	0.22	-1.12	0.26	-0.68	0.19
Measure of social cognition	0.15	0.39	0.38	0.70	-0.63	0.93
Blinding	**-0.33**	0.149	-2.22	**0.03**	-0.63	-0.04
IR Reliability	-0.21	0.149	-1.43	0.15	-0.50	0.079
Statistical Bias	0.29	0.26	1.13	0.26	-0.22	0.82

Note *–bold indicates statistical significance.*

The funnel plot of included studies is presented in Appendix 3C. Visual inspection indicated approximate symmetry. Further inspection revealed some studies demonstrated markedly different effect estimates, with two studies demonstrating significant difference between PT and FT groups in the expected direction (to the left of the funnel plot presented in Appendix 3C) and three studies demonstrating significant difference in PT and FT children in the unexpected direction (to the right of the funnel plot), suggesting higher RJA (point + gaze) skills in PT children. Egger's test (Egger et al., 1997) was not appropriate due to >10 studies included in analysis.

Initiating Joint Attention[92] Results

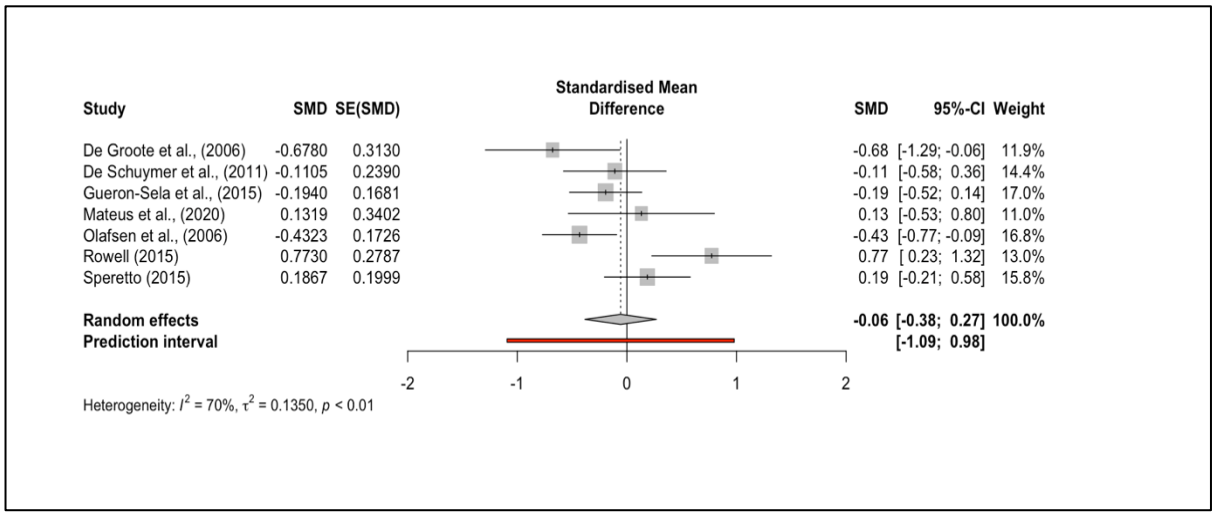
A total of seven studies measuring IJA were included in the meta-analysis. This included 201 PT children and 249 FT children. The random effects model was chosen as substantial heterogeneity was indicated (Higgins I^2=69.60%, tau^2=0.14, p=<.01). The model

[92] Initiating joint attention (point + gaze) will be referred to using the acronym IJA from here onwards.

suggested a weighted average SMD = -0.06, z= -0.34, p = 0.73 and a 95% confidence interval of between -0.38 and 0.27. Forest plot is presented in Figure 6.

Figure 6

Figure showing forest plot for studies measuring initiating joint attention.



Note *– A negative value indicates preterm children demonstrated reduced initiating joint attention behaviours compared to full-term children.*

Baujat scatterplot and the leave-one-out procedure (presented in Appendix 4C and Appendix 5C respectively) were inspected. Speretto (2015) appeared at the far-right corner of the scatterplot indicating this study may be exerting a substantial influence on the overall analysis and might be discrepant from the rest of the literature. The leave-one-out analysis did not result in a significant reduction in overall heterogeneity when omitted (a reduction in Higgins of I^2=71.0%). Consistent with visual inspection of the forest plot in Figure 6, Speretto (2015) did not appear to be notably inconsistent with the rest of the literature.

As Higgins I^2 >10% a series of meta regressions analysis were conducted for each of the eight types of study quality as shown in Table 11. FT sample identification reached statistical significance suggesting this quality domain was associated with changes in the effect

size estimate of differences between PT and FT children's IJA behaviours. Higher quality studies in these domains predicted higher effect size estimates.

Table 11

Table showing the impact of study level risk of bias for studies measuring initiating joint attention.

					Confidence interval	
Quality domain	**Estimate**	**SE**	**Z**	**P**	**Lower**	**Upper**
PT sample identification	0.29	0.48	0.59	0.56	-0.67	1.25
PT sample description	0.07	0.20	0.36	0.72	-0.32	0.47
FT sample identification	**0.47**	0.17	2.75	**0.01**	0.13	0.80
Comparability of groups	-0.00	0.17	-0.02	0.98	-0.33	0.33
Measure of social cognition	-0.30	0.34	-0.90	0.37	-0.96	0.36
Blinding	0.21	0.19	1.10	0.27	-0.16	0.59
IR Reliability	-0.06	0.17	-0.34	0.73	-0.38	0.27
Statistical Bias	0.08	0.29	0.28	0.78	-0.49	0.65

Note *–Bold indicates statistical significance.*

Visual inspection of the funnel plot (presented in Appendix 3D) suggested approximate symmetry. Further inspection revealed some studies demonstrated markedly different effect estimates, with two studies demonstrating significant difference between preterm and full-term groups in the expected direction (to the left of the funnel plot presented in Appendix 3D) and three studies demonstrating significant difference in PT and FT children in the unexpected direction (to the right of the funnel plot), suggesting improved IJA skills in preterm children. Egger's test (Egger et al., 1997) was not appropriate due to >10 studies included in analysis.

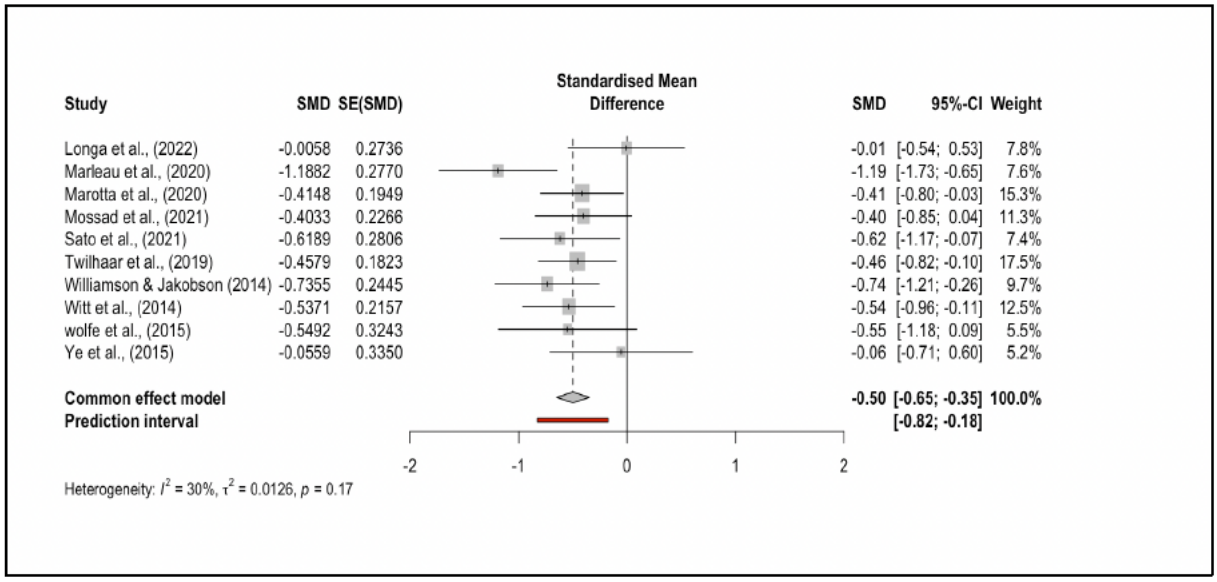
Affect Recognition

A total of ten studies measuring affect recognition were included in the meta-analysis. This included 358 PT children and 345 FT children. The fixed effects model was chosen due to effects being consistent within a normal distribution on the QQ plot (see Appendix 2E) and a low level of heterogeneity (Higgins I^2=29.60%, tau^2 = 0.013, p = 0.17). The model suggested a weighted average *SMD* = -0.49, z= -6.53, p <.01, and a 95% confidence interval of between

-0.65 and -0.35. Forest plot is presented in Figure 7. As Higgins I^2 was <50 % Baujat scatter plot and leave-one-out analysis was not inspected.

Figure 7

Figure showing forest plot of studies measuring affect recognition.



Note *– A negative value indicates preterm children demonstrated reduced affect recognition compared to full-term children.*

A series of meta regression analyses were conducted for each of the eight domains of study quality as shown in Table 12 (as Higgins I^2 >10%). None of the domains of quality showed a statistically significant association with the effects reported in the included studies. However, FT sample identification did show a *trend* toward statistical significance with higher quality studies in this domain predicting a lower effect size estimate between groups.

Table 12

Table showing the impact of study level risk of bias for studies measuring affect recognition.

					Confidence interval	
Quality domain	**Estimate**	**SE**	**Z**	**P**	**Lower**	**Upper**
PT sample identification	-0.03	0.13	-0.29	0.77	-0.29	0.21
PT sample description	-0.16	0.19	-0.85	0.39	-0.54	0.21
FT sample identification	**-0.14**	0.09	-1.67	**0.09**	-0.31	0.03
Comparability of groups	0.11	0.12	0.96	0.34	-0.12	0.34
Measure of social cognition	-0.27	0.21	-1.29	0.19	-0.67	0.14
Blinding	-	-	-	-	-	-
IR Reliability	-	-	-	-	-	-

Statistical Bias	-0.09	0.14	-0.73	0.47	-0.37	0.17

Note -*Blinding and inter-rater reliability in included studies within this meta-analysis had the same quality scores and so estimates not computed during analysis.*

Visual inspection of the funnel plot (presented in Appendix 3E) demonstrated approximate symmetry. Further inspection indicated that most of the studies fell in the area of non-statistical significance (white shading) rather than the areas of significance (red and orange shading), and the funnel appeared symmetrical. This was confirmed via the Eggers test (Egger et al., 1997) which revealed a non-significant effect (p=0.97) and thus no clear evidence of publication bias in the distribution of effect.

Theory of Mind (ToM)[93]

A total of eleven studies measuring ToM were included in the meta-analysis. This included 397 PT children and 407 FT children.

Initial analysis including *all* studies identified no evidence of non-linearity, however there was suggestion of a possible outlier when examining QQ plots for both fixed and random effects model, which are presented in Appendix 6A. A substantial level of heterogeneity was observed (Higgin's I^2 = 77.2%, tau^2 = 0.28, p= <.01) and examination of the Baujat scatterplot revealed a possible outlier (presented in Appendix 6B). The robustness of the results was therefore tested via outlier and sensitivity analysis by the leave-one-out method (presented in Appendix 6C), which revealed an influential study (Marleau et al., 2020), with heterogeneity reducing to Higgin's I^2 = 0.0%, tau^2 = 0.01, p= <.01when this study was omitted. This study was re-examined to identify any risks of bias which might account for this discrepancy. Marleau et al. (2020) measured cognitive ability finding significant differences between groups and so differences in ToM between PT and FT children cannot be reliability attributed to prematurity. This study also appeared different from the rest of the literature; in that it

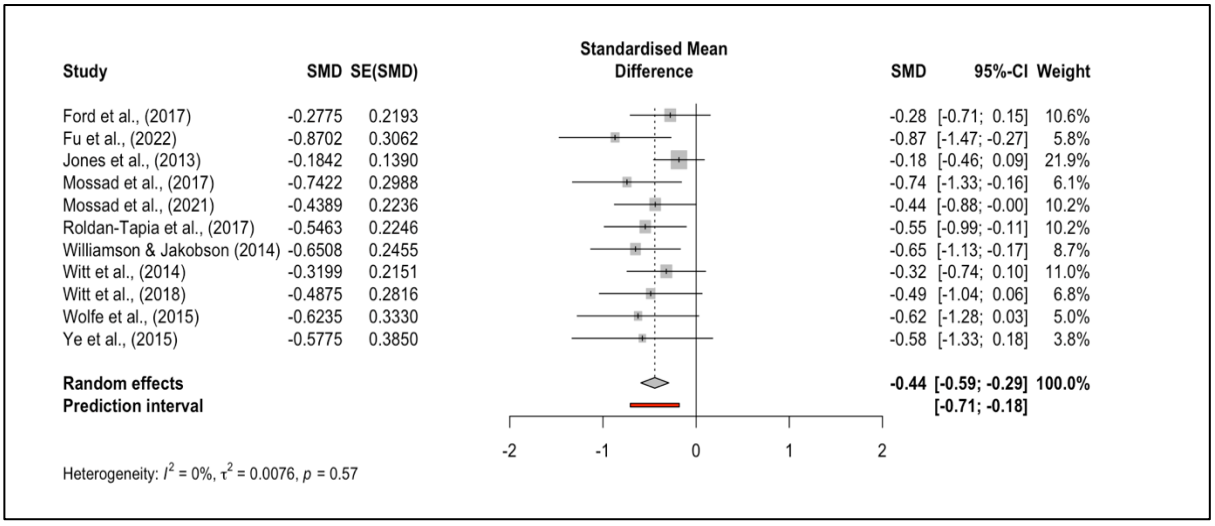
[93] Theory of mind will be referred to using the acronym ToM from here onwards.

suggested a large difference in ToM between the groups as compared to other studies. There was therefore a clear rationale to exclude this outlier from the analysis. The distribution of primary study effects, *after* removal of Marleau et al. (2020) are presented in QQ plots in Appendix 2F. There was no evidence of non-linearity the distribution of the standardised mean difference within either the fixed effects model or the random effect model after this study as removed.

After removal of identified outlier, the fixed effects model was used due to effects being consistent within a normal distribution on the QQ plot and a low level of heterogeneity (Higgins I^2=0.00%, tau^2 = 0.01, p = 0.57). The model suggested a weighted *SMD* = -0.42, z=-6.06, p= <.01and a 95% confidence interval of between -0.57 and -0.29. See Figure 8 for forest plot. As heterogeneity was <10% further exploration of the impact of study quality was not necessary.

Figure 8

Figure showing forest plot for studies measuring theory of mind.



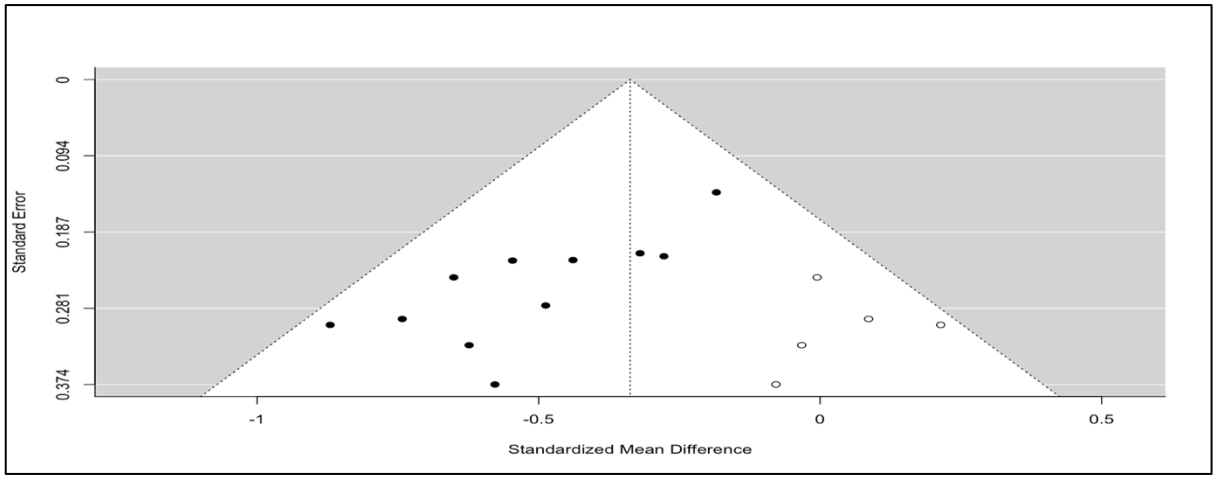
Note – *A negative value indicates preterm children demonstrated reduced ToM compared to full-term children.*

Visual inspection of the funnel plot suggested *some* asymmetry in the distribution of effect (presented in Appendix 3F). Closer inspection demonstrated that most of the studies fell in the area of non-statistical significance (white shading) rather than the areas of significance

(red and orange shading), hence the funnel asymmetry was unlikely to be attributed to publication bias (Palmer et al., 2008). Eggers test for a regression intercept (Egger et al., 1997) did however give a *p* value of $p = <.01$, indicating possible publication bias. The effect of publication bias was simulated using a trim-and-fill procedure (Duval & Tweedle, 2000). The trim-and-fill procedure yielded a corrected random effects model of effect = -0.33 (95% CI -0.47 to -0.19). The corrected random effects model evidences an approximately -24.69% decrease relative to the uncorrected estimate. Rosenthal method was not required as the trim-and-fill procedure provided an appropriate correction. This correction resulted in a reduction in the overall effect however this remained in the expected direction (i.e., PT children were still performing worse than FT children on measures of ToM). Figure 9 presents the funnel plot *after* trim-and-fill procedure.

Figure 9

Funnel plot of the standardised mean differences in theory of mind between preterm and full-term children after trim-and-fill procedure. The 95% confidence interval of the expected distribution of differences in ToM is shown as an inverted "funnel".



Narrative Review

Studies measuring other aspects of social cognition were retained for narrative review. The small number of studies measuring multiple aspects of social cognition, with different

measurement methods did not allow for meta-analytic statistical methods. Studies measuring different aspects of joint attention (such as point and gaze behaviours separately) have also been summarised here. Due to the small number of studies measuring these behaviours in this way, it was not possible to combine for meta-analysis.

Social Attention and Perception

Studies were categorised into five areas under the overarching domain named social attention and perception. This included social preference (meta-analytic results already described), face/eye preference (*K*=4), face recognition (*K*=2), imitation (*K*=2), biological motion perception (*K*=2).

Face Recognition. Perez-Roche et al. (2017) found PT children demonstrated poorer immediate face recognition but comparable face memory to FT children. After adjusting for birthweight, prematurity was no longer associated with differences in outcomes for both face recognition and face memory. Frie et al. (2016) found PT children demonstrated a different pattern of cortical face recognition process compared with FT children.

Imitation. Two studies measuring imitation in PT children as compared to FT children were included in the final list of papers. Haley et al. (2010) used a deferred imitation task (Barr et al., 1996) to evaluate social learning. They hypothesised that PT children would show less imitation behaviour than FT children but found no differences between groups. Linares (1986) developed an imitation battery containing four types of imitation. The social gestures category was used as a measure of social communication/imitation and found no significant differences between PT and FT children in imitation behaviours.

Face and Eye Preference. Sekigawa-Hosozawa et al. (2017) used dynamic social scenes with multi-dimensional scaling to depict the relationship between variables. PT children were classified into two groups, one which demonstrated a similar MDS plane to FT children and one which was most similar to autistic children included in the study. Face viewing

proportion time was higher for the FT and PT children with a similar MDS plane. In comparison, PT children with a similar MDS pattern to autistic children spent less time viewing faces in the videos used in the study. There were no significant differences between groups in proportion of time spent viewing faces. Yamamoto et al. (2020) found PT children spent significantly less time viewing the eyes of faces, and less time viewing facial stimuli generally compared to FT children. Pereira et al., (2017) found differences in preference for a natural or distorted face in the first few hours of life, with PT children demonstrating no preference between a natural or distorted face. FT children demonstrated a preference for natural faces over distorted faces.

Biological Motion Perception. Two papers measured biological motion perception defined as the ability to perceive and be sensitive to cues of biological/social motion (Johansson, 1973; Taylor et al., 2009) and these met criteria for inclusion in this study as per the definition of social cognition which has been given. Taylor et al. (2009) measured sensitivity to global form, global motion and biological motion and found PT children demonstrated reduced sensitivity for biological motion and global motion, but no impairment for global form when compared to FT children. Analysis did not identify any correlations between the three measures, however.

Williamson et al. (2015) tested whether children could perceptually group local motion cues to extract information related to body structure. The authors suggest these are higher order processes which are needed for action recognition and person identification. PT children found all tasks for biological motion perception more difficult than FT children. Additionally, the four different tests were not correlated with one another, which was also the case in Taylor et al. (2009).

Joint Attention

Responding to Joint Attention (Gaze and Point Behaviours). Rowell (2015) measured RJA (point behaviours) in isolation from gaze behaviours and found significant differences, with PT children underperforming compared to FT children. Table 6 shows information for Rowell (2015), in addition to statistical information for RJA (point behaviours) in isolation from gaze behaviours extracted from the study. Yamaoka and Takada (2016) created the joint attention observation scale (JAOS) and found no differences between PT and FT children in RJA when experimenters used gaze and pointing behaviours together. They found significant differences between groups, with preterm children underperforming, when only one "RJA clue" was given. PT children did not show as many RJA behaviours when experimenters used only eye orientation in attempts to establish joint attention.

Initiating Joint Attention. De Schuymer et al. (2011) and Rowell (2015) reported IJA point behaviours separately from gaze behaviours. Yamaoka and Takada (2016) measured IJA gaze behaviours in a way which was not possible to combine with other studies and so it was retained for narrative review. Rowell (2015) found PT children demonstrated more IJA gaze behaviours than FT children. However, PT children displayed fewer IJA point behaviours than FT children. De Schuymer et al. (2011) found PT children to demonstrate fewer IJA point and gaze behaviours compared to FT, but this did not reach statistical significance. Yamaoka and Takada (2016) found inconsistent results, with PT children underperforming in IJA gaze behaviours compared to FT children on some items in the JAOS and presenting with no differences on other items measuring the same construct.

Social Reasoning & Empathy

A subset of papers were grouped together under the heading of *social reasoning & empathy.* These papers explore the processing of social cues and use of social knowledge (Beer

& Ochsner, 2006). Three studies measuring social reasoning and one study measuring empathy were included in the final list of papers.

Marotta et al. (2020) used a task which required children to understand and explain social situations as depicted in pictures and vignettes and to describe what occurs in each image and how the protagonist feels and why (Ferri et al., 2015; Grob et al., 2009; Marotta et al., 2020). The groups did not differ significantly in their social reasoning abilities.

Lejeune et al. (2016) used a task which assessed children's ability to judge, identify and reason about another's behaviour in relation to conventional and moral rules (Barisnikov, 2005; Lejune et al., 2016). This task provided a global score and three subscale scores based on whether children could identify a situation as correct or not (Judgment score), identify what was correct or incorrect about a situation (identification score) and whether they can identify why a situation is correct/incorrect (social reasoning score). Overall, PT children scored significantly lower (with moderate effect size) on the global social reasoning score, suggesting increased difficulty in their understanding of socially appropriate behaviours than FT peers. Further analysis demonstrated PT children had difficulties in detection of the pertinent social cues in the task, typically used when making judgements about social situations (as described by their identification subscale score). These skills are important in identifying the social rule transgression (Lejune et al., 2016).

Williamson and Jakobson (2014) used the child and adolescent social perception measure - CASP (Magill-Evans et al., 1995). The emotion score of this measure has been included within the affect recognition meta-analysis in this paper. The CASP provides additional information of relevance here. A non-verbal cue score assesses a child's ability to identify specific face, body, situational and voice cues which are often used when determining an individual's emotional state (Williamson & Jakobson, 2014). After controlling for cognitive ability, PT children performed significantly worse in the identification of situation cues.

Campbell et al. (2015) investigated empathic competence using the Kids Empathic Development scale – KEDS (Reid et al., 2013). This task is a measure of the cognitive, affective, and behavioural components of empathy. PT children performed sufficiently worse compared to FT children, even after controlling for cognitive ability.

All results presented here have been tabulated to provide a clear summary of results across all social cognitive domains. This is presented in Table 13. Study-level effects for differences between PT and FT children on each social-cognitive domain where meta-analysis was conducted are reported in Appendix 7. This information has been visually presented already throughout the results section in forest plots for each meta-analysis.

Table 13

Table summarising outcomes across social cognitive domains included in meta-analysis and narrative review.

Domain of social cognition	Meta/ Review	Effect size	Confidence interval (CI)	Summary of outcome *Effect is in comparison to full-term children*
Early social attention and preference				
Social preference		**-0.35**	-0.69 to -0.11	Reduced preference in PT children (small effect). Range of values in CI is large but negative effect remains.
Face recognition		-	-	Reduced face recognition in PT children & different cortical processes employed by PT children.
Face/eye preference		-	-	Overall reduction with PT children spending less time viewing faces/eyes
Imitation		-	-	No differences in imitation between PT & FT children
Biological motion perception		-	-	Reduced sensitivity to biological motion perception in PT children
Joint attention behaviours				
RJA (gaze following)		-0.03	-0.44 to 0.38	Negligible difference. Range of values in CI is large reaching small/medium increase and decrease in GF abilities in PT children.
RJA (gaze + point)		0.05	-0.89 to 0.18	Negligible difference. Range of values in CI is large, ranging from a large difference (-0.8918) to small/negligible effect (0.1798)
RJA (point)		-	-	PT demonstrate reduced pointing RJA behaviours
Initiating joint attention behaviours				
IJA		0.06	-0.38 to 0.27	Negligible difference in IJA between PT & FT children. CI range is large demonstrating small negative effect and small positive.
IJA (point)		-	-	PT show reduced pointing IJA behaviours.

Domain of social cognition	Meta/ Review	Effect size	Confidence interval (CI)	Summary of outcome *Effect is in comparison to full-term children*
Affect recognition		**-0.49**	-0.65 to -0.35	Reduced affect recognition in PT children (moderate effect). CI range within negative effect
ToM		**-0.43**	0.57 to -0.29	Reduced TOM skills in PT children (moderate effect). CI range within negative effect
Social reasoning & empathy		-	-	Overall reduction in social reasoning & empathy skills in preterm children

Note – *Boxes in green indicate narrative review, whilst those in blue indicate separate meta-analysis; Bold highlights significant decreases in preterm children compared to full-term children.*

Discussion

This systematic review and meta-analysis aimed to investigate whether children born PT demonstrate differences in social cognition compared to their FT peers. It is important to discuss the results of each social cognitive domain separately to be able to comment on individual factors important within each analysis. However, as highlighted in the Introduction, the focus on one skill in isolation limits broader interpretations. Discussions have therefore been structured to offer individual commentary alongside a broader interpretation of results. Table 13 presents a summary of results that allows for direct comparison.

Early Social Attention.

The pooled effect size estimate of studies investigating *social preference* suggested PT children demonstrated decreased preference for social images compared to FT children, with a small effect size (Cohen, 1960), however a small number of studies was included in analysis *(k*=3*)*. Two studies were longitudinal measuring social preference across multiple time points. When looking at the data across time, Dean et al. (2021) found PT children "catch up" with their full-term peers by the age of 5. Imafuku et al. (2021) showed differences remained consistent at six, 12 and 18 months of age. There are no studies investigating social preference between the age of 18 months to 5 years which limits our knowledge of the trajectory of differences in social preference. Longitudinal studies, employing multiple time points and robust paradigms are essential in developing a more comprehensive understanding of these initial differences.

Attending to other aspects of our world which are also conceptualised as essential for social development include attention to eye/face preference (Perez-Roche et al., 2017). Very few studies explored this but overall, the narrative review indicated differences in attending to these social aspects (eyes/face), with PT children spending less time preferentially attending to eyes/faces in comparison to FT children. A reduction in preference to stimuli considered

important for social development such as face/eyes, alongside a reduction in preference for social stimuli when compared to non-social stimuli (Dean et al., 2021; Imafuku et al., 2017: Imafuku et al., 2021) is interesting and warrants further exploration. Methodologies used to explore social preference are particularly important to consider. Studies using static paradigms (as in Dean et al., 2021) may fail to represent real-life social stimuli thereby limiting conclusions, and perhaps mask underlying differences across different groups which may be present in real-life social situations (Risko et al., 2012).

Joint Attention

Overall effect size estimates within each domain of joint attention suggested no differences between PT and FT children. There was however high variation across studies and wide confidence intervals suggesting unreliability or heterogeneity due to undetected differences between studies. This was demonstrated across meta-analyses measuring different joint attention behaviours. Within the gaze following meta-analysis, one study represented just under half of included samples (Pena et al., 2014), all of which suggested superior gaze-following abilities in PT children compared to those born FT. It is important to highlight the identification of the PT sample in this study was rated as poor quality (see Table 6 for presentation of quality assessment rating). Meta-regression of quality ratings did not however indicate any change in estimates based on this aspect of quality. Given the over-representation of this sample in the analysis, it is important to hold this in mind, particularly as we cannot establish how representative this sample is of the PT population.

The pooled effect size estimate for RJA (gaze + point) indicated no evidence of a difference between PT and FT children. Interestingly, when gaze and point behaviours are combined the weighted average effect and confidence interval lies more in the expected direction - below zero (see Figure 5 presenting forest plot), and individual studies more consistently demonstrate fewer RJA behaviours in PT children compared to FT children. This

contrasts to the pattern seen in gaze-following alone (see Figure 4 presenting forest plot). Only one study measured point behaviours in isolation from gaze responding to joint attention behaviours (Rowell, 2015). Interestingly, Rowell (2015) found significant differences, with PT children underperforming in response to point behaviours compared to FT children, whilst PT children performed better at RJA (gaze behaviours) although this did not reach statistical significance.

The pooled effect size estimate for IJA (gaze + point/showing) indicated no difference between PT and FT children. There was high variation across studies with some indicating no differences in abilities and others indicating a reduction in abilities in PT children. Likewise, studies measuring IJA gaze and point/showing behaviours independently suggested PT children perform less point/showing behaviours than FT children.

Looking at the analyses of different aspects of joint attention behaviours together, the current analyses extends Mateus et al.'s, (2019) meta-analysis with the inclusion of additional studies and by measuring gaze-following in isolation from other behaviours associated with responding to joint attention. We revealed similar pooled effect size estimates and the same methodological limitations making further exploration of heterogeneity impossible. The measurement and reporting of different joint attention behaviours such as gaze and point/show are important to understand possible nuanced differences between PT and FT children. More work is required to better understand potential differences between separate components of joint attention behaviours and possible covariates such as gestational age and cognitive ability.

Affect Recognition

The pooled effect size estimate for affect recognition indicated PT children showed lower affect recognition abilities compared to FT children. This difference approached a moderate level of effect (Cohen, 1960); however, this may not be a reliable estimate. Preterm children included in these studies were predominantly very preterm (28-<32 weeks GA),

however this must be considered alongside the inconsistences in the reporting of demographic data described. One study (Marotta et al., 2022) measured affect recognition in moderate to late preterm children in isolation from earlier GAs (34-<37 weeks GA). The forest plot (see Figure 7 for presentation of forest plot) suggests preterm children in Marotta et al. (2022) showed similar reductions in affect recognition as compared to FT children as those of PT children born at earlier GAs. The low heterogeneity suggests this body of studies reported a coherent and consistent effect size and therefore we can be more confident that children born at younger GAs demonstrate differences in affect recognition abilities compared to FT children.

Theory of Mind (ToM)

The pooled effect size estimates of ToM indicated PT children showed a reduction in ToM abilities compared to their FT peers. This difference approached a moderate level of effect (Cohen. 1960). PT children included in these studies were all either extremely preterm (<28 weeks GA) or very preterm (28-<32 weeks GA). Importantly the mean ages fell very close to 28 weeks GA and therefore results here can only be extrapolated to younger GAs. The low heterogeneity across studies suggests consistency within the literature, therefore we can be more confident that children born at younger GA's demonstrate differences in ToM compared to FT children. Given the arguably more robust differences identified in ToM abilities, it is interesting that at age 5 Dean et al. (2021) observed no differences in social preference, yet differences in affect recognition and ToM were observed in the current meta-analyses. More work is required to better understand the role of social preference in the development of these later social cognitive skills.

For both affect recognition and ToM, despite the low heterogeneity, it is important to remain cautious. The differences found between PT and FT children may not be a reliable estimate of true differences attributed to prematurity as we were not able to test other moderators due to insufficient data.

Social Reasoning and Empathy

The limited number of studies measuring later social cognitive skills, such as social understanding and empathy make concrete conclusions about differences between PT and FT children difficult. The lack of studies identified measuring these social cognitive domains are likely impacted by the majority of studies including children of lower ages. Studies were generally suggestive of impairment of social reasoning abilities in PT children. Particularly interesting were the difficulties in detection of pertinent social cues in social reasoning tasks in PT children and also reduced biological motion perception compared to FT children. When these results are considered together, it is interesting that there is a suggestion of differences in PT children's abilities to use certain cues in making sense of their social worlds. Given affect recognition and ToM abilities are shown to be impaired, further exploration of these later social cognitive skills may be warranted.

Methodological Limitations

Limitations of Included Studies

The quality of a meta-analysis is only as high as the quality of the studies it includes (Todorova et al., 2019). When analysing the impact of aspects of quality via meta-regressions, very few reached statistical significance across meta-analyses and therefore changes in the quality of a study did not appear to contribute enough to influence the results within each domain. This is interesting, however there are many possible explanations for this. The low statistical power of the analyses due to the relatively small number of papers included could explain this, therefore we must be cautious of interpreting null findings here. Additionally, it could also be that important aspects of methodological quality either were not covered in the quality framework tool developed for this study *or* the large amounts of variability found in results is related to something other than quality – such as the heterogeneity across gestational ages. This, alongside the poor measurement and reporting of important descriptors across

studies (such as cognitive ability, health status, GA, SES and ethnicity) is reasonable grounds for careful description and categorisation of PT *and* FT samples in future studies. Currently, the PT cohort to which these results can be extrapolated is not clear *and* the impact of different GA categories on the different domains of social cognition assessed cannot yet be clearly understood.

Broadly, tasks used to measure social cognition across domains were high quality with clear descriptions of tasks and the use of well validated and replicable paradigms. Despite this, variability across different measures of the same construct cannot be ruled out as having a probable impact on heterogeneity. In the ToM literature concerns about the reliability and validity of tasks and their utility in locating differences in high-risk samples such as those with cognitive impairment (Cashion, 2009; Jones et al., 2013) have been noted. One study (Marleau et al., 2020) was removed from the ToM analysis which found significant differences in cognitive ability in their sample and results appeared to be inconsistent with remaining studies in this social cognitive domain. Interestingly however, when Marleau et al. (2020) controlled for IQ, differences in ToM remained. The authors suggested that the more detailed battery of tasks used in this study may have allowed for more nuanced identification of differences. Therefore, methodology may also be important to consider, particularly to understand subtle differences across different social-cognitive skills.

Aim 3 of this meta-analytic review (to identify participant characteristics that may be associated with differences in social cognition in PT and FT children) was not possible to meet. The lack of exploration of moderators due to insufficient data is problematic and limits our understanding of the influence of other important variables which may account for social cognitive differences outside of prematurity – such as cognitive ability, fine motor skills, age at assessment, ethnicity and socio-economic status to name a few.

Despite these limitations, this meta-analysis has been the first to assemble research across multiple social cognitive domains to begin to quantify differences in social cognition between PT and FT children. This in-depth synthesis and analysis have enabled identification of key problems in this research area which will hopefully contribute to re-directing future studies exploring social cognition in PT children and adolescence.

Limitation of Current Meta-Analysis

The current meta-analysis is subject to several limitations. Studies were categorised into social cognitive domains according to definition. Due to the broad range of methodologies used within each meta-analysis, we cannot be certain that they are indeed all measuring the same construct in the same way. This may be particularly important in studies measuring different aspects of joint attention behaviours. The high heterogeneity observed may suggest the meta-analytic results here are not useful in summarising differences between groups in a meaningful way. Additionally, the low quality of included studies may bias results and as such study inclusion may have benefited from a stricter criterion regarding study quality (Lee, 2019).

Certain aspects of the review process may also be subject to bias. The quality framework was devised for the analysis based on aspects of quality considered important by the research team. This may be problematic as it has not been tested against broader aspects of validity. The broadness of the search terms chosen for the literature search (predominantly social cognition) over the use of more specific terms may have had an impact on the identification of all relevant studies measuring constructs of interest. The diverse range of definitions and terms related to social cognition alone (Rose-Krasnor, 1997; Rubin et al., 1995; Yeates et al., 2007), may provide argument for the inclusion of additional, more specific search terms for different aspects of social cognition. However, the use of more discrete search terms *could* perhaps risk excluding relevant papers, whereas the broader approach chosen may work to be over-inclusive. Given the large number of papers the current search returned, in addition

to the selected terms being guided by those used by similar reviews in the area, there is a strong argument for the search strategy employed here. Additionally, only one reviewer undertook searches which may have introduced bias into the process of study selection.

Despite these limitations, this meta-analysis has been the first to assemble research across multiple social cognitive domains to *begin* to quantify differences in social cognition between PT and FT children. This in-depth synthesis and analysis has enabled identification of key problems in this research area which will hopefully contribute to re-directing future studies exploring social cognition in PT children and adolescence.

Clinical Implications

Awareness of prematurity as a risk factor for social cognitive differences has important clinical implications. Early identification and intervention are essential in the treatment and prognosis of children with social difficulties (Clark et al., 2017; Fuller & Kaiser, 2019; Luu et al., 2020). ToM is suggested to be correlated with real-life social skills (Frith et al., 1994, Fletcher-Watson et al., 2014) and early intervention for PT children with difficulties in this area may lessen the social and communication difficulties which often accompany this and have a positive impact on future development (Fletcher-Watson et al., 2014; Luu et al., 2020). Increased monitoring/assessment of early social cognitive skills important for social development in children born prematurely is suggested given the differences identified in affect recognition and ToM abilities.

Clinical implications with regards to intervention for those with difficulties in affect recognition and ToM need to be considered in context. Meta-analysis has shown interventions targeting ToM and other related skills such as emotion recognition and joint attention behaviours demonstrate positive effects in autistic children (Fletcher-Watson et al., 2014). However, current evidence is poor with regards to the ability of available interventions to

maintain improvements over time and generalize to other settings (fletcher-Watson et al., 2014).

Another important clinical implication of possible social cognitive differences associated with prematurity, relates to autism *assessment* and screening. The suggestion of a unique prematurity-based phenotype of autism (Johnson & Marlow, 2011) may indicate that social cognitive differences in PT children as compared to FT children are more subtle and nuanced, which may not be picked up at a standardised developmental assessment. This meta-analysis supports characterisation of possible subtle social cognitive differences, the identification of which may support tailored interventions. Alongside the relatively new guidelines on follow-up for premature birth (NICE, 2017), this work may provide further evidence to support the importance of enhanced follow up for children born prematurely.

Future Directions

Results here suggest important differences is social cognitive skills such as affect recognition and ToM in children born at earlier gestational ages. This is important as these differences may result in challenges in social functioning and in the development of social competence. The inconclusive evidence found in earlier social cognitive skills means we know very little about the development and trajectory of these foundational skills which are important for later developing skills such as affect recognition and ToM. Many questions remain unanswered such as whether there is a maturational delay of these skills? Do these differences level out over time? *or* do differences emerge later in development?

Future studies in this field must acknowledge the crucial limitations described here. To drive knowledge forward about differences in social cognition between PT and FT children, three research recommendations are suggested:

1. Well defined and reported key demographics of both PT and FT children will allow for analysis of important factors which may influence the relationship between prematurity

and social cognitive differences. Crucially, clearly defined gestational age ranges in line with categories widely reported *in addition* to measuring and reporting of cognitive ability, ethnicity, and socio-economic status.

2. Emerging studies to focus on well-defined *earlier* social cognitive skills to understand trajectories to later social cognitive skill development and nuances in development of skills which may contribute to discrete divergent trajectories of development.
3. Longitudinal studies spanning early infancy to later adolescence are crucial to better understand nuances in PT developmental trajectories.

Conclusions: Social Cognition and Prematurity

The most compelling evidence here is that affect recognition and ToM abilities may be impaired in children born at younger GAs. However, we must be cautious with conclusions directly attributing these differences to prematurity alone, without additional research. ToM plays a critical role in the development of social competence (Becht et al., 2021; Blakemore & Mills, 2014; Burnett et al., 2011; Fu et al., 2022) and is a skill which relies on earlier social cognitive skills and continues to develop across adolescence. Montagna and Norsarti (2016) describe the developmental cascades model from PT birth to neural alterations in the social brain network which then go on to influence social cognitive functioning (Fu et al., 2022). The substantial variability across studies measuring *earlier* social cognitive skills evidenced here demonstrates how current research tells us very little about the developmental trajectory of social cognitive skills in PT children. Crucially, older preterm gestational ages are largely neglected in the literature, however, methodological limitations across studies also limit knowledge of earlier social cognitive skills in those born at earlier gestational ages as well.

References

Aarnoudse-Moens, C. S. H., Weisglas-Kuperus, N., van Goudoever, J. B., & Oosterlaan, J. (2009b). Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. *Paediatrics* 124, 717–728. Doi: 10.1542/peds.2008-2816.

Abell, F., Happe´, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development, *Cognitive Development,* 15, 1–16. Doi: 10.1016/S0885-2014(00)00014-9.

Agrawal, S., Rao, S.C., Bulsara, M.K., & Patole, S.K. (2018). Prevalence of Autism Spectrum Disorder in preterm infants: a meta-analysis, *Paediatrics,* 142 (3). Doi: 10.1542/peds.2018-0134 e20180134.

Alaerts, K., Nackaerts, E., Meyns, P., Swinnen, S. P., & Wenderoth, N. (2011). Action and emotion recognition from point light displays: an investigation of gender differences. *PLOS One,* 6:e20989. Doi: 10.1371/journal.pone. 0020989.

Arpi, E., & Ferrari, F. (2013). Preterm birth and behaviour problems in infants and preschool-age children: a review of the recent literature. *Developmental Medicine & Child neurology,* 55, 788–796. Doi: 10.1111/dmcn.12142.

Barisnikov, K., Van der Linden, M., & Hippolyte, L. (2005). *Tâche de Résolution Sociale.* Geneva (unpublished).

Baron-Cohen, S., O'Riordan, M., Jones, R., Stone, V., & Plaisted, K. (1999). A new test of social sensitivity: Detection of faux pas in normal children and children with Asperger syndrome, *Journal of Autism and Developmental Disorders,* 29, 407–418. Doi: 10.1023/a:1023035012436.

Barr R, Dowden A, & Hayne H. (1996). Developmental changes in deferred imitation by 6- to 24-month-old infants, *Infant Behavior and Development*, 19:159–170. Doi: 10.1016/S0163-6383(96)90015-6.

Becht, A.I., Wierenga, L.M., Mills, K.L., Meuwese, R., Van Duijvenvoorde, A., Blakemore, S. J., Guroglu, B., & Crone, A. E. (2021). Beyond the aver- age brain: individual differences in social brain development are associated with friendship quality, *Social Cognitive and Affective Neuroscience,* 16(3), 292–301. Doi: 10.1093/scan/nsaa166.

Bergwerff, C., Luman, M., Meffert, H., Blair, R., & Oosterlaan, J (2016). Measuring social cognition in school-aged children using a morphed facial emotion recognition task. In: Bergwerff C, ed. Food for thought: novel insights into childhood ADHD. Amsterdam: Vrije Universiteit; 2016. p. 125-45.

Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., & Anand, K. J. S. (2002). Cognitive and behavioural outcomes of school-aged children who were born preterm: a meta-analysis. *JAMA,* 288, 728–737. Doi: 10.1001/jama.288. 6.728.

Bhutta, A., Cleves, M., & Casey, P. (2014) Cognitive and behavioural outcomes of school-aged children who were born preterm: a meta-analysis, *The Journal of the Medical Association,* 288:728–37. Doi:10.1001/jama.288.6.728.

Blakemore, S.-J., Mills, K.L. (2014). Is adolescence a sensitive period for sociocultural processing? *Annual Review of Psychology*, 65, 187–207. Doi: 10.1146/annurev-psych-010213-115202.

Blencowe, H., Cousens, S., Chou, D., Oestergaard, M., Say, L., Moller, A. B., Kinney, M., & Joy, L. (2013). Born too soon: The global epidemiology of 15 million preterm births, *Reproductive Health,* 10 (Suppl 1): S2. www.reproductive-health-journal.com/content/10/S1/S2.

Blencowe, H., Cousens, S., Oestergaard, M. Z., Chou, D., Moller, A.B., & Narwal. R. (2012). National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications, *Lancet,* 379 (9832): 2162e72. Doi: 10.1016/S0140-6736 (12) 60820-4.

Bos, F. A., Van Braeckel, K.N.J., Hitzert, M. M., Tanis, C. J., & Rose, E. (2013). Development of fine motor skills in preterm infants, *Developmental Medicine & Child Neurology,* 55 (4): 1-4. Doi: 10.1111/dmcn.12297.

Boyle, E.,., Johnson, S., Manktelow, B., Seaton, S. E., Draper, E. S., Smith, L. K., Dorling, J., ,arlow, N., Petrou, S., & Field, D. J. (2015). Neonatal outcomes and delivery of care for infants born late preterm or moderately preterm: a prospective population-based study, *Archives of Disease in Childhood. Fetal & Neonatal Edition,* 100: F479-F485. Doi: 10.1136/archdischild-2014-307347.

Bradshaw, J., Schwichtenberg, A. J., & Iverson, J. M. (2022). Capturing the complexity of autism: Applying a developmental cascades framework, *Child Development Perspectives,* 16: 18-26. Doi: 10.1111/cdep.12439.

Burnett, S., Sebastian, C., Cohen Kadosh, K., & Blakemore, S. J. (2011). The social brain in adolescence: Evidence from functional magnetic resonance imaging and behavioural studies, *Neuroscience & Biobehavioural Reviews,* 35, 1654-1664. Doi: 10.1016/j.neubiorev.2010.10.011.

*Campbell, C., Horlin, C., Reid, C., McMichael, J., Forrest, L., Brydges, C., French, N., & Anderson, M. (2015). How do you think she feels? Vulnerability in empathy and the role of attention in school-aged children born extremely preterm, *British Journal of Developmental Psychology,* 33, 312-323. Doi: 10.1111/bjdp.12091.

Capobianco, M., & Cerniglia, L. (2017). Early language development in preterm children without neurological damage: a longitudinal study, *F1000 Research,* 6: 2169. Doi: 10.12688/f1000research.13314.1.

Carlsson, J., Lagercrantz, H., Olson, L., Printz, G., & Bartocci, M (2008). Activation of the right fronto-temporal cortex during maternal facial recognition in young infants. *Acta Paediatrics,* 97:1221-5. Doi: 10.1111/j.1651-2227.2008.00886.

Cashion, L. (2009). Theory of mind performance in middle childhood: Australian normative and validation data, *The Australian Educational & Developmental Psychologist,* 26 (2): 138-53. Doi: 10.1375/aedp.26.2.138.

Castelli, F., Happe´, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns, *NeuroImage, 12*, 314–325. Doi: 10.1006/nimg.2000.0612.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16* (4), 231–239. Doi: 10.1016/J.TICS.2012.02.007.

Choudhury, S., Blakemore, S. J., & Charman, T. (2006). Social cognitive development during adolescence, *SCAN,* 1, 165-174. Doi: 10.1093/scan/nsl1024.

Clark, M. L. E., Barbaro, J., & Dissanayake, C. (2017). Continuity and change in cognition and autism severity from toddlerhood to school age, *Journal of Autism & Developmental Disorders,* 47: 328-39. Doi: 10.1007/s10803-016-2954-7.

Clark, M. L. E., Vinen, Z., Babaro, J., & Dissanayake, C. (2018). School age outcomes of children diagnosed early and later with autism spectrum disorder, *Journal of Autism & Developmental Disorders,* 48: 92-102. Doi: 10.1007/s10803-017-3279.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*, 37–46. Doi: 10.1177/001316446002000104.

Coursel, A., & Wagner, E. E. (1986). Effect of positive findings on submission and acceptance rates: A note on meta-analysis bias. *Professional Psychology: Research & Practice,* 17 (2), 137-137. Doi: 10.1037/0735-7028.17.2.136.

Covernton, E., Richards, C., Surtees, A., &, Laverty, C. (2021). Do children born preterm present with differences in social cognition as compared to their full-term peers?. *PROSPERO*, CRD42021229456. Available

from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021229456.

Crick,N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment, *Psychological Bulletin,* 115 (1): 74-101. Doi: 10.1037/0033-2909.115.1.74.

Dalrymple, K.A., Gomez, J., & Duchaine, B (2013). The Dartmouth Database of Children's Faces: Acquisition and Validation of a New Face Stimulus Set. *PLoS ONE, 8*, e79131. Doi: 10.1371/journal.pone.0079131.

Dawson, G. (2008). Early behavioural intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and Psychopathology*, *20*(03), 775–803. Doi: 10.1017/ S0954579408000370.

Dean, B, M., Ginnell, L., Boardman, J.P., & Fletcher-Watson, S (2021). Social cognition following preterm birth: A systematic review, *Neuroscience & Biobehavioural Reviews,* 124, 151-167. Doi: 10.1016/j.neubiorev.2021.01.006.

*Dean, B., Ginnell, L., Ledsham, V., Tsanas, A., Telford, E., Sparrow, S., Fletcher-Watson, S., & Boardman, J. P. (2021). Eye tracking for longitudinal assessment of social cognition in children born preterm, *The Journal of Child Psychology & Psychiatry,* 62: 4, 470-480. Doi: 10.1111/jcpp.13304.

*De Groote, I., Roeyers, H., & Warreyn, P. (2006). Social-communicative abilities in young high-risk preterm children, *Journal of Developmental & Physical Disabilities,* 18 (2). Doi: 10.1007/s10882-006-9011-y.

Dennis, M., Agostino, A., Taylor, H.G., Bigler, E. D., Rubin, K., Vannatta, K., Gerhardt, C. A., Stancin, T., & Yeates, K. O (2013a). Emotional expression and socially modulated emotive communication in children with traumatic brain injury. *Journal of the*

International Neuropsychological Society, 19(1), 34–43. Doi: 10.1017/S1355617712999884,

Dennis, M., Barnes, M.A., Wilkinson, M., & Humphreys, R.P. (1998). How children with head injury represent real and deceptive emotion in short narratives. *Brain & Language*, 61(3), 450–83. Doi: 10.1006/brln.1997.1886.

Dennis, M., Simic, N., Bigler, E.D., Abildskov, T., Agostino, A., Taylor, G. H., Rubin, K., Vannatta, K., Gerhardt, C. A., Stancin, T., & Yeates, K. O (2013b). Cognitive, affective, and conative theory of mind (ToM) in children with traumatic brain injury, *Developmental Cognitive Neuroscience*, 5, 25–39. Doi: 10.1016/j.dcn.2012.11.006.

Dennis, M., Simic, N., Taylor H. G., Bigler, E.D., Rubin, K., Vannatta, K., Gerhardt, C. A., Stancin, T., Roncadin, C., & Yeates, K. O (2012). Theory of mind in children with traumatic brain injury, *The International Neuropsychological Society,* 18:908–916. Doi: 10.1017/S1355617712000756.

*De Schuymer, L., De Groote, I., Beyers, W., Striano, T., & Roeyers, H. (2011). Preverbal skills as mediators for language outcome in preterm and full-term children, *Early Human Development,* 87, 265-272. Doi: 10.1016/j.earlhumdev.2011.01.029.

*Davis, R., Donati, G., Finnegan, K., Boardman, J. P., Dean, B., Fletcher-Watson, S., Forrester, G. S. (2021). Social gaze in preterm infants may act as an early indicator of atypical lateralization, *Child Development,* 93: 869-880. Doi: 10.1111/cdev.13734.

*Degnan, A. J., Wisnowski, J. L., Choi, S., Ceschin, R., Bhushan, C., Leahy, R. M., Corby, P., Schmithorst, V. J., & Panigraphy, A. (2015). Altered structural and functional connectivity in late preterm preadolescence: An anatomic seed-based resting state networks related to posteromedial and lateral parietal cortex, *PLOS One.* Doi: 10.1371/journal.pone.0130686.

Devine, R.T., Apperly, I.A. (2022). Willing and able? Theory of mind, social motivation, and social competence in middle childhood and early adolescence, *Developmental Science*, 25(1), e13137. Doi: 10.1111/desc.13137.

Downs, A., Strand, P., & Cerna, S. (2007). Emotion understanding in English- and Spanish-speaking preschoolers enrolled in Head Start, *Social Development,* 16, 410–439. Doi: 10.1111/j.1467-9507.00391.

Ekman, P. (1976). *Pictures of Facial Affect*; Psychologists Press: Palo Alto, CA, USA.

*Egotubov, A., Atzaba-Poria, N., Meiri, G., Marks, K., & Gueron-Sela, N. (2020). Neonatal risk, maternal sensitive-responsiveness and infants' joint attention: Moderation by stressful contexts, *Journal of Abnormal Child Psychology,* 48: 453-466. Doi: 10.1007/s10802-019-00598-3.

*Fenoglio (2019). *Social-cognitive development in infants born preterm,* [Doctoral thesis, University of Minnesota].

Fenoglio, A., Georgieff, M. K., & Elison, J. T. (2017). Social brain circuity and social cognition in infants born preterm, *Journal of Neurodevelopmental Disorders,* 9: 27. Doi: 10.1186/s11689-017-9206-9.

Field, A.P., & Gillett, R (2010). How to do a meta-analysis, *British Journal of Mathematical & Statistical Psychology, 63*: 665-94. Doi: 10.1348/000711010X502733.

Fletcher-Watson, S., Findlay, J.M., Leekam, S.R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene, *Perception,* 37, 571–583. Doi: 10.1068/p5705.

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

Fletcher-Watson, S., Leekam, S.R., Benson, V., Frank, M.C., & Findlay, J.M. (2009). Eye-movements reveal attention to social information in autism spectrum disorder, *Neuropsychologia,* 47, 248–257. Doi: 10.1016/j.neuropsychologia.2008.07.016.

Fletcher-Watson, S., McConnell, F., Manola, E., B& McConachie, H. (2014). Interventions based on the thory of mind cognitive model for autism spectrum disorders (ASD), *Cochrane Database of Systematic Reviews,* (3): CD008785. Doi: 10.1002/14651858.CD008785.pub2.

*Ford, R. M., Griffiths, S., Neulinger, K., Andrews, G., Shum, D. H. K., & Gray, P. H. (2017). Impaired prospective memory but intact episodic memory in intellectually average 7- to 9-year-olds born very preterm and/or very low birth weight, *Child Neuropsychology,* 23 (8), 954-979. Doi: 10.1080/09297049.2016.1216091.

*Frie, J., Padilla, N., Aden, U., Lagercrantz, H., & Bartocci, M. (2016). Extremely preterm-born infants demonstrate different facial recognition processes at 6-10 months of corrected age, *The Journal of Pediatrics*

Frith, C. D. (2008). Social cognition, *Philosophical Transactions of the Royal Society,* 363, 2033-2039. Doi: 10.1098/rstb.2008.0005.

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. Doi: 10.1098/rstb.2002.1218.

Frith, C. D., & Frith, U. (2007). Social cognition in humans, *Current Biology,* 17 (16): R724-32. Doi: 10.1016/j.cub.2007.05.068.

Frith, U., Happe, F., & Siddons, F. (1994). Autism and theory of mind in everyday life, *Social Development,* 3 (2): 108-24. Doi: 10.1111/j.1467-9507.1994.tb00031.

*Fu, X., Hung, A., De Silva, A. D., Busch, T., Mattson, W. I., Hoskinson, K. R., Taylor, H. G., & Nelson, E. E. (2022). Development of the metalizing network structures and theory

of mind in extremely preterm youth, *Social Cognitive & Affective Neuroscience,* 17, 977-985. Doi: 10.1093/scan/nsac027.

Fuller, E. A., & Kaiser, A. P. (2019). The effects of early intervention on social communication outcomes for children with autism spectrum disorder: a meta-analysis, *Journal of Autism & Developmental Disorders,* 50: 1683-700. Doi: 10.1007/s10803-019-03927.

Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. (2009). Faces attract infants' attention in complex displays. Infancy, 14, 550–562.

*Gomez, E. F., Vasung, L., Meskaldji, D. E., Lazeyras, F., Borradosi-Tolsa, C., Hagmann, P., Barisnikov, K., Thiran, J. P., & Huppi, P. S. (2015). Structural brain connectivity in school-age preterm infants provides evidence for impaired networks relevant for higher order cognitive skills and social cognition, *Cerebral Cortex,* 25: 2793-2805. Doi: 10.1093/cercor/bhu073.

Goren, C.C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern dis- crimination of face-like stimuli by new-born infants, Paediatrics, 56:544-9. PMID: 1165958.

Greenland, S (1994). Can meta-analysis be salvaged? *American Journal of Epidemiology,* 140: 783-7. Doi: 10.1093/oxfordjournals.aje.a117326.

Grob, A., Meyer, C. S., & Hagmann-von Arx, P. (2009). *Intelligence and development scales.* Bern, Switzerland: Hans Huber.

Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Research*, *41*, 1475–1482. Doi: 10.1016/s0042-6989(00)00317-5.

*Gueron-Sela, N., Atzaba-Poria, N., Meiri, G., & Marks, K. (2015). The caregiving environment and developmental outcomes of preterm infants: Diathesis stress or differential susceptibility effects? *Child Development,* 86 (4), 1014-1030. Doi: 10.1111/cdev.12359.

Guo, B. Q., Li, H. B., Zhai, D. S., & Yang, L. Q. (2022). Prevalence of autism spectrum disorder by birth wright, gestational age, and size for gestational age: a systematic review, meta-analysis, and meta-regression, *European Child & Adolescent Psychiatry.* Doi: 10.1007/s00787-022-02078-4.

*Haley, D. W., GrunAU, R. E., Weinberg, J., Keidar, A., & Oberlander, T. F. (2010). Physiological correlates of memory recall in infancy: vagal tone, cortisol, and imitation in preterm and full-term infants at 6 months, *Infant Behaviour & Development,* 33 (2), 219-234. Doi: 10.1016/j.infbeh.2009.10.007.

Henning, A. T., Hofer, G., & Aschersleben, D. (2012). *Erweiterte 'Theory of Mind'-Skala für 3-bis 11-jährige Kinder,* in [Extended Theory-of-Mind Scale for 3- to 11-year-olds], Peterson, Wellman, and Slaughter. Saarbrücken: Universität des Saarlandes.

Heywood, C. A., Gadotti, A., & Cowey, A. (1992). Cortical area V4 and its role in the perception of color, *Journal of Neuroscience*, *12*, 4056–4065. Doi: 10.1523/JNEUROSCI.12-10-04056.1992.

Higgins J.P.T, Thomas J, Chandler J, Cumpston M, Li T, Page M., & Welch V.A (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook.

Higgins, J.P.T., Thompson, S.G., Deeks, J., & Altman, D.G. (2003). Measuring inconsistency in meta-analysis, *British Medical Journal,* 327:557. Doi: 10.1136/bmj.327.7414.557.

Hirschberger, R. G., Kuban, K. C. K., O'Shea, T. M., Joseph, R. M., Heeren, T., Douglass, L. M., Stafstrom, C. E., Jara, H., Frazier, J. A., Hirtz, D., Rollins, J. V., & Paneth, N. (2018). Co-occurrence and severity of neurodevelopmental burden (cognitive impairment, cerebral pal- sy, autism spectrum disorder, and epilepsy) at age ten years

in children born extremely preterm, *Paediatric Neurology*, 79: 45–52. Doi: 10.1016/j.pediatrneurol.2017.11.002.

*Hood, K. M. (2009). Social competence at age 4 years, of children born very preterm [Doctoral thesis, University of Canterbury].

Howlin, P., Baron-Cohen, S., & Hadwin, J. (1999). Teaching children with autism to mind-read: A practical guide. Chichester: Wiley.

Howson, C. P., Kinney, M.V., McDougall, L., & Lawn, J.E. (2013). Born toon soon: preterm birth matters, *Reproductive Health,* 10 (1): 1–9. www.reproductive-health-journal.com/content/10/S1/S1.

Johnson, S., Hollis, C., Kochhar, P., Hennessy, E., Wolke, D., & Marlow, N. (2010). Autism Spectrum disorders in extremely preterm children, *The Journal of Paediatrics,* 156 (4), 525–531. Doi.org/10.1016/j.jpeds.2009.10.041 e522.

Johnson, S., Hollis, C., Kochhar, P., Hennessy, E., Wolke, D., & Marlow, N (2010). Psychiatric disorders in extremely preterm children: longitudinal finding at age 11 years in the EPICure study, *Journal of Child & Adolescent Psychiatry,* 49: 453–63. Doi:10.1016/j.jaac.2010.02.002. e1.

Johnson, S., & Marlow, N. (2011). Preterm birth and childhood psychiatric disorders. *Paediatric Research.* 2011;69 (5 Pt 2):11R–8R. Doi: 10.1203/PDR.0b013e31821faa0.

Johnson, S., Waheed, G., Manktelow, B. N., Field, D. J., Marlow, N., Draper, E. S., & Boyle, E.M. (2018). Differentiating the preterm phenotype: Distinct profiles of cognitive and behavioural development following late and moderately preterm birth, *The Journal of Pediatrics,* 193, 85-92.e1. Doi: 10.1016/j.jpeds.2017.10.002.

*Jones, K. M., Champion, P. R., & Woodward, L. J. (2013). Social competence of preschool children born very preterm, *Early Human Development,* 89, 795-802. Doi: 10.1016/j.earlhumdev.2013.06.008.

Kanel, D., Vanes, L.D., Pecheva, D., Hadaya, L., Falconer, S., Counsell, S.J., Edwards, D.A., & Nosarti, C. (2021). Neonatal White Matter Microstructure and Emotional Development during the Preschool Years in Children Who Were Born Very Preterm. *Eneuro*, 8 (5). Doi: 10.1523/ENEURO.0546-20.2021.

Klin A (2000). Attributing social meaning to ambiguous visual stimuli in higher-functioning Autism and Asperger syndrome: the social attribution task, *Journal of Child Psychology & Psychiatry*, 41: 831–46. Doi: 10.1111/1469-7610.00671.

Kohls, G., Chevallier, C., Troiani, V., and Schultz, R. T. (2012). Social wanting dysfunction in autism: neurobiological underpinnings and treatment implications. *Journal of Neurodevelopmental. Disorders,* 4:1. www.jneurodevdisorders.com/content/4/1/10.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-second edition*. Harcourt Assessment.

*Imafuku, M., Kawai, M., Niwa, F., Shinya, Y., & Myowa, M. (2021). Longitudinal assessment of social attention in preterm and term infants: Its relation to social communication and language outcome, *Infancy,* 26, 617-634. Doi: 10.1111/infa.12402.

*Imafuku, M., Kawai, M., Niwa, F., Shinya, Y., Iangawa, M., & Myowa-Yamakoshi, M. (2017). Preference for dynamic human images and gaze-following abilities in preterm infants at 6 and 12 months of age: An eye tracking study, *Infancy,* 22 (2), 223-239. Doi: 10.1111/infa.12144.

*Landry, S. H. (1986). Preterm infants' responses in early joint attention interactions, *Infant Behaviour & Development,* 9, 1-14. Doi: 10.1016/0163-6383(86)90034-2.

Laverty. C., Surtees, A., O' Sullivan, R., Sutherland, D., Jones, C., & Richards, C. (2021). The prevalence and profile of autism in individuals born preterm: a systematic review and meta-analysis, *Journal of Neurodevelopmental Disorders,* 13: 41. Doi: 10.1186/s11689-021-09382-1.

Lee, Y. H. (2019). Strengths and limitations of meta-analysis, *Korean Journal of Medicine,* 94 (5),391-395. Doi: 10.3904/kjm.2019.94.5.391.

*Lejeune, F., Reveillion, M., Monnier, M., Huppi, P. S., Tolsa, C. B., & Barisnikov, K. (2016). Social reasoning abilities in preterm and full-term children aged 5-7 years, *Early Human Development,* 103, 49-54. Doi: 10.1016/j.earlhumdev.2016.07.010.

Levels and trends in child mortality: Report 2017, Estimates developed by the UN inter-agency group for child mortality estimation. New York: United Nations Children's Fund; 2017.

Liberate, A., Altman, D. G., Tetlaff, J., Mulrow, C., Gotzsche, P. C., Loannidis, J.P.A., Clarke, M., Devereaux, P. J., Kleijnen, J., Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care intervention: explanation and elaboration, *PLOS Medicine,* 21; 6 (7): e1000100. Doi: 10.1371/journal.pmed.1000100.

Liddle, B., & Nettle, D. (2006). Higher-order theory of mind and social competence in school-age children, *Journal of Cultural and Evolutionary Psychology,* 4(3), 231–244. Doi:10.1556/ JCEP.4.2006.3–4.3.

*Linares, L. O. (1986). Early mental representation and the emergence of language in preterm and full-term infants. [Doctoral thesis, Yeshiva University].

*Longa, L. D., Nosarti, C., & Farroni, T. (2022). Emotion recognition in preterm and full-term school-age children, *International Journal of Environmental Research & Public Health,* 19, 6507. Doi: 10.3390/ijerph19116507.

Luu, J., Jellett, R., Yaari, M., Gilbert, M., & Barbaro, J. (2020). A comparison of children born preterm and full-term on the autism spectrum in a prospective community sample, *Frontiers in Neurology,* 11: 597505. Doi: 10.3389/fneur.2020.597505.

Magill-Evans, J., Koning, C., Cameron-Sadava, A., & Manyk, K. (1995). The child and adolescent social perception measure, *Journal of Nonverbal Behaviour,* 19, 151–169. Doi: 10.1007/BF02175502.

*Marleau, I., Vona, M., Gagner, C., Luu, T. M., Beauchamp, M. H. (2020). Social cognition, adaptive functioning, and behaviour problems in pre-schoolers born extremely preterm, *Child Neuropsychology.* Doi: 10.1080/09297049.2020.1797656.

*Marotta, A., Rea, M., Ferri, R., & Casagrande, M. (2020). Investigating socio-emotional cognition in late preterm children: A case control study, *European Journal of Developmental Psychology,* 17: 3, 365-378. Doi: 10.1080/17405629.2019.1637339.

Masten, A. S & Cicchetti, D. (2010). Developmental cascades: part 1, *Development & Psychopathology,* 22 (3), 491-495. Doi: 10.1017/S0954579410000222.

Mateus, V., Osorio, A., & Martins, C. (2019). Effects of prematurity on infants' joint attention abilities: A meta-analytic study, *Infant & Child Development,* 28 e2145. Doi: 10.1002/icd.2145.

*Mateus, V., Vieira, E. D. A. P., Martins, C., Pachi, P. R., & Osorio, A. (2020). Joint attention abilities in Brazilian preterm and full-term infants, *Infant Behaviour & Development,* 60, 101451. Doi: 10.1016/j.infbeh.2020.101451.

Mazur, E., & Ritz, E. (1984). Patterns of gestural, vocal and verbal imitation performance in infancy, *Merrill- Palmer Quarterly,* 30 (4) 369-392.

McCall, R. (1974). Exploratory manipulation and play in the human infant, *Monographs of the Society for Research in Child Development,* 39 (2): 1-88. PMID: 4444723.

McCall, R., Parke, R., & Kavanaugh, K. (1977). Imitation of live and televised models by children one to three years of age, *Monographs of the Society for Research in Child Development,* 42 (5): 1-94. PMID: 615290.

Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience, *Philosophical Transactions of the Royal Society B Journals,* 29; 358 (1431): 491-500. Doi: 10.1098/rstb.2002.1261.

Meyer, C. S., Hagmann-von Arx, P., & Grob, A. (2009). Die Intelligence and Development Scale Sozial-Emotionale Kompetenz (IDS-SEK) Psychometrische Eigenschaften eines Tests zur Erfassung sozial-emotionaler Fähigkeiten. [The intelligence and development scale of social-emotional competence (IDS-SEK): Psychometric properties of a test for assessing social-emotional abilities]. *Diagnostica,* 55(4), 234–244. Doi: 10.1026/0012-1924.55.4.234.

Montagna, A., & Nosarti, C. (2016). Socio-emotional development following very preterm birth: Pathways to pathology, *Frontiers in Psychology,* 7: 80. Doi: 10.3389/fpsyg.2016.00080.

*Mossad, S., Muscat, C., Pang, W. W., & Taylor, M. (2020). Emerging atypical connectivity networks for processing angry and fearful faces in very preterm born children, *Human Brain Mapping,* 41: 3794-3806. Doi: 10.1002/hbm.25088.

*Mossad, S. I., Smith, M. L., Pang, E. W., & Taylor, M. J. (2017). Neural correlates of "theory of mind" in very preterm born children, *Human Brain Mapping,* 38: 5577-5589. Doi: 10.1002/hbm.23750.

*Mossad, S. I., Vandewouw, M. M., Smith, M. L., & Taylor, M. J. (2021). The preterm social brain: altered functional networks for theory of mind in very preterm children, *Brain Communications,* 1, 1-13. Doi: 10.1093/braincomms/fcaa237.

*Mossad, S. I., Young, J. M., Wong, S. M., Dunkley, B. T., Hunt, B. A. E., Pang, E. W., & Taylor, M. J. Taylor. (2021). Very preterm brain at rest: Longitudinal social-cognitive

network connectivity during childhood, *Social Cognitive & Affective Neuroscience,* 00, 1-10. Doi: 10.1093/scan/nsab110.

Moser, K., Macfarlane, A., Chow, Y. H., Hilder, L., & Dattani, N. (2007). Introducing new data on gestation-specific infant mortality among babies born in 2005 in England and Wales, *Health Statistics Quarterly,* (35):13–27. PMID: 17894197.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy, *Child Development,* 78 (3), 938-954. Doi: 10.1111/j.1467-8624.2007.01042.x.

Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., & Seibert, J (2003). A manual of the abridged Early Social Communication Scales (ESCS). Available through the University of Miami Psychology Department, Coral Gables, Florida.

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition, *Current Directions in Psychological Science,* 16 (5): 269-274. Doi: 10.1111/j.1467-8721.2007.00518.

Nakano, T., Tanaka, K., Endo, Y., Yamane, Y., Yamamoto, T., Nakano Y., Ohta, H., Kato, N., & Kitazawa, S. (2010). Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour, *International Journal of Biological Sciences,* 277 (1696): 2935–43. Doi: 10.1098/rspb.2010.0587.

National Institute of Health Care and Excellence. Developmental follow-up of children and young people born preterm (NICE Guideline NG72). 2017 Online at: www.nice.org.uk/guidance/ng72.

Nelson, P. B., Adamson, L. B., & Bakeman, R. (2008). Toddlers' joint engagement experience facilitates pre-schooler's acquisition of theory of mind, *Developmental Science*, 11 (6): 847–852. Doi: 10.1111/j.1467-7687.2008.00733.

Newsome, W. T., & Pare, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT), *Journal of Neuroscience*, 8 (6): 2201–2211. Doi: 10.1523/JNEUROSCI.08-06-02201.1988.

Nosarti, C., Reichenberg, A., Murray, R. M., Cnattingius, S., Lambe, M. P., Yin, L., MacCabe, J., Rifkin, L., & Hultman, C. M. (2012). Preterm birth and psychiatric disorders in young adult life, *Archives of General Psychiatry,* 69 (6): E1–8. Doi: 10.1001/archgenpsychiatry.2011.1374.

*Olafsen, K. S., Ronning, J. A., Handegard, B. H.M Ulvund, A. E., Dahl, L. B., & Kaaresen, P. I. (2012). Regulatory competence and social communication in term and preterm infants at 12 months corrected age. Results from a randomised controlled trial. *Infant Behaviour & Development,* 35, 140-149. Doi: 10.1016/j.infbeh.2011.08.001.

*Olafsen, K. S., Ronning, J. A., Kaaresen, P. I., Ulvund, S. E., Handegard, B. H., & Dahl, L. B. (2006). Jooint attention in term and preterm infants at 12 months corrected age: The significance of gender and intervention based on a randomised controlled trial, *Infant Behaviour & Development,* 29, 554-563. Doi: 10.1016/j.infbeh.2006.07.004.

Parker, J., Rubin, K. H., Erath, S., Wojslawowicz, J. C., & Buskirk, A. A. (2006). Peer relationships and developmental psychopathology. In D. Cicchetti & D. Cohen (Eds.), *Developmental psychopathology: Risk, disorder, and adaptation* (Vol. 2, 2nd ed., pp. 419–493). New York: Wiley.

Pavlova, M (2011). Biological motion processing as a hallmark of social cognition, *Cerebral Cortex,* 22 (5): 981-95. Doi: 10.1093/cercor/bhr156.

*Pena, M., Arias, D., & Dehaene-Lambertz, G. (2014). Gaze following is accelerated in healthy preterm infants, *Psychological Science,* 25 (10), 1884-1892. Doi: 147.188.128.74.

Perner, J., & Wimmer, H. (1985). “John thinks that Mary thinks that. . .” attribution of second-order beliefs by 5- to 10-year-old children, *Journal of Experimental Child Psychology,* 39 (3), 437–471. Doi:10.1016/0022-0965(85)90051-7.

Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The mind behind the message: advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome, *Child Health & Development,* 83, 469–485. Doi: 10.1111/j.1467-8624.2011.01728.

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

*Pereira, S. A., Junior, A. P., Da Costa, M. F., De V Monteiro, M., De Almedia, V. A., Da Fonseca Filho, G. G., Arrais, N., & Simion, F. (2017). A comparison between preterm and full-term infants’ preference for faces, *Jornal De Pediatria,* 93 (1): 35-39. Doi: 10.1016/j.jped.2016.04.009.

*Perez-Roche, T., Altemir, I., Gimenez, G., Prieto, E., Gonzalez, I., Lopez Pison, J., & Pueyo, V. (2017). Face recognition impairment in small for gestational age and preterm children, *Research in Developmental Disabilities,* 62, 166-173. Doi: 10.1016/j.ridd.2017.01.016.

Rao, P. A., Beidel, D. C., & Murray, M. J. (2008). Social skills interventions for children with Asperger’s syndrome or high-functioning autism: a review and recommendations, Journal *of Autism & Developmental Disorders,* 38, 353–361. Doi: 10.1007/s10803-007-0402-4.

Ream, M. A., & Lehwald, L. (2018). Neurologic consequences of preterm birth, *Current Neurology & Neuroscience Reports,* 18: 48. Doi: 10.1007/s11910-018-0862-2.

Reid, C., Davis, H., Horlin, C., Anderson, M., Baughman, N., & Campbell, C. (2013). The Kids’ empathic development scale (KEDS): A multi-dimensional measure of empathy

in primary school aged children, *British Journal of Developmental Psychology,* 31(2), 231–256. Doi:10.1111/bjdp.12002.

Reynolds, C. R., & Bigler, E. D. (1994). Test of memory and learning. Austin, TX: PRO-ED.

Ritchie, K., Bora, S., & Woodward, L. J. (2015). Social development of children born very preterm: a systematic review. *Developmental Medicine & Child Neurology,* 57, 899–918. Doi: 10.1111/dmcn.12783.

Robinson, K.E., Fountain-Zaragoza, S., Dennis, M., Taylor, G. H., Bigler, E. D., Rubin, K., Vannatta, K., Gerhardt, C. A., Stancin, T., Yeates, K. O. (2014). Executive functions and theory of mind as predictors of social adjustment in childhood traumatic brain injury, *Journal of Neurotrauma,* 31(22), 1835–42. Doi: 10.1089/neu.2014.3422.

*Roldan-Tapia, M. D., Moreno-Rios, S., & Canovas-Lopez, R. (2017). Thinking about social and non social alternative possibilities in premature pre-schoolers, *Journal of Clinical & Experimental Neuropsychology,* 39: 8, 725-737. Doi: 10.1080/13803396.2016.1257703.

Rose-Krasnor, L. (1997). The nature of social competence: A theoretical review, *Social Development,* 6, 111-135. Doi: 10.1111/j.1467-9507.1997.tb00097.x

Rosenthal, R (1979). The file drawer problem and tolerance for null results, *Psychological Bulletin,* 86: 638-41. Doi: 10.1037/0033-2909.86.3.638.

*Rowell, L. N. (2015). Investigating joint attention behaviours and episodes in 18-month toddlers born very low birth weight compared to toddlers born normal birth weight. [Doctoral thesis, University of New Mexico]. https://digitalrepository.unm.edu/psy_etds/123.

Rubin, K. H., Bukowski, W., & Parker, J. (2006). Peer interactions, relationships, and groups. In N. Eisenberg (Ed), *Handbook of child psychology: Social, emotional, and personality development* (6th ed., pp. 571–645) New York: Wiley.

Rubin, K. H., Booth, C., Krasnor, L. R., & Mills, R. S. L. (1995). Social relationships and social skills: A conceptual and empirical analysis. In S. Shulman (Ed.), *Close relationships and socioemotional development* (pp. 63–95). Norwood, NJ: Ablex Publishing.

Rubin, K. H., & Krasnor, L. R. (1986). Social-cognitive and social behavioural perspectives on problem solving. In M. Perlmutter (Ed.), *Cognitive perspectives on children's social and behavioural development: The Minnesota Symposia on Child Psychology* (Vol. 18, pp. 1–68). Hills- dale, NJ: Erlbaum.

Rubin, K. H., & Rose-Krasnor, L. (1992). Interpersonal problem-solving. In V. B. Van Hassett & M. Hersen (Eds.), *Handbook of social development* (pp. 283–323). New York: Plenum.

Ryan, N.P., Anderson, V.A., Bigler, E.D., Dennis, M., Taylor, H. G., Rubin, K. H., Vannatta, K., Gerhardt, C. A., Stancin, T., Beauchamp, M. H., Hearps, S., Catroppa, C., Yeates, K. O. (2021). Delineating the nature and correlates of social dysfunction after childhood traumatic brain injury using common data elements: evidence from an international multi-cohort study, *Journal of Neurotrauma,* 38(2), 252–60. Doi: 10.1089/neu.2020.7057.

*Sato, J., Safar, K., Vandewouw, M. M., Bando, N., O'Connor, D. L., Unger, S. L., & Taylor, M. J. (2021). Altered functional connectivity during face processing in children born with very low birth weight, *Social Cognitive & Affective Neuroscience,* 1182-1190. Doi: 10.1093/scan/nsab070.

Saunders, D. (2011). Components of biological motion perception (Unpublished dissertation). Queen's University, Kingston, ON.

Serenius, F., Ewald, U., Farooqi, A., Fellman, V., Hafström, M., Hellgren, K., Marsal, K., Ohlin, A., Olhager, E., Stjernqvist, K., Stromberg, B., Aden, U., & Kallen, K. (2016).

Neurodevelopmental outcomes among extremely preterm infants 6.5 years after active perinatal care in Sweden. *JAMA Pediatics,*170 (10): 954–63. Doi: 10.1001/jamapediatrics.2016.1210.

*Sekigawa-Hosozawa, M., Tanaka, K., Shimizu, T., Nakano, T., Kitazawa, S. (2017). A group of very preterm children characterized by atypical gaze patterns, *Brain & Development,* 39, 218-224. Doi: 10.1016/j.braindev.2016.10.001.

Singh, G. K., Kenney, M. K., Ghandour, R. M., Kogan, M. D., & Lu, M. C. (2013). Mental health outcomes in US children and adolescents born prematurely or with low birthweight. *Depression Research & Treatment,* 2013: 570743. Doi: 10.1155/2013/570743.

Slaughter, V., Imuta, K., Peterson, C.C., & Henry, J.D. (2015). Meta- analysis of theory of mind and peer popularity in the preschool and early school years, *Child Development,* 86(4), 1159–74. Doi: 10.1111/cdev.12372.

Soto-Icaza, P., Aboitiz, F., & Billeke, P. (2015). Development of social skills in children: neural and behavioural evidence for the elaboration of cognitive models, *Frontiers in Neuroscience,* 9 (333). Doi: 10.3389/fnins.2015.00333.

Spittle, A. J., Treyvaud, K., Doyle, L. W., Roberts, G., Lee, K. J, Inder, T. E., Cheong, J. L. Y., Hunt, R. 1., Newnham, C. A., & Anderson, P. J. (2009). Early emergence of behavior and social-emotional problems in very preterm infants, *Journal of the American Academy of Child & Adolescent Psychiatry,* 48: 909–18. Doi:10.1097/ CHI.0b013e3181af8235.

*Speretto, R. G. (2015). The influence of gestational age on social attention and language in the second year of life. [Doctoral thesis, Cardiff University].

Sterne JAC, Gavaghan D, Egger M (2000) Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *Journal of Clinical*

Epidemiology 53: 1119–29. Sterne JAC, Gavaghan D, Egger M (2000) Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology* 53: 1119–29. Doi: 10.1016/s0895-4356(00)00242-0.

Striano, T., & Rochat, P. (1999). Developmental link between dyadic and triadic social competence in infancy, *British Journal of Developmental Psychology,* 17, 551–62. Doi: 10.1348/026151099165474.

Striano, T., & Stahl, D (2005). Sensitivity to triadic attention in early infancy, *Developmental Science,* 8: 333–43. Doi: 10.1111/j.1467-7687.2005.00421.

Taylor, G. H. (2022). Neurodevelopmental origins of social competence in very preterm children, *Seminars in Fetal & Neonatal Medicine,* 25. 101108. Doi: 10.1016/j.siny.2020.101108.

*Taylor, N. M.M Jakobson, L. S., Maurer, D., & Lewis, T. L. (2009). Differential vulnerability of global motion, global form, and biological motion processing in full-term and preterm children, *Neuropsychologia,* 47, 2766-2778. Doi: 10.1016/j.neuropsychologia.2009.06.001.

Treyvaud, K,, Ure, A.,, Doyle, L.W., Lee, K. J., Rogers, C. E., Kidokoro, H., Inder, E. T., & Anderson, P. J. (2013). Psychiatric outcomes at age seven for very preterm children: rates and predictors, *Journal of Child Psychology & Psychiatry,* 54: 772–9. Doi:10.1111/jcpp.12040.

*Telford, E. J. (2017). The effect of preterm birth on white matter tracts and infant cognition. [Doctoral thesis, University of Edinburgh].

Telford, E. J., Cox, S., Fletcher-Watson, S., Anblagen, D., Sparrow, S., Pataky, R., & Boardman, J. P. (2017). A latent measure explains substantial variance in white matter

microstructure across the new-born human brain, *Brain Structure and Function,* 222 (9), 4023–4033. Doi: 10.1007/s00429-017-1455-6.

*Telford, E. J., Fletcher-Watson, S., Gillespie-Smith, K., Pataky, R., Sparrow, S., Murry, E. C., O'Hare, A., & Boardman, J. (2016). Preterm birth is associated with atypical social orienting in infancy detected using eye tracking, *The Journal of Child Psychology & Psychiatriy,* 57: 7, 861-868. Doi: 10.1111/jcpp.12546.

Tordorova, G. K., Mcbean Hatton, R. E., & Pollick, F.E. (2019). Biological motion perception in autism spectrum disorder: a meta-analysis, *Molecular Autism,* 10: 49. Doi: 10.1186/s13229-019-0299-8.

Turner, R. M., Bird, A.M., & Higgins, J.P.T. (2013). The impact of study size on meta-analyses: Examination of underpowered studies in Cochrane reviews, *PLOS One,* 8 (3): e59202. Doi: 10.1371/journal.pone.0059202.

*Twilhaar, E. S., De Kieviet, J. F., Bergwerff, C. E., Finken, M. J. J., Van Elburg, R. M., & Oosterlaan, J. (2019), *The Journal of Pediatrics,* 213: 66-73. Doi: 10.1016/jpeds.2019.06.045.

Uljarevic, M., Frazier, T. W., Jo, B., Phillips, J. M., Billingham, W., Cooper, M. N., & Hardan, A. Y. (2021). Relationship between social motivation in children with autisim spectrum disorder and their parents, *Frontiers in Neuroscience,* 15: 660330. Doi: 10.3389/fnins.2021.660330.

Vogel, J.P., Chawanpaiboon, S., Moller, A.B., Watananirun, K., Bonet, M., & Lumbiganon, P. (2018). The global epidemiology of preterm birth, *Best Practice & Research Clinical Obstetrics & Gynaecology,* 52, 3-12. Doi: 10.1016/j.bpobgyn.2018.04.003.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind task. *Child Development*, *75*(2), 523–541. Doi: 10.111/j.1467-8624.2004.00691.

*Williamson, K. E. (2014). Social cognition and social outcomes in children born at very low birth weight [Doctoral thesis, University of Manitoba].

*Williamson, K. E., & Jakobson, L. S. (2014)a. Social attribution skills of children born preterm at very low birthweight, *Development & Psychopathology,* 26, 889-900. Doi: 10.1017/S0954579414000522.

*Williamson, K. E., & Jakobson, L. S. (2014)b. Social perception in children born at very lowe birthweight and its relationship with social/behavioral outcomes, *The Journal of Child Psychology & Psychiatry,* 55: 9, 990-998. Doi: 10.1111/jcpp.12210.

*Williamson, K. E., Jakobson, L. S., Saunders, D. R., Troje, N. F. (2015). Local and global aspects of biological motion perception in children born at very low birth weight, *Child Neuropsychology,* 21: 5, 603-628. Doi: 10.1080/09297049.2014.945407.

*Witt, A., Theurel, A., Tolsa, C. B., Lejeune, F., Fernandes, L., De Jonge, L. V. H., Monnier, M., Graz, M. B., Barisnikov, K., Gentaz, E., & Huppi, P. S. (2014). Emotional and effortful control abilities in 42-month-old very preterm and full-term children, *Early Human Development,* 90, 565-569. Doi: 10.1016/j.earlhumdev.2014.07.008.

*Witt, S., Weitkamper, A., Neumann, H., Lucke, T., & Zmyj, N. (2018). Delayed theory of mind development in children born preterm: A longitudinal study, *Early Human Development,* 127, 85-89. Doi: 10.1016/j.earlhumdev.2018.10.005.

*Wocadlo, C., & Rieger, I. (2006). Social skills and nonverbal decoding of emotions in very preterm children at early school age, *European Journal of Developmental Psychology,* 3: 1, 48-70. Doi: 10.1080/17405620500361894.

*Wolfe, K. R., Vannatta, K., Nelin, M. A., Yeates, K. O. (2015). Executive functions, social information processing, and social adjustment in young children born with very low birth weight, *Child Neuropsychology,* 21 (1), 41-54. Doi: 10.1080/09297049.2013.866217.

Woolfe, T., Want, S. C., & Siegal, M (2002). Signposts to development: theory of mind in deaf children, *Child Development,* 73(3):768–78. Doi: 10.1111/1467-8624.00437.

*Yamamoto, M., Konishi, Y., Kato, I., Koyano, K., Nakamura, S., Nishida, T., & Kusaka, T. (2021). Do low birth weight infants not see eyes? Face recognition in infancy, *Brain & Development,* 43, 186-191. Doi: 10.1016/j.braindev.2020.09.002.

*Yamaoka, N., & Takada, S. (2016). Joint attention development in low-risk very low birth weight infants at around 18 months of age, *Kobe Journal of Medical Sciences,* 62 (4). E89-E98. Doi: PMC5432063.

*Ye, A. X., AuCoin-Power, M., Taylor, M. J., & Doesburg, A. M. (2015). Disconnected neuromagnetic networks in children born very preterm, *NeuroImage: Clinical,* 9, 376-384. Doi: 10.1016/j.nicl2015.08.016.

Yeates, K. O. (2000). Closed-head injury. In K. O. Yeates, M. D. Ris, & H. G. Taylor (Eds.), *Paediatric neuropsychology: Research, theory, and practice* (pp. 92–116). New York: Guilford.

Yeates, K. O., Bigler, E. D., Dennis, M., Gerhardt, C. A., Rubin, K. H., Stancin, T., Taylor, G. H., & Vannatta, K. (2007). Social outcomes in childhood brain disorder: A heuristic integration of social neuroscience and developmental psychology, *Psychological Bulletin,* 133 (3), 535-556. Doi: 10.1037/0033-2909.133.3.535.

*Zivan, M., Morag, I., Yarmolovsky, J., & Geva, R. (2021). Hyper-reactivity to salience limits social interaction among infants born pre-term and infant siblings of children with ASD, *Frontiers in Psychiatry,* 12, 646838. Doi: 10.3389/fpsyt.2021.646838.

Zmyj, N., Witt, S., Weitkamper, A., Neumann, H., & Lucke, T. (2017). Social cognition is children born preterm: A perspective on future research directions, *Frontiers in Psychology,* 8:455. Doi: 10.3389/fpsyg.2017.00455.

End of Chapter One

CHAPTER TWO:

A PILOT AND FEASIBILITY STUDY EXPLORING ASSOCIATIONS BETWEEN IMPLICIT AND EXPLICIT MEASURES OF SOCIAL COGNITION *AND* DIFFERENCES BETWEEN CHILDREN BORN MODERATE-TO-LATE PRETERM AND THEIR FULL-TERM PEERS.

Abstract

OBJECTIVES: Part one of this pilot and feasibility study investigated associations between implicit and explicit measures of social cognition and part two investigated differences in social cognition between moderate-to-late preterm and full-term children aged 18 to 36 months. The aim of part one and part two was to estimate the effect size of the correlation between measures and the magnitude of differences between groups respectively.

METHOD: Implicit social cognition was measured using eye tracking technology using two well-validated preferential looking paradigms. Explicit social cognition was measured using the Early Social Cognition Scale, a battery of well-validated tasks measuring skills considered important precursors to later developing theory of mind abilities. Children also completed the Mullen Scales of Early Learning as broad measure of cognitive ability.

RESULTS: Results from this pilot study suggest preliminary evidence of large correlations between implicit and explicit social cognition are unlikely when measured cross-sectionally *and* important differences may exist in social preference between moderate-to-late preterm children and full-term children.

CONCLUSIONS**:** Strong correlations between implicit and explicit measures of social cognition may be unlikely, therefore, given the large, estimated sample size required for a fully powered study, further studies exploring associations between implicit and explicit measures *cross-sectionally* may not be feasible/ useful. Exploring differences in social preference between moderate-to-late preterm and full-term children may be beneficial, given that estimated sample sizes for a fully powered study are likely feasible.

Introduction

The capacity to preferentially attend to social aspects of our world (referred to as social preference from here onwards) plays a key role in social development (Pavlova, 2012; Sifre et al., 2018). The social motivation hypothesis (Chevalier et al., 2012; Dawson, 2008; Kohls et al., 2012) suggests that reduced preference to social content may place a child on an alternative developmental pathway (Fletcher-Watson & Happe, 2019) and be an early marker for divergent social cognition (Dean et al., 2021; Gogate, 2020; Imafuku et al., 2019). This has been demonstrated in autistic children, in whom reduced social preference is a key feature (Wen et al., 2021). Therefore, a reduced motivation towards social stimuli may be a key mechanism which explains the social and communication characteristics associated with an autism diagnosis (Bradshaw et al., 2022; Uljarevic et al., 2021) and other divergent developmental outcomes. Bradshaw et al. (2022) suggest these early, more subtle differences in social attention may have a cascading effect on social development and so, social preference and how it is associated with earlier social cognitive skills may be important to investigate further to understand the mechanisms underpinning divergent neurodevelopmental outcomes.

Eye tracking instruments are an objective way of exploring what an individual selectively pays attention to (Deubel & Schneider, 1996; Ernst & Paulus, 2005; Wolf & Ueda, 2021) and many social attention paradigms are available. Such methodologies assume that longer gaze duration to social stimuli is indicative of increased interest/positive value over non-social stimuli (Dubey et al., 2022; Myowa-Yamakoshi et al., 2003 Shimojo et al., 2003; Shimojo et al., 2010). Advancements in eye tracking technology have increased the feasibility of exploring gaze behaviours with younger children (Hessles & Hooge, 2019) and allow for research of attentional preferences without the need for language or more advanced development required in many direct observational methods (Eckstein et al., 2017).

Whilst direct observational methods of social cognition are reliable when conducted by trained individuals (Hintze & Matthews, 2004), such methodologies are time consuming, resource intensive and can be confounded by many factors such as experimenter bias (Dubey et al., 2022). As such, the analysis of gaze behaviours can be complimentary to these approaches and can offer important information related to social development without reliance on verbal and intellectual development and experimenter skill.

Implicit and Explicit Social Cognition

Quantifying social preference alongside other behaviours considered important in social cognitive development can offer insights about how nuanced and *earlier* attentional preferences might be associated with development. Research investigating associations between implicit and explicit measurement of theory of mind (ToM from here onwards) has suggested performance between implicit and explicit tasks is strongly correlated (Oktay-Gur et al., 2018). As such, children (aged 3 to 6 years) who perform well in implicit ToM tasks perform equally as well in explicit ToM tasks. However, our understanding of how *social preference* is associated with performance on measures of *early* social cognitive skills (such as precursors to ToM) is limited. Studies exploring this relationship have found reduced social preference to be associated with reduced joint attention behaviours (Bacon et al., 2020; Schietecatte et al., 2012;), reduced social interaction behaviours (Rijn et al., 2018) and increased difficulties in social skills (Pierce et al., 2016). There are inconsistences in these findings however, with correlations only apparent across some social cognitive skills and not others (Schietecatte et al., 2012). Most studies of these *earlier* social cognitive skills have focussed on atypical samples (such as autistic children), without a typically developing comparison group (Bacon et al., 2020; Schietecatte et al., 2012). Further, eye tracking methodologies measuring social preference are mostly unvalidated and social outcomes are often measured via non-experimental methods such as parent-report (Dubey et al., 2022; Falck-

ytter et al., 2010; Jones et al., 2008; Klin et al., 2007; Rijn et al., 2018; Speer et al., 2007). Investigating this further with well validated measures is important to better understand how these methods of measurement are associated and possible mechanisms underlying unique developmental pathways.

Prematurity

Children born prematurely are more likely to experience divergent social development (Dean et al., 2021; Montagna & Nosarti, 2106; Zymj et al., 2017). Preterm birth, defined by the World Health Organisation (WHO, 1977) as birth occurring before 37 weeks gestational age (GA from here onwards) affects approximately 11% of births worldwide (Blencowe et al., 2012; Vogel et al., 2018) and is increasing in prevalence in higher income countries (Blencowe et al., 2012). Research has indicted a range of outcomes across development in comparison to children born full-term (FT from here onwards) which includes differences in social development (Johnson & Marlow, 2017). Of relevance here are increased rates of autism (Agarwal et al., 2018; Fenoglio et al., 2018; Guo et al., 2022; Johnson et al., 2010; Laverty et al., 2021; Singh et al., 2013) and attention deficit hyperactivity disorder – ADHD (Bhutta et al., 2002; Johnson et al, 2010; Norsarti et al., 2012; Treyvaud et al., 2013) which both include divergent social and communication features in diagnostic criteria. Unsurprising then, are findings of differences related to a broad range of social cognitive skills important for social development in preterm children, compared to FT peers (Arpi & Ferrari, 2013; Ritchie et al., 2015; Zimj et al., 2017). Therefore, a better understanding of social preference and earlier social cognitive skills in children born prematurely is of particular clinical importance.

Moderate to Late Prematurity

Most preterm births occur moderate to late (32-37 weeks GA), representing 75% of births before 37 weeks' GA (Boyle et al., 2015; Moser et al., 2007). Despite this, research investigating outcomes of children born prematurely has tended to focus on earlier GAs.

Research with those born moderately to late preterm (MLP from here onwards) suggests adverse developmental patterns *comparable* with those born at earlier GAs but at lower rates (Hornman et al., 2017), in addition to a distinct behavioural phenotype compared to FT peers. Such outcomes include poorer cognitive functioning (Gamarra-Oca et al., 2022; Morgan & Boyle., 2017), impaired language development (Johnson et al., 2018), increased general health problems (Boyle et al., 2012) and increased fine motor difficulties (Kerstiens et al., 2010). Furthermore, increased rates of neurodivergence, such as ADHD (De Jong et al., 2012) and an increase in sensory impairments and cerebral palsy (Tammela, 2022) are also reported to extend to this sub-category of prematurity. More recently, MLP children have been found to *also* be at increased risk for autism diagnosis when using diagnostic tools and clinical assessments compared to FT peers (Guo et al., 2022), the rate of which was found to be near comparable risk to those born <32weeks GA. Whilst these data are important to understand the risk profile of MLP children, diagnostic assessments in prevalence estimates are limited as they cannot identify sub-threshold differences in areas associated with autism such as social development, which may still result in divergent outcomes or clinical needs (Laverty et al., 2021). Therefore, as differences in social development are a key feature of autism and ADHD *and* data suggests increased prevalence of such neurodivergence in MLP children, understanding early social development in this group is important.

Social Development and Moderate to Late Prematurity

The systematic review and meta-analysis described in Chapter One of this thesis included studies employing direct observational methods across all GAs. When pooled for meta-analysis, preterm children performed more poorly in comparison to FT children in early social attention, affect recognition and ToM (see pages 47-57 of Chapter One). Inconsistent findings were revealed in studies measuring different aspects of joint attention. Additionally other social cognitive skills, such as social reasoning and empathy, were not able to be pooled

for analysis. There were two crucial findings from this systematic review and meta-analysis. The measurement and reporting of social cognitive constructs used across studies was inconsistent *and* studies failed to investigate the MLP cohort in isolation from earlier gestations across multiple social cognitive domains. Therefore, results of these studies may be confounded with the medical heterogeneity evident in earlier GAs (Meldrum et al., 2013) and there is a clear scientific and clinical need to undertake robust assessments of well-defined early social abilities in MLP cohorts.

Of the few studies of social cognition with a MLP focus, joint attention skills have been shown to be reduced when compared to FT children (Fenoglio, 2019; Gueron-Sela et al., 2015; Speretto 2015). The remaining studies focus on *later* social cognitive skills such as affect recognition (Longa et al., 2022; Marotta et al., 2019), with MLP demonstrating a reduced capacity as compared to FT children. Marotta et al., (2019) found the same MLP cohort to perform similarly on later skills, such as social understanding (Marotta et al., 2019). There are no studies exploring social attention, ToM, or earlier precursors to ToM outside of joint attention with MLP children. Thus, our knowledge of differences in social cognition in MLP children compared to FT children is largely unknown and this is therefore an area warranting further research.

Problems with Current Research

Understanding the limitations of current research exploring experimentally measured social development with preterm populations is important to guide future studies with MLP children. Whilst eye tracking has been suggested as an excellent method to explore social preference in children born preterm (Fletcher-Watson & Hampton, 2018), many of the limitations predominantly lie in chosen paradigms. Studies using static paradigms (Davis et al., 2022; Dean et al., 2021; Telford et al., 2016; Zivan et al., 2021) may fail to represent real-life social stimuli thereby limiting conclusions, and perhaps mask underlying differences across

different groups which may be present in real-life social situations (Risko et al., 2012). An important example highlighting the latter is demonstrated in a study comparing typically developing *adults* with autistic adults. Speer et al. (2007) demonstrated that the typically developing group presented with a bias towards attending to eyes in *dynamic* social videos but found no difference between groups when only *static* social images were used (Speer et al., 2007). This highlights the importance of the choice of paradigm when investigating social preference. Different images and modes of presentation may be measuring different phenomena and different mechanisms may underlie the preference for social stimuli across different paradigms.

Other studies using more complex dynamic paradigms with children born prematurely such as the human geometric preference task (Sekigawa et al., 2017) and biological motion tasks (Williamson et al., 2015) arguably offer greater ecological validity however their use is more limited in available research with MLP children. As such, future research should utilise well validated eye-tracking paradigms alongside well-validated behavioural tasks of social cognition.

Further problems with research in this area include a focus on the measurement of discrete domains of social cognition in isolation, which limits broader interpretations of social cognitive development. Zmyj et al. (2017) highlight the importance of investigating social cognition in more depth, outside of focusing on these discrete domains such as joint attention. This will support the development of a more nuanced picture of social strengths and weaknesses resulting from being born MLP. Exploration of *earlier* social cognitive skills is important to support with early identification of divergent development in MLP children. A focus on early social cognitive skills will also confer the opportunity to evaluate the developmental cascade model which suggests the importance of researching *how* observable features of difference emerge rather than *when* (Bradshaw et al., 2022).

The Early Social Cognition Scale

A battery of early social cognitive tasks, taken from well-established social cognitive measures by Tomasello and colleagues (Ellis et al., 2020; Tomasello 2014; Wellman 2014) has recently been developed (Ellis et al., 2020). The early social cognitive scale (ESCogS hereon - Ellis et al., 2020) investigates early social cognitive skills in depth, beyond discrete skills such as joint attention, and allows for a better understanding of the mechanisms through which MLP children acquire these earlier skills. Use of the ESCogS also affords the opportunity to contrast performance by MLP children with GA-matched, FT peers. Research in these earlier social cognitive skills will contribute to understanding of long-term outcomes for MLP children and help with earlier identification of difference/divergence in development of these skills, the latter of which has been argued as essential to increase the benefit of suitable interventions (Boardman & Fletcher-Watson, 2017; Dean et al., 2021; Meredith, 2015). Such research may support the development of tailored interventions which directly match the behavioural profiles of MLP children. Thus far, no studies have explored the skills included in the ESCogS with infants born prematurely.

Rationale

In summary, early social preference is suggested to play a key role in social development; however, our understanding of how such preference is associated with performance on explicit measures of social cognition is limited by a focus on atypical development and poor measurement of related constructs. Understanding how measures of early social preference are associated with other measures of social cognition more broadly is important in order to better understand how visual attention impacts other aspects of social development. A group in which the investigation of social preference and early social cognitive skills might be of particular importance is those born MLP. Being born prematurely results in increased differences across multiple areas of development, including higher rates of autism,

of which social communication differences are a key feature. Research investigating sub-threshold differences in social cognitive skills with MLP children has been neglected. A focus on exploration of social development, beyond the diagnostic taxonomy of autism, of those born within this GA is important. Particularly, a better understanding of the *earlier* development of social cognitive skills will have great clinical value in terms of earlier identification of possible divergent development in MLP children compared to those born FT. Given the growing number of preterm births and the greater proportion of MLP births, such a focus will support in identification of a potentially at-risk group which may require additional support and early intervention.

The Current Study

This study furthers research investigating social cognition in the following ways. Part one investigates how implicit measures of social preference (using eye-tracking methods) and explicit measures of social cognition (ESCogS) may be associated using well-validated paradigms for both constructs. Part two investigates performance in MLP children compared to FT children using well-validated paradigms which focus on these *earlier* social cognitive skills.

Objectives and Hypotheses

Part One

1. To use a sample of children aged 18-36 months to estimate the effect size for the correlation between implicit measures of social cognition (preference for social content) and explicit measures of social cognition (as measured by the ESCogS).
2. To use the obtained effect size (Spearman's Rho) to determine the sample size needed for an appropriately powered future study to test the significance of this relationship.

It was hypothesised that scores on the ESCogS would be positively correlated with social preference scores. That is, children who pass more tasks on the ESCogS would also have higher percentage looking time to social images.

Part Two

1. To compare a small sample of MLP children with FT children aged 18-36 months (matched on adjusted age and cognitive ability) on implicit (social preference) and explicit (ESCogS) and measures of social cognition.
2. To use obtained effect size (Cohen's d) to determine sample size required for an appropriately powered future study to test significance of this relationship.
3. To explore correlations between implicit (social preference) and explicit (ESCogS) and measures of social cognition between groups.

Methodology

Ethical Approval

Ethical approval was granted from the Science, Technology, Engineering and Mathematics Ethical Review Committee at the University of Birmingham (ERN_09-048AP9; See Appendix 1 for ethical approval). The data collected under this approval were part of a larger programme of work exploring social cognition in children born prematurely. The original protocol was amended to remove the Autism Diagnostic Observation Schedule – ADOS (Gotham et al., 2006) due to uncertainties post-pandemic with regards to re-commencing infant testing. Removal of the ADOS allowed for the testing procedure to be reduced from two days to one day, conferring feasibility advantages for recruited families.

Recruitment Strategy

Participants were recruited by contacting local nurseries, play groups and sure start centres, introducing the research project to nursery staff, disseminating posters and advertising on social media platforms. The charity Bliss (an organisation for families of children born prematurely or sick) were approached for project support and the research was advertised through their platforms. Recruitment also occurred via the Richards lab and Birmingham Baby lab social media platforms.

Participants

Participants were aged 18 to 36 months and included a MLP group ($32^{+0/6}$ - $36^{+6/7}$ weeks GA), and FT comparison group (37^{+0} - $41^{+6/7}$ weeks GA). The age of children born MLP was adjusted for GA and so MLP children in the study were eligible up to approximately 37 months chronological age. Age correction is standard practice in children born prematurely and allows for comparisons to peers of the same developmental age, rather than chronological age (NICE, 2017). Children born post mature (born >42 weeks' GA) were not eligible for the study as research suggests increased risk for divergent neurodevelopmental outcomes in this

population (Mackay et al., 2010; Moster et al., 2010; Rolschau et al., 2020; Yang et al., 2010). To be eligible for inclusion, children were required to have no siblings with idiopathic autism due to an increased likelihood of social-cognitive differences in siblings of autistic individuals (Pisula & Ziegart-Sadowska, 2015).

Part One

Part one was a correlation analysis and so it was important for participants with complete data sets to be included. Of the 27 children who participated in the research, *N*=9 did not have complete data sets for both eye tracking paradigms and ESCogS. For one child, this was due to insufficient eye signal quality. For three participants, lack of data was a result of a technical failure of the eye tracking equipment. The remaining five children did not settle during the eye tracking task. This included intolerability of the eye tracking sticker and not wishing to sit on their parent's lap and watch the video. The remaining 18 participants with both eye tracking data and ESCogS were included in part one data analysis. Five MLP children (37.5%) were excluded due to lack of eye tracking data in comparison to four FT children (21.1%).

Part Two

Part two explored *differences* between MLP and FT children. Out of the 27 children who participant in the study, *N*=8 were MLP and *N*=19 were FT. MLP children were matched with eight FT children to explore differences between groups. The process for matching is described in data processing and statistical analysis (see pages 128-130 of the current chapter).

Table 1 details the participant and family characteristics of the MLP and FT samples. This table also presents information according to the different analyses conducted.

Table 1

Table describing important participant demographics, health, birth and family characteristics for all participants, part one and part two analysis.

Participant & Family Characteristics[94] **Mean, (SD), *Range***	**All participants**	**Part one analysis**	**Part two analysis**		**t /X2**[95]	**p**[96]
	All participants (n=27)	**All participants (n=18)**[97]	**Term group (n=8)**	**Preterm group (n=8)**		
Chronological age (months)	25.48 (4.4) *18.43-34.59*	24.59 (3.7) *18.42-33.30*	25.4 (4.7) *18.4-33.3*	27.51 (4.7) *19.88-34.59*	-.89	.19
Corrected age (months)	25.17 (4.4) *18.06-34.40*	24.34 (3.9) *18.06-34.40*	25.5 (5.1) *18.2-34.4*	26.27 (4.8) *18.06-33.55*	-.29	.73
Gestational age (weeks)	38.35 (2.5) *32.00-41.28*	38.49 (2.6) *32.00-41.28*	39.6 (0.9) *38.4-41.3*	35.01 (1.6) *32.00-36.71*	6.86	**<.001**
Gender (% male)	59.3	55.6	50	62.5	.25	.50
Birthweight (kg)	3.88 (2.4) *1.75-14.40*	3.19 (.71) *1.75-4.04* **N=17*	3.4 (0.3) *2.9-4.0*	4.0 (4.6) *1.75-14.40* **N=7*	-.35	.37
Birth height (cm)	72.9 (7.1) *33.0-51.0* **N=8*	41.16 (7.5) *33.00-51.0* **N=6*	41.0 (11.3) *33-49* **N=2*	44.3 (21.6) *42.00-48.00* **N=2*	-	-
Head circumference (cm)	33.69 (2.8) *29-38*	32.92 (2.9) *29-37*	36.5 (N/A) *N/A*	31.16 (2.2) *29-35*	-	-

[94] *Indicates number of responses values are based upon when data were missing. Information is based on total samples size unless otherwise indicated.
[95] T-tests and chi square tests were run only when there were enough data in preterm and full-term groups to do so.
[96] Bold indicates significant differences between the MLP and FT group.
[97] Information provided based on the n-18 who completed both the ESCogS.and GeoPref tasks.

Participant & Family Characteristics[94] Mean, (SD), *Range*	All participants	Part one analysis	Part two analysis		t /X2[95]	p[96]
	All participants (n=27)	All participants (n=18)[97]	Term group (n=8)	Preterm group (n=8)		
	*N=12	*N=6	*N=1	*N=5		
Birth order N (% yes)	*N=11	*N=9	*N=3	*N=2	-	-
1st born.	6 (22.2)	6 (33.3)	3 (37.5)	1 (12.5)		
2nd born.	4 (14.8)	3 (16.7)	-	1 (12.5)		
3rd born.	1 (3.7)	-	-	-		
Apgar (1 minute)	7.10 (1.9)	6.71 (2.2)	7 (2.8)	6.2 (1.9)		
	4-9	*4-9*	*5-9*	*4-9*	.45	.32
	*N=17	*N=7	*N=2	*N=5		
Apgar (5 minutes)	8.50 (1.3)	8.43 (1.5)	9 (-)	8.2 (1.8)		
	5-9	*5-9*	9	*5-9*	.59	.19
	*N=17	*N=7	*N=2	*N=5		
Apgar (10 minutes)	9.25 (1.5)	7.00 (-)	-	9.0 (1.7)		
	7-10	*7-7*	-	*7-10*	-	-
	*N=23	*N=1	*N=8	*N=3		
Delivery Mode N (%)						
Vaginal	19 (70.4)	14 (77.8)	8; 100	3 (37.5)	7.27	**.026**
Caesarean section	3 (11.1)	-	-	4 (50)		
(scheduled)	5 (18.5)	4 (22.2)	-	1 (12.5)		
NICU Stay N (% yes)	-	-	0 (0)	8 (100)	16.00	**<.001**
Birth complications	-	-			-	-
N (% yes)			6 (75)[98]	3 (37.5)[99]		
Ventilator N (% yes)	-	-	0%	4 (50)	6.23	**.03**
Intracranial haemorrhage	-	-	0%	1 (12.5)	1.07	.50
N (% yes)						

[98] N=4 forceps, n=1 high infection rate, n=1 significant amount of blood loss.

[99] N=1 child not growing and had breathing problems resulting in caesarean, n=1 caesarean due to one twin not growing in womb, not complications once born, n=1 spontaneous labour – footling breech.

Participant & Family Characteristics[94] Mean, (SD), *Range*	All participants	Part one analysis	Part two analysis		t /X2[95]	p [96]
	All participants (n=27)	All participants (n=18)[97]	Term group (n=8)	Preterm group (n=8)		
Infant surgery during first year of life, N (% YES)			1 (12.5)[100]	0 (0)	-	-
Epidural N (% yes)	-	-	6 (75)	1 (12.5)	6.349	**.041**
Hospitalisation during pregnancy N (% yes)	-	-	2 (25)	5 (62.5)	2.29	.13
Significant periods of illness during pregnancy N (% yes)	-	-	0 (0%)	1 (12.5)	1.07	-.50
Drinking and/or smoking during pregnancy N (% yes)	-	-	0 (0%)	0 (0)	-	-
Breastfed N (% yes)	23 (85.20)	16 (88.90)	7 (87.50)	6 (75.00)	.41	.50
Maternal age at childbirth (years)	32.41 (5.19) *15-44*	32.22 (5.94) *15-44*	32.1 (8.1) *15-44*	31.4 (2.9) 26-36	.25	.40
Maternal education N (%)						
5 + GCSE's, NVQ 2	2 (7.4)	2 (11.1)	1 (12.5)	-	2.67	.26
3 + 'A' Levels, NVQ 3, BTEC	1 (3.7)	-	-	-		
National or equivalent	10 (37.0)	7 (38.9)	3 (37.5)	2 (25.0)		
Polytechnic/University degree, NVQ 4 or equivalent						
Masters/Doctoral degree, NVQ5	14 (51.9)	9 (50.0)	4 (50)	6 (75.0)		
Family Income N (% yes)						
<£25,000	1 (3.7)	1 (5.6)	-	-	6.48	.22
£25,001 to £35,000	4 (14.8)	4 (22.2)	1 (12.5)	2 (25.0)		
£35,001 to £45,000	2 (7.4)	2 (11.1)	1(12.5)	-		
£45,001 to £55,000	3 (11.1)	2 (11.1)	2 (25.0)	-		
£55,001 to £65,000	6 (22.2)	3 (16.7)	0 (0)	3 (37.5)		

[100] Surgery was for tongue tie revision.

Participant & Family Characteristics[94] Mean, (SD), *Range*	All participants	Part one analysis	Part two analysis		t /X2[95]	p[96]
	All participants (n=27)	All participants (n=18)[97]	Term group (n=8)	Preterm group (n=8)		
£65,001 or more	11 (40.7)	6 (33.3)	4 (50.0)	3 (37. 5)		
Child's ethnicity N (%)	*N=7	*N=4	*N=1	*N=5	-	-
White	6 (22.2)	3 (16.7)	1 (12.5)	4(50)		
Black, African, Caribbean/Black British	1 (3.7)	1 (5.6)	-	1(12)		

Materials

Health and Background Questionnaire

This questionnaire was developed with input from Clinical Psychologists, Neonatologists and experts in the field of early developmental research to describe the family and birth characteristics of participants. Questions fell within three broad categories of maternal history, birth experience and socio-economic information. See Table 1 for information collected from families and Appendix 2 for the health and background and questionnaire.

Modified Checklist for Autism in Toddlers-Revised with Follow Up (M-CHAT-R/F - Robins et al., 2009).

This widely used parental screening questionnaire screens children between the ages of 16 and 30 months for early signs of autism spectrum disorder (ASD) (Robins & Fein, 2018) and was used in the present study to describe the sample and identify participants with possible autism via parental report. Scores range from 0-20 and can be categorized as low, medium, or high risk. Parent-report scores falling within the 'medium risk' to 'high risk' category (<2) were explored further using a set of follow-up questions as per MCHAT-R/F measure protocol. The MCHAT-R/F is reliable for preterm populations when administered in full (Kim et al., 2016). See Appendix 3 for M-CHAT-R/F.

Play-Based Measures

Participants completed the early social cognition scale – ESCogS (Ellis et al., 2020) and the Mullen Scales of Early Learning – MSEL (Mullen, 1995).

ESCogS (Ellis et al., 2020; 2021). The ESCogS is a scaled, play-based battery of tasks developed by Ellis et al. (2020; 2021) taken from well-established social cognitive measures developed by Tomasello and colleagues (Ellis et al., 2020; Tomasello 2014; Wellman 2014). These tests of early social cognition, considered precursors to ToM abilities, assess six skills falling into four broad areas (helping, understanding of cues - gaze and point behaviours, re-

enactment of intended acts, cooperation). See Table 2 for a detailed summary of ESCogS tasks including typical age of acquisition.

Gutman scaling analysis conducted by Ellis et al. (2020;2021) indicated a co-efficient of reproducibility of 0.96, where a value >.90 indicated a reproducible scale (Ellis et al., 2020;2021; Green, 1956). The Index of consistency for the co-efficient was 0.5. Green (1956) suggest a value of >.50 indicates a co-efficient of reproducibility above what we would expect by chance (Ellis et al., 2020;2021), suggesting the tasks included in the ESCogS represent a developmental sequence of skill acquisition.

Table 2

Table summarising the Early Social Cognition Scale tasks, original authors, and description of individual task procedures, as described in Ellis et al (2020).

Task (Skill measured)	Original Task Author	Description of the task	Description of what is required to pass task	Typical age of acquisition[101]
The helping task (Helping behaviour) [102]	Warneken & Tomasello (2006, 2007)	Task originally derived from 'out of reach' tasks which explore a child's understanding of another person's intention. There are two tasks involving a pen and paper bells. There are two conditions of each (control and experimental). 1. ***Pen trial:*** In the control condition E1 accidently drops a pen and does not reach for it. In the experimental condition, the E2 accidently drops the pen and reaches for it unsuccessfully. 2. ***Paper bells trial:*** In the control condition the E1 places six polystyrene 'bells' on a table in front of the child and picks them up with large tongs. Three bells are placed in a container and E1 then stops and does not place any more in the box. In the experimental trial, three bells are placed out of E2 reach and they unsuccessfully attempt to reach for the three cones.	To pass this task the child is required to demonstrate correctly interpreting the experimenter's intention to access an object. This involves either passing the pen or paper bells back to the experimenter. To pass the task the child is required to pass just one of the two tasks (experimental condition).	14
The re-enactment of intended acts task (intention)	Bellagamba & Tomasello (1999)	Task originally derived from tasks assessing a child's ability to infer other's intentions based on their goal-oriented actions. There are three identical experimental trials utilising different object pairs: For each of the three trials, E1 models an intention to place a small loop on a peg. The experimenter 'fails' three times, showing the child this failure to the left, right and in front. Each object pair is then handed to the child and experimenter says, "can you try?". The three object pairs are as follows:	To pass this task, the child is required to infer another's intentions from their goal-oriented (but unsuccessful) actions. This involves successfully performing the target	18

[101] Age shown in months.

[102] The helping task includes a control condition which are included to examine whether successful performance by an infant was reflective of an understanding of experimenter's *intention* rather than either reinstating the original situation or taking possession of the objects (pen and bells) for themselves. This is explored statistically in preliminary analysis of the ESCogS as per original protocol (Ellis et al., 2020)

Task (Skill measured)	Original Task Author	Description of the task	Description of what is required to pass task	Typical age of acquisition[101]
		1. Loop and peg 2. String of beads and a plastic cup 3. Square board with a hole	action on at least two out of the three trials.	
The understanding others communicative cues tasks[103]	Behne et al., (2005)	This task includes two communicative cues (gestures pointing and gestures gaze). There are two control and two experimental conditions for both gaze and point cues and thus a total of eight trials. These cues are delivered in a counterbalanced way. The task is set up as a hiding game and includes a warm-up phase. Materials for the task includes three pairs of containers (of varying shapes and sizes), three toys to hide and a green piece of cardboard used as a hiding screen. Two experimenters are required for this task. The warm-up phase involves hiding a toy in one of the three pairs of containers without the screen. This orientates the child to the aim of the game, and they can see the container in which the experimenter hides the toy. This task involves two experimenters sat opposite the chid. In both the warm-up phase, control and experimental trials, E1 gains the child attention and says, "I'm going to hide it". The toy is placed in one of the pairs of containers and then E1 says, "where's the toy?". ***Gestures control trials (gaze and point):*** E1 uses a distracted point and absent gaze to the container with the hidden toy. ***Gestures experimental trials (gaze and point):*** E2 delivers an exaggerated point or an ostensive gaze to the container with the hidden toy, expressing their intent to direct attention to the box through raised eyebrows/or alternating gaze.	To pass this task, the child is required to understand the use of pointing and eye gaze to direct attention to an object. This involves correctly selecting the container with the hidden toy in the experimental gaze and point trials.	Gaze cues (18) Point cues (24)

[103] The communication tasks include a control condition for both point and gaze cues which are included to examine whether successful performance by an infant is reflective on an understanding of the experimenters point and gaze cues, rather than a response to the low-level attentional cues. This is explored statistically in preliminary analysis of the ESCogS as per original protocol (Ellis et al., 2020)

Task (Skill measured)	Original Task Author	Description of the task	Description of what is required to pass task	Typical age of acquisition[101]
The cooperation tasks. (Shared intentionality)	Warneken et al., (2006)	This task involves two games which both assess a child's ability to develop shared intentionality to achieve a joint goal with the experimenter. Both games follow the same format: 1. There is an initial 'show' phase whereby two experimenters demonstrate each game. 2. This is followed by four experimental trials. 3. Trial one and two involve familiarising the child with the game and playing with experimenter one 4. Trails three and four involve E1 interrupting the game with a 15 second pause in which the child's response behaviours are of interest. E1 then re-engages with the child to complete the game. More detail on each game is described below:	To pass this task, the child is required to achieve a joint goal with the experimenter and evidence an attempt to re-engage with the experimenter in an interrupted joint activity. Co-ordination and engagement are both coded for the 15 second period of interruption. This is then scored against a predefined coding schema from ESCogS original protocol. A median score is then calculated across trials. A pass is coded if child re-engaged with experimenter at least once *and* their median cooperation score is three.	24
Co-operation tubes		***Co-operation tubes:*** This is a problem-solving task using a large cardboard tube with handles which can be pulled apart and a small toy to hide inside. Both experimenters play the game with the child watching. The cardboard tube is pulled apart, and the toy hidden inside. Both experimenters work together to pull the handles on each side to retrieve the toy. E1 then plays the game twice with the child. During experimental trials three & four, after working with the child to hide the toy in the tube, E1 drops the handle interrupting the game. E1 waits 15 seconds and continues once this time has passed or the child re-engages with the experimenter (indicating they understand the need to work with experimenter one to complete the game).		
Co-operation trampoline		***Co-operation trampoline:*** This is a social game identical to the co-operation tubes game above with different equipment (a trampoline). A red foldable fabric trampoline and a small toy is placed in the centre. Two people are required to be able to bounce the toy on the trampoline. After following the same structure as above, E1 drops the trampoline (the toy cannot bounce – i.e the game cannot be achieved by one person). E1 waits 15 seconds and continues once this time has passed or the child re-engages with the experimenter (indicating they understand the need to work with E1 to complete the game).		

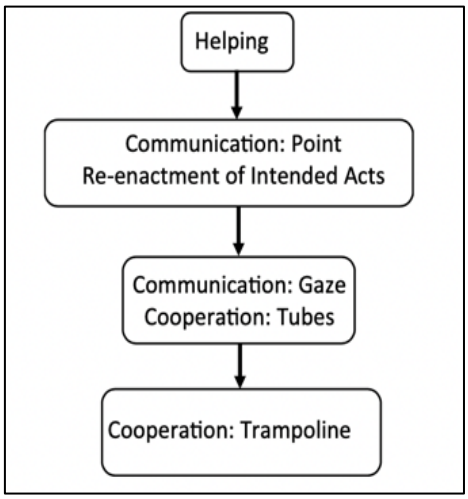
Note *– E1=Experimenter one; E2=Experimenter two.*

Ellis et al. (2020) details the developmental sequence in which typically developing children acquire these early social cognitive skills, of which, 88% of participants conformed to a four-stage scalable pattern. The scale demonstrated a co-efficient of reproducibility of 0.96 and index of consistency 0.5. Figure 1 depicts the four-stage cumulative scale, where tasks are ordered by level of difficulty. Ellis et al. (2020) additionally provide normative data for ESCogS for a sample of autistic children.

All tasks were administered as described in Ellis et al. (2020) original protocol[104]. Each task was scored as either pass or fail and total scores range from 0 to 6.

Figure 1

Figure showing the four-stage scale of early social cognitive skills from Ellis et al. (2020) – open access journal with permission to use with citation.



Mullen Scales of Early Learning - MSEL (Mullen, 1995). The MSEL is an individually administered, norm-referenced measure of cognitive functioning for children from birth to 68 months (Mullen, 1995). Four of the five Mullen scales were administered (visual

[104] Minor discrepancies were noted in the supplementary material from Ellis et al. (2021), the original footage from development of the scale available at the Richards lab and the scoring sheet for the ESCogS provided by Ellis et al. (2020). The original studies in which the tasks were developed were consulted, in addition to contact with the developer of the ESCogS for clarification. The discrepancies were predominately related to the roles of experimenter one and experimenter two and not what the tasks were measuring, A summary of these minor discrepancies is available in Appendix 4.

reception, fine motor, receptive language and expressive language) in order to compute individual subscale scores and the early learning composite (a summative measure of children's age equivalent abilities) as described in Mullen (1995). The gross motor scale was not administered due to the large number of tasks children were asked to complete and to prevent fatigue. The gross motor scale is not required for computation of the composite score.

Social Preference Measures

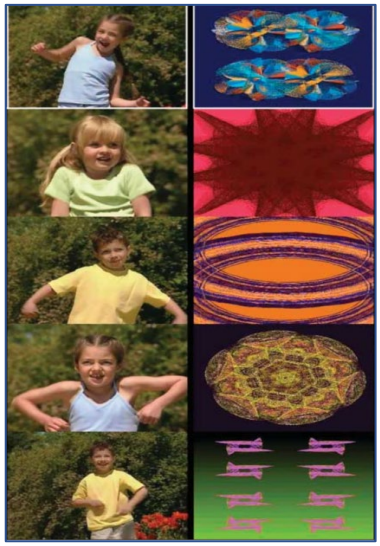
Participants completed two brief eye tracking tasks, both preferential looking paradigms using the EyeLink Portable Duo, TOBII T120 eye tracker (SR Research LTD) apparatus. All stimuli in the study, as described below, were programmed and run with Experiment Builder (Software licence from SR Research) and were presented on a 17" screen. The apparatus measures fixations and number of saccades in response to the visual stimulus presented. Areas of interest (AOI's) were presented side by side and described in further detail below.

The complex Geometric Preference Task - GeoPref (Pierce et al., 2011). This task is a 1-minute video clip depicting a dynamic social image (DSI), children engaging in dynamic yoga movements, and a dynamic geometric image (DGI), dynamic screen saver recordings. The DSI and DGI were presented side by side. To control for any preference mediated by spatial location, the side of presentation of DGI and DSI scenes were split with 50% of both MLP and FT groups shown a video with DSI on the left, as per recommendations by paradigm developers (Peirce et al., 2011). See Figure 2 for GeoPref stimuli. The stimuli in this task have demonstrated good discriminative validity in a range of studies between autistic and non-autistic samples (Moore et al., 2018; Pierce et al., 2011; Pierce et al., 2016). More recently, a large-scale validation study demonstrated excellent specificity in detecting autistic indidivuals from non-autistic indidivuals based on higher levels of visual attention towards geometric

images (Wen et al., 2022). Permission for use was requested and a contractual agreement was in place with the creator of the stimuli for use in this study. See Appendix 5 for agreement.

Figure 2

Figure depicting Geometric Preference task stimuli from Pierce et al. (2011) – permission to use as per contractual agreement.



Biological motion preference task -BioMotion. This task was developed by Annaz et al. (2012) using point-light display stimuli developed by Johansson (1973). The stimuli are presented side by side and include three stimuli depicting a biological motion walker, scrambled display, and a spinning top. On each trial a fixation cross appeared at the centre of the screen for 2 seconds (s). After the fixation cross disappeared, a *pair* of stimuli were displayed side by side for 6s. There were nine trials with biological/phased walker (referred to as condition one - C1 hereon) and nine trails with biological/spinning top (referred to as condition two – C2 hereon). See Figure 3 for stimuli. The location of the biological/non-biological motion stimulus on each trial was randomized to control for any preference mediated by spatial location as per recommendations by Annaz et al. (2012). The stimuli in this task have demonstrated good discriminative validity between a range of studies with autistic and non-

autistic samples (Mason et al., 2021). Permission for use was requested and granted for use in this study by task developers. See Appendix 6 for agreement.

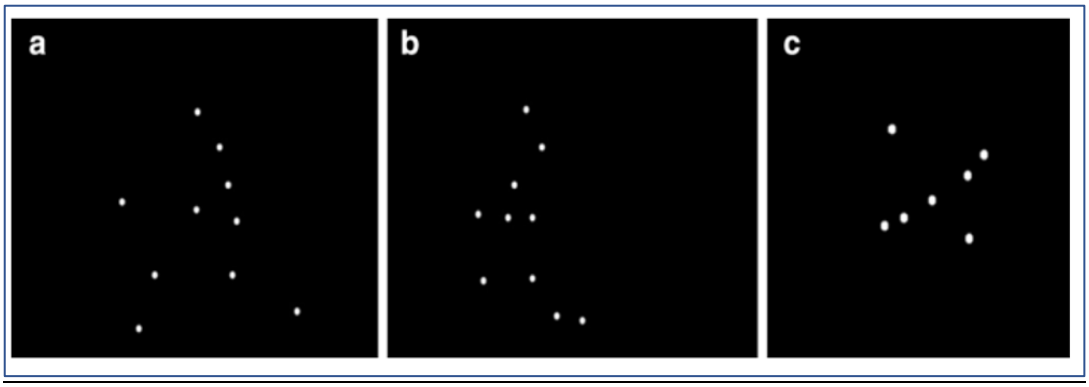
Figure 3

Figure depicting stimuli biological motion task (condition one and condition two) from Annaz et al. (2012)- permission to use as per contractual agreement.

(d) Biological motion walker (biological stimulus present across C1 and C2)

(e) Scrambled (C1)

(f) Spinning top (C2)



Procedure

Parents and their child were invited into the University of Birmingham Baby Lab for 2–3 hours. Upon giving consent, parents completed questionnaires whilst children completed the play-based tasks. Assessments were completed by researchers trained in the administration and scoring of tasks. At the end of assessment, children were provided with a small gift (typically a book, certificate, and Baby Lab T shirt). Parents were given a debrief form including researcher contact information and an individualised feedback report detailing performance on play-based tasks was posted to families.

For a subset of participants *(N*=14*)* whose data were collected as part of the larger study, of which this project was a part, assessments took place over two days. The ESCogS was the first task completed on day two of the original protocol.

Play-Based Measures

All assessments were video recorded to allow for further analysis if required. All assessments were completed in the same laboratory setting. Prior to administration, experimenters played with each child for 10-15 minutes with a standardised set of toys, to ensure they felt comfortable with the research setting. Children completed the ESCogS (Ellis, 2020; 2021) first, so the social cognition task was delivered without extensive social 'warm-up'. In line with the authors' guidance on administration, participants completed the tasks in one of six potential orders (see Appendix 7 for the possible administration orders). The ESCogS requires two experimenters to complete included tasks. A third researcher was present in the room for the purposes of video recording. A description of the ESCogS tasks is presented in Table 2.

Infants completed the MSEL (Mullen, 1995) second, with all tasks being administered in the standardised way as per testing manual, ensuring basal and ceiling scores were reached for each subscale. Breaks were taken between tasks when the child/caregiver indicated this was required.

Eye Tracking Procedure

The eye tracking procedure occurred in a separate dimly lit room, with the necessary equipment set-up prior to the start of the experiment. The tracker was set to sample at a rate of 2000 Hz. To maximise eye tracking data quality there were two critical environmental factors as per recommendations from SR Research LTD. These included the child's distance from the monitor and head level to eye tracking camera. Children were seated approximately 30cm from a 19 LED monitor (screen resolution 1280 x 1024) and children's eyes were positioned in the upper quarter of the monitor to optimise the trackable range of the eye tracker camera.

Children were required to wear a target sticker in the middle of their forehead with an approximately equal distance between the eyes. The target sticker allows for the eye tracking camera to track both eyes of the child without restricting movement. Prior to placing the sticker on the child's head, the procedure was shared with parents to consider the best way to place the sticker on the child's forehead with minimal distress. Children were seated on their parent's lap and a number of steps were required to maximise data quality.

Pupil and Corneal Reflection (CR) Threshold Values

Appropriate pupil-CR thresholds were essential for image processing algorithms to function correctly to produce good quality data. Reasonable optimal thresholds were obtained from those recommended by SR Research LTD platform and were obtained prior to calibration. To obtain appropriate pupil-CR thresholds, the display screen visible to the child was set up to allow the child to see their own face and eyes whilst the experimenter adjusted the equipment to the obtain the necessary thresholds. To complete this task, the child must be looking at the camera, and the child being able to see their face on screen, as well as their caregiver optimised interest to the screen.

Calibration, Validation, and Drift Correction Procedures.

To maximise the accuracy and precision of the eye tracking data, data were not recorded until a five-point calibration validation procedure had been repeated and average accuracies across the five positions on the screen was less than half a degree. Calibration targets involved moving circles to capture child's attention; stimuli were sourced from SR Research LTD. See Figure 4 for image of calibration stimuli. To engage the child with the calibration targets, the experimenter would verbally direct the child to look at the image on the screen and encouraged them to look at where the image was moving on the screen.

After calibration, a validation check was performed to ensure the accuracy of the calibration. This involved one target presented in the centre of the screen. When both of the

child's eyes were looking at the target, the validation check was complete. As calibration is only a *point in time* accuracy check, drift correction procedures were automated between trial recordings to check collected data remained spatially accurate throughout recordings. Drift correction therefore was automated at the beginning and end of the GeoPref task and then between trial recordings for the BioMotion task (C1 and C2). There were 9 trials for C1 and 9 trails for C2. Drift correction involved a target presented at the centre of the screen. After each trial the drift correction procedures automatically requested for the experimenter to check the positioning of the child's eyes. The next trail would start only when both of the child's eyes were looking at the target. The experimenter pressed the button to confirm to the eye tracking computer the spatial location of the child's right and left eyes.

The two social viewing tasks were presented to children with the GeoPref task first and BioMotion task second (C1 followed by C2). Figure 5 depicts the visual order of administration of tasks. When children were no longer engaging with the screen, became agitated, or indicated they no longer wanted to watch the video, the experiment was ended and viewing data up until that point was recorded by Data Viewer. Therefore, data was only recorded for trials the child watched. As the BioMotion task was presented second, not all children watched all trials for C1 and C2.

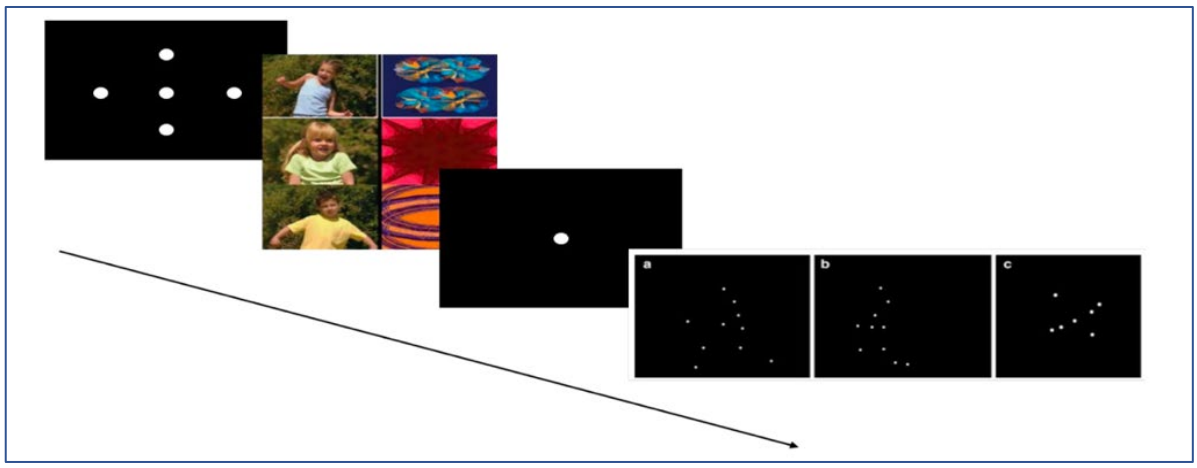
Figure 4

Figure depicting infant calibration image from SR Research LTD – Image open access through SR Research LTD.



Figure 5

Figure depicting order of presentation of eye tracking stimuli.



Data Processing and Statistical Analysis

Data analysis was separated into two parts with different aims. Data processing (eye tracking tasks) and exploratory analysis (ESCogS) were completed for part one/two as described below.

Social Preference

Eye tracking data were processed and exported using EyeLink Data Viewer software package (SR Research Ltd., version 4.3.210). SPSS software version 29 (SPSS Inc, 2013) was used to perform statistical analyses. During data processing, areas of interest (AOI) were coded for social and non-social images for each task. Dwell time for each AOI in milliseconds was tabulated for each participant. Dwell time includes all fixations and saccades within each AOI and is an excellent metric to convey level of interest (Albert & Tullis, 2023). For eye tracking paradigms which involve multiple trials (as is the case for BioMotion tasks in this study), Data Viewer calculates dwell time for each AOI divided by the number of trials using each interest area. Dwell time for each AOI was then divided by the total dwell time for both AOI's to derive proportion of time spent on each AOI (i.e. % social, % non-social). This was important to not include time the child spent looking away from both AOIs and therefore a more accurate representation of preferential looking at social and non-social images.

ESCogS Data

Preliminary analysis was conducted on the ESCogS as per Ellis et al. (2020) protocol for *all participants.* Three tasks in the battery (helping and communication point and gaze tasks) involve both control and experimental conditions. Ellis et al. (2020) analysed these separately to inspect whether behaviours were a result of correct interpretation of intention instead of a reinstatement of the original situation (helping task) or due to low level attentional cues (communication tasks). Refer to Table 2 for ESCogS task descriptions. In line with Ellis et al., (2020) scoring protocol, for the helping task, performance was explored by looking at the percentage of children who passed the experimental condition, that also passed the control condition. For the communication tasks, performance was explored using Pearson chi-square tests. Assumptions of this test require expected frequencies to be greater than five to rely on significance values (Field, 2017; Howell, 2012) which is not the case here. Due to this, Fisher's exact test was therefore used to compute significance (Field, 2013). A nonsignificant result here would mean children who passed communication trials, did so because they were correctly interpreting the experimenters' communicative cues and not simply attending to low level attentional cues. This preliminary analysis was not conducted for part two data analysis due to small sample size for the test to be computed meaningfully.

For the ESCogS, total scores were calculated by scoring each individual task using the pass/fail criteria outlined in Ellis et al. (2020) protocol. Performance comparison of *all participants* who completed the ESCogS was compared against the four-stage cumulative scale described in Ellis et al. (2020) sample of typically developing children. This was explored via group scaling conducted in Excel whereby tasks were placed in hierarchical order (according to the scale described by Ellis et al., 2020), and performance for each participant was mapped onto this to identify those participants who conform to the scale and those who present with different patterns of performance.

Part One Data Analysis Plan

A correlation matrix was used to describe associations between each social preference task (GeoPref, BioMotion – C1 and C2) and performance on the ESCogS (total scores, range 0-6). The ESCogS is a criterion-referenced measure whereby scores are not normally distributed and therefore non-parametric methods were used for statistical analysis. Spearman's rho was used to calculate an estimate of the correlation between performance on each measure of social preference and ESCogS.

Part Two Data Analysis Plan

Preliminary analysis and data processing for both social preference tasks and the ESCogS were performed as described above. MLP children were matched with FT children on corrected age first and then MSEL composite scores. Independent t-tests were then performed to explore differences between groups on these domains. Differences between other important demographics (such as ethnicity, socio-economic status and measures of health status) were also explored using t-tests and chi-square as appropriate. Test statistic values and *p* values are reported in Table 1.

Between group differences on ESCogS (total scores) were analysed with non-parametric methods (rationale as described above). Kolmogorov-Smirnov test was used to calculate an estimate of the difference between groups. This test was used instead of Mann Whitney U because it has better power than the Mann Whitney when samples are less than 25 per group (Field, 2013).

Between group differences on measures of social preference were analysed using independent samples t-tests. Cohen's d was used to calculate the estimate of the difference between groups.

Associations between social preference (implicit) and ESCogS (explicit) measures between MLP and FT groups were presented visually in scatterplots to explore the possible strength and direction of any relationships that might exist.

Across part one and two of the analysis, G*Power Version 3.1 (Erdfelder, Faul, & Buchner, 2005) was used to perform a power analysis to determine the sample size necessary for a full-scale study, given the effect size from the current sample. Where results indicated a correlation *or* difference in the unexpected direction power analyses were not conducted. Additionally, given the unreliability of statistics obtained from small samples, significance values have not been reported throughout (Hung et al., 2017).

Results

Preliminary Analysis

Helping Task

Of the 77.8% *(N*=21*)* children who passed the experimental condition of the helping task, only 18.5% *(N*=5*)* also handed over the pen to the experimenter in the control condition and only 14.8% *(N*=4*)* also handed over the bells to the experimenter in the control condition. From those that passed the experimental condition, 18.5% *(N*=5*)* took possession of the bells and 14.8% *(N*=4*)* took possession of the pen before handing over to the experimenter during experimental trials.

Communication Task

Pearson chi square test revealed performance did not differ significantly from chance for either task; control point -$X^2(2) = 1.32$; control gaze $X^2(2) = 1.33$.

Performance on the ESCogS

To provide a comparison of children's performance to a normative sample, Table 3 summarises individual task pass rates in comparison to Ellis et al., (2020) typically developing sample. Table 4 shows ESCogS tasks in the order of difficulty/sequence in which children typically pass these tasks as presented by Ellis et al., (2020). A total of 15 children (55.5%) conformed to this sequence, compared to 88% in Ellis et al., (2020) sample.

Table 3

Table showing individual pass rates on the Early Social Cognition Scale for all participants in comparison to Ellis et al.'s (2020) typically developing sample.

Early Social Cognition Scale (Task)	Number of children passing task (%)	
	Current sample *(n=27)*	TD[105] sample from Ellis et al., (2020 *(n=88)*
Helping	21 (77.8%)	76 (88%)
Communication: Point	10 (37%)	58 (67%)

[105] Typically developing sample

Re-enactment of intended acts	11 (40.7%)	54 (63%)
Communication gaze	8 (29.6)	37 (43%)
Cooperation: Tubes	6 (22.2)	32 (37%)
Cooperation: Trampoline	6 (22.2%)	19 (22%)

Table 4

Table showing group scaling and participant characteristics according to the Early Social Cognition Scale compared to Mullen composite score and age.

	Patterns					
Pattern	**0**	**1**	**2**	**3**	**4**	**Other patterns**
Helping						-
REI[106] or communication point						-
Communication: Gaze or Co-operation: Tubes						-
Trampoline						-
Number of cases	1	1	5	5	3	12
Age (mean)[107]	18.4	20.8	24.5	23.8	23.6	28.1
Age (SD)	N/A	N/A	3.1	3.6	1.3	4.6
Age (Range)	N/A	N/A	22.0-29.6	21.1-28.8	22.1-26.6	19.9-34.6
Mullen composite score (mean)	122.0	101.0	91.8	85.0	115.7	106.3
Mullen composite score (SD)	N/A	N/A	12.2	17.8	10.1	26.7
Mullen composite (Range)	N/A	N/A	77-111	68-108	105-125	63-141

***Note** – Green = passed task, red = task failed*

As can be shown in Table 4, older children were more likely to pass more demanding tasks identified in the sequence. There was however crossover in age observed in patterns three and four. A total of 12 children demonstrated other patterns of performance which are described in further detail in Table 5.

[106] Re-enactment of intended acts task
[107] Age in months; NA = not applicable

Table 5

Table showing group scaling and participant characteristics according to the Early Social Cognition Scale compared to Mullen composite score and age, for participants demonstrating ***other*** *patterns of performance on the Early Social Cognition Scale.*

	Other observed patterns				
Pattern	**5**	**6**	**7**	**8**	**9**
Helping					
REI[108] or communication point					
Communication: Gaze or Co-operation: Tubes					
Trampoline					
Number of cases	1	3	6	1	1
Age (mean)[109]	34.5	23.8	27.4	33.3	33.1
Age (SD)	N/A	4.5	2.7	N/A	N/A
Age (Range)	N/A	19.9-28.8	24.3-32.3	N/A	N/A
Mullen composite score (mean)	69	78	102.7	117	124
Mullen composite score (SD)	N/A	15.5	29.8	N/A	N/A
Mullen composite (Range)	N/A	63-94	63-133	N/A	N/A

Note – Green = passed task, red = task failed

ESCogS Associations with MSEL and Corrected Age. A Spearman correlation coefficient was computed to assess the relationship between ESCogS total scores, MSEL composite scores and corrected age for all 27 participants completing the ESCogS. Spearman's rho yielded a very small positive correlation between ESCogS and MSEL composite

[108] Re-enactment of intended acts task
[109] Age in months; NA = not applicable

performance (*rho* = -.09), with a CI of -.25 to 1.00. Spearman's rho yielded a small positive correlation between ESCogS and corrected age (*rho* = .128) with a CI of -.21 and 1.00.

Social Preference. Out of 27 participating children, *N*=18 viewed the GeoPref, *N*=16 viewed the BioMotion (C1) and *N*=10 viewed the BioMotion (C2). Table 6 shows summarised data from all three tasks. Table 7 shows individual participant data including number of trials completed.

Table 6

Table showing mean dwell time and percentage dwell time to social images across the three eye tracking tasks for all participants (Geometric preference task, biological motion task – condition one and biological motion task – condition two).

Eye tracking task	**Dwell time to social stimuli (in s)**			**Dwell time % looking time to social AOI**		
	Mean	SD	Range	Mean	SD	Range
GeoPref *(n=18)*	34.8	16.3	12.3 -59.2	79.4	16.9	38.2-98.7
BioMotion (C1) *(n=16)*	1.97	1.33	0.35-5.5	54.2	22.1	23.4-100
BioMotion (C2) *(n=10)*	1.57	1.07	0.0-3.5	38.6	26.9	0-100

Note *-s = seconds*

Part One Results

Part One aimed to describe associations between implicit (social preference) and explicit (ESCogS) measures of social cognition and to conduct power analyses to estimate the minimum sample size required to produce a valid significance test for a fully powered future study, were the same associations observed. Table 7 shows individual participant characteristics, subsequent scores across measures and which eye tracking tasks were completed.

Table 7

Table showing individual participant characteristics across measures. Including sex, chronological age, corrected age, Mullen composite scores, Modified Checklist for Autism in Toddlers-Revised scores, Early Social Cognitive Scale scores, percentage time viewing social area of interest and number of trials viewed for each eye tracking paradigm.

ID[110]	sex	Chron. Age (months)	Correct. Age (months)	Gest. Age (weeks)	Mullen Total[111]	MCHAT Total[112]	ESCogS Total (0-6)	Geo Pref	BioMotion (C1)		BioMotion (C2)	
								%time viewing social image	%time viewing social image	Trials[113]	%time viewing social image	Trials[114]
1	male	24.6	24.6	40.14	125	0	4	78.4	52.2	9	36	4
2	female	22.2	21.7	37.43	111	0	4	84.5	77.2	9	-	
3	male	26.3	26.4	40.14	133	0	3	94.3	86.4	9	-	
4	female	22.1	22.3	41.14	117	1	4	72.2	34	9	31.2	8
5	female	20.8	21.0	41.14	101	0	1	93.2	100	1	-	
6	male	24.3	24.4	40.29	124	0	3	92.8	50.9	9	38.6	8

[110] Participant numbers in **bold** indicate the full-term children which were included in part two analysis.
[111] Mullens composite score. Asterisks show when adjusted age was used to compute Mullens (when this was different from the use of chronological age). Mullens composite score shown here.
[112] Asterisks show MCHAT score after administering follow up questions – total MCHAT score shown here.
[113] Number of trials viewed (out of 9)
[114] Number of trials viewed (out of 9)

ID[110]	sex	Chron. Age (months)	Correct. Age (months)	Gest. Age (weeks)	Mullen Total[111]	MCHAT Total[112]	ESCogS Total (0-6)	Geo Pref	BioMotion (C1)		BioMotion (C2)	
								%time viewing social image	%time viewing social image	Trials[113]	%time viewing social image	Trials[114]
7	male	18.4	18.2	38.86	122	0	0	93.5	-	-	-	
8	female	28.8	28.7	39.71	141	0	2	65.8	33.4	9	100	1
9	male	33.3	34.4	40.4	117	1*	4	83.7	63.0	5	-	-
10	female	21.7	21.5	39.14	76	0	4	89.2	71.1	9	40.4	8
11	female	22.9	22.5	38.42	77	1*	1	91.7	-	-	-	
12	male	24.0	24.0	41.28	105	0	4	77.5	32.3	2	-	
13	female	28.1	28.7	39	121	0	2	65.8	43.2	9	63.3	8
14	*male*	26.1	25.4	*36.71*	63	0	3	64.5	54.7	9	0	2
15	*male*	26.5	25.7	*36.71*	73	1	3	50.6	69.2	9	28.8	8
16	*male*	28.3	27.5	*36.42*	68	0	4	95.1	46.3	9	28.1	3
17	*female*	19.9	18.1	*32*	94	0	1	38.2	29.6	9	19.5	8
18	*male*	24.6	23.2	*34*	77	0	2	98.7	23.4	4	-	

Note – *Green = indicated social preference task* ***was viewed*** *by participant; red = social preference task* ***not viewed*** by participant.

GeoPref and ESCogS

A total of 18 participants viewed the GeoPref and completed the ESCogS. A Spearman correlation coefficient was computed to assess the relationship between percentage looking time to social images on the GeoPref and total score on the ESCogS. A positive correlation was hypothesised. Spearman's rho yielded a very small *negative* correlation between social preference and ESCogS performance (*rho*= -.05), with a CI of -1.00 to 0.37. See Figure 6A for scatterplot. Expected associations were not observed, therefore power analysis was not conducted.

BioMotion(C1) and ESCogS

A total of 16 participants viewed the BioMotion (C1) and completed the ESCogS. A Spearman correlation coefficient was computed to assess the relationship between percentage looking time to social images (biological motion) and total score on the ESCogS. A positive correlation was hypothesised. Spearman's rho yielded a *very small positive* correlation between social preference and ESCogS performance (*rho*=.18), with a CI of -.28 to 1.00. See Figure 6B for scatterplot.

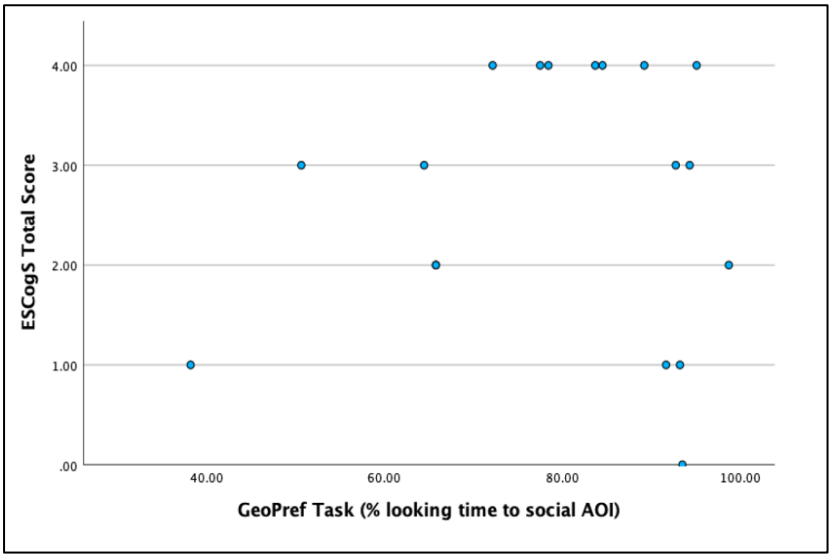
Power Analysis. Based on a one-tailed hypothesis, a sample size of 182 would yield 80% power for detecting a correlation of $\rho = .18$.

BioMotion (C2) and ESCogS

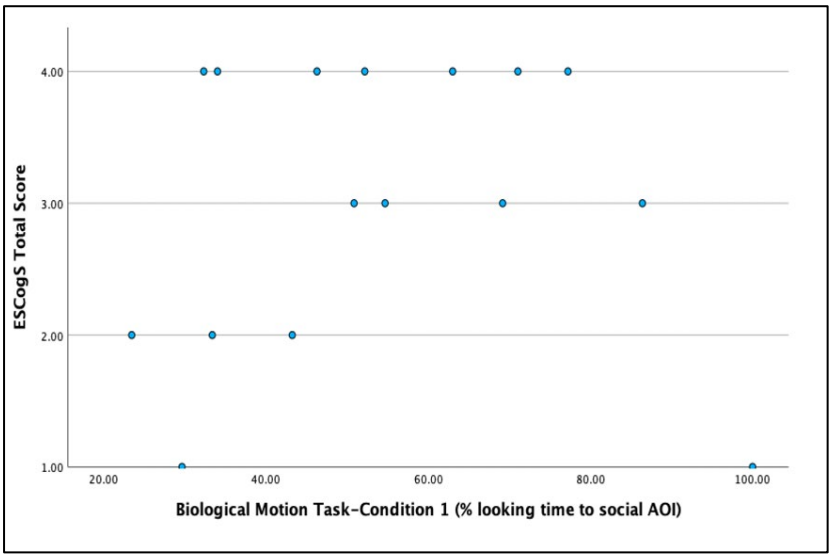
A total of 10 participants viewed the BioMotion (C2) and completed the ESCogS. A Spearman's correlation coefficient was computed to assess the relationship between percentage looking time to social images (biological motion) and total score on the ESCogS. A positive correlation was hypothesised. Spearman's rho yielded a very small *negative* correlation between social preference and ESCogS performance (*rho*= -.08), with a CI of -1.00 to 0.51 A scatter plot of individual children's performance is shown in Figure 6C. Expected associations were not observed; therefore, power analysis was not conducted.

Figure 6: Figure showing scatterplot of social preference measures correlated with the Early Social Cognition Scale.

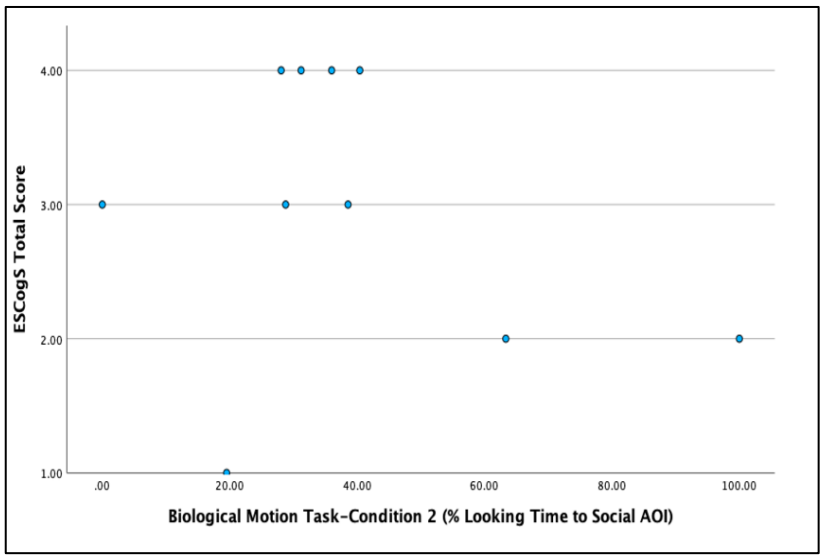
A. Geometric Preference task and the Early Social Cognition Scale.



B. Biological Motion task (condition one) and the Early Social Cognition Scale.



C. *Biological Motion task (condition two) and the Early Social Cognition Scale.*



Part Two Results

Part Two aimed to describe differences between MLP and FT children on social preference and ESCogS and to conduct a power analysis to determine minimum sample size required to produce a valid significance test for a fully powered future study. Associations between implicit (social preference) and explicit (ESCogS) measures of social cognition between groups were also explored visually.

MLP and FT children were comparable based on adjusted age, t (14) = -.29, p = .77 and MCHAT scores, t (14) = .25, p = .81. Groups did however differ significantly on MSEL composite scores, with MLP children demonstrating lower scores compared to FT children, t (14) = 2.38, p = .03. MLP and FT children also differed significantly on gestational age as expected. Other characteristics indicating a significant difference were NICU stay and the use of ventilator after birth, both of which MLP children required increased use (See Table 1 for full demographic characteristics).

ESCogS.

Pattern of Performance on ESCogS. To provide a comparison of ESCogS performance between MLP and FT children, group scaling in line with the difficulty/sequence

reported by Ellis et al., (2020) was mapped for each group. Three children (38%) in *both* MLP and FT groups conformed to this sequence, compared to 88% in Ellis et al., (2020) sample. Appendix 8 details participants in each group who conformed to this sequence and Appendix 9 demonstrates other patterns of performance on the ESCogS between groups.

Table 8

Table showing Early Social Cognitive Scale individual task pass rates for moderate to late preterm children and full-term children.

Task	Number of children passing task (%)	
	PT children (n=8)	MLP children (n=8)
Helping	6 (75%)	4 (50%)
Communication: Point	3 (38%)	4 (50%)
Re-enactment of intended acts	1 (13%)	3 (38%)
Communication gaze	4 (50%)	3 (38%)
Cooperation: Tubes	6 (75%)	3 (38%)
Cooperation: Trampoline	1 (13%)	3 (38%)

ESCogS Performance Between Groups. ESCogS total score performance in MLP children (*MDN*=3) did not differ from FT children *(MDN*=2.5), KS=.50. As the expected direction of difference was not observed, therefore power analysis was not conducted.

Performance on Social Preference Tasks.

GeoPref. On average, MLP children spent less time viewing social stimuli (*M*=69.2, *SE*=11.9) than FT children (*M*=82.7, *SE*=4.2). This difference, -13.3, Bca 95% CI [-45.91, 19.38] represented a medium effect size (Cohen, 1960), *d*= -0.71.

Power Analysis. An a priori power analysis conducted for sample size estimation, derived from the effect size reported here (.71). with a significance criterion of *alpha*=.05 and power =.80, the minimum sample size needed with this effect size is *N*= 52 (*N*=26 per group) for a one-tailed independent samples t-test of mean differences.

BioMotion (C1). On average MLP children spent less time viewing social stimuli (*M*=44.6, *SE*=8.3) than FT children (*M*=54.9, *SE*=9.0). This difference, -10.26 Bca 95% CI [-38.45, 17.93] represented a medium effect size, *d*= -0.49.

Power Analysis. An a priori power analysis conducted for sample size estimation, based on the effect size reported here (.49), with a significance criterion of *alpha*=.05 and power =.80, the minimum sample size needed with this effect size is *N*= 102 (*N*=51 per group) for a one-tailed independent samples t-test of mean differences.

BioMotion (C2). Only three MLP children viewed the second condition of the BioMotion task, therefore planned statistical analyses were not conducted.

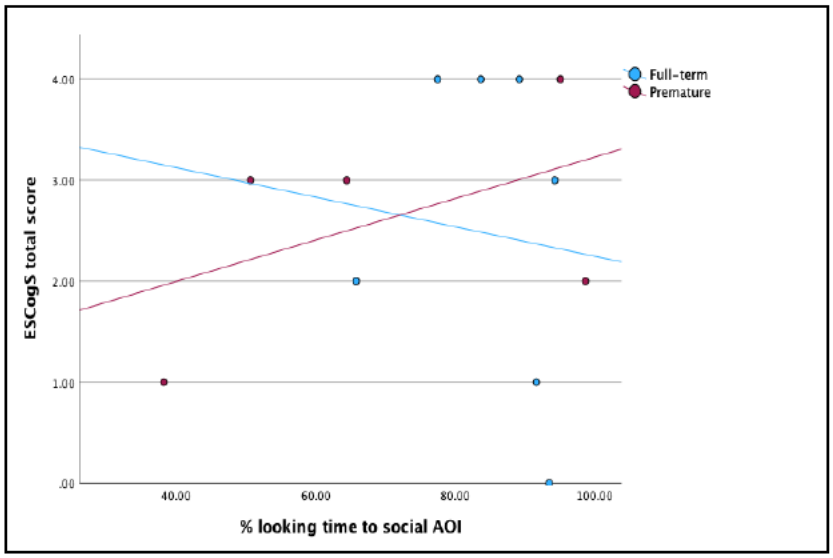
Correlations Between Social Preference and ESCogS Between Groups

Correlations of social preference (GeoPref and BioMotion –C1) and ESCogS total scores in MLP and FT children were visually presented using scatterplots (see Figure 7). Statistical analysis was not conducted due to the small preterm sample size (*N*=5) who completed both measures of social preference and ESCogS.

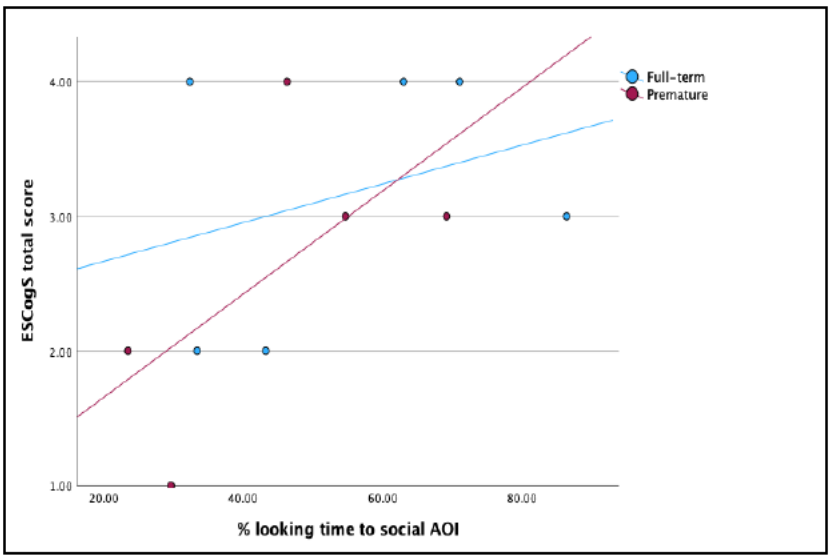
Figure 7

Figure showing scatterplot of social preference tasks and the Early Social Cognition Scale scores with line of best fit for moderate to late preterm children and full-term children.

A. Geometric Preference task and the Early Social Cognition Scale



B. Biological Motion task (condition one) and the Early Social Cognition Scale



Discussion

Part one of this study investigated how implicit and explicit measures of social cognition may be associated and part two investigated differences in performance in MLP children compared to FT children on *earlier* social cognitive skills. Both studies were the first to explore associations and differences using the GeoPref, BioMotion and ESCogS.

Preliminary Analysis (ESCogS)

Preliminary analysis of ESCogS helping task suggested a smaller proportion of children who passed the experimental condition also performed the target act in the control condition. Analysis of communication (point and gaze) tasks indicated that the control cues had not been sufficient to direct children's attention to the location of the toy. Whilst we cannot be certain, analysis suggests that children understood the experimenter's intention and their actions were indicative of helping behaviour and that children were using experimental cues to select the location of the toy.

The sample did not consistently conform to the same sequence of acquisition as described by Ellis et al., (2020), with only 55% performing in this pattern compared to 88% in the sample reported by Ellis et al. (2020). However, the number of children passing each task was broadly similar to that of Ellis et al.'s, (2020) typically developing sample. Given the unreliability that comes from small sample sizes, it is probable that the current results are impacted by this. Interestingly however, Ellis et al. (2020) reported a sample of autistic children (N=21) of which 90.48% conformed to the ESCogS scalable pattern of performance, with differences demonstrated in a *delay* of these skills but the same pattern of acquisition. In other work, Ellis et al. (2022) describe alternative developmental sequences for different syndrome groups, which may indicate that children who do not conform to this sequence may represent an atypical population. Given the age of children in this study, the possible impact of the covid pandemic on children's social development may also be an important factor to consider. Given

the size of our sample, it is not possible to draw conclusions about the mechanisms that led to the observed pattern of performance.

It might therefore be reasonable to consider the evidence for the ESCogS scale. The index of consistency is 0.5 which represents the cut off for a scalable pattern (Green, 1956). It is possible that the scale is unstable, particularly when observing the tasks in the middle of the scale. Alternative developmental sequences of the scale for the current sample were described in Table 5. Pattern numbers six and seven were relatively common and may suggest some ambiguity in the scalability of the middle tasks. This may indicate that the ESCogS is not as robust a scale in terms of sequence of acquisition and that the total ESCogS score may be a more reliable indicator of social cognitive skills in the current sample. There were only very small correlations suggested between ESCogS performance and the MSEL composite scores and corrected age. It is possible that an alternative latent variable may be impacting social cognitive development in this sample. Replication of the pattern obtained by Ellis et al. (2020) pattern of performance in a larger typically developing sample is required.

Part One

This was the first study to explore the relationship between implicit and explicit measures of social cognition using well validated tasks. Whilst we must be cautious with pilot data, results here may suggest there is little evidence that a strong correlation exists between the implicit and explicit measures used in this study. In line with previous findings, our results show inconsistencies, with only one implicit task demonstrating a very small positive association with explicit social cognitive skills as hypothesised, and others demonstrating very small negative correlations (Bacon et al., 2020; Dubey et al., 2022; Rijn et al., 2018; Schietecatte et al., 2012). The lack of observed consistent associations is, however, interesting and important. Differences in early social preference are suggested to place a child on an alternative developmental trajectory, indicating that this may be a mechanism through which

other divergent social behaviours emerge. The pilot cross-sectional data presented here does not rule out the possibility of this theory. It might be that early social preference predicts *later* social cognition, but not cross-sectionally. Well powered, longitudinal studies will be essential to understand this better.

As part one suggested strong correlations between implicit and explicit measures of social cognition may be unlikely, scatterplots of associations between MLP and FT children on measures of implicit and explicit social measures were interesting. From visual inspection, social preference scores in the graph appear more strongly correlated with ESCogS scores compared to FT children. We must however be cautious with interpretations of the scatterplots presented and therefore cannot draw any conclusions from these descriptive statistics.

Part Two

Part two investigated differences in implicit and explicit measures of social cognition in MLP and FT children with a small pilot sample. MLP children spent less time viewing social stimuli in both tasks both with medium effect sizes. Power analysis estimated a total sample of 52 would be required for a full powered study for differences in GeoPref between groups and 102 for BioMotion (C1), if the same effect sizes were maintained. There were no differences in ESCogS between groups.

There is no previous research investigating differences in social preference and ESCogS in MLP children, and as such, these pilot data make an important contribution to the literature. In children born prematurely at earlier GAs, Dean et al., (2020) found differences in social preference between preterm children and their FT peers, however these differences did not persist when the same children were tested again at the age of 5. Additionally, these findings are limited by the use of static eye tracking paradigms, which may lack ecological validity (Dean et al., 2020). The results in the present study, drawn from well-validated eye tracking paradigms, *could* be consistent with a difference in social preference in MLP children between

the ages of 18 to 36 months compared to FT children. However, with the small sample presented here, we were not able to demonstrate this statistically. Sample size estimates for a fully powered study exploring differences in MLP and FT children in measures of social preference are reasonable and practical with the support of NHS recruitment pathways to ensure a MLP sample is representative and to ensure numbers recruited for a fully powered study are practically possible. Whilst it is important to highlight that the effect size obtained in a small sample provides an unreliable estimate of the actual effect size. If a medium effect size were indicated in a larger sample size, this would likely be quite meaningful. As per the social motivation theory, if MLP children find social stimuli less rewarding, this will likely impact real-word exposure to social interactions with peers and impact development of later social competence. This might then be useful mechanistic information to inform support provided in education and home settings.

The pilot data here showed no evidence of trends for differences between MLP and FT children on explicit measures of *early* social cognition, measured with the ESCogS. The limited studies of social cognition with MLP children suggest reduced joint attention skills (Fenoglio, 2019; Gueron-Sela et al., 2015; Speretto et al., 2015) and reduced affect recognition abilities (Longa et al., 2022; Marotta et al., 2019), which are at odds with the pilot data presented here. It is important to hold in mind other potential pathways which might link prematurity and differences in social cognition as compared to full-term children. Many aspects of development are inextricably linked with social cognition. The development of other skills such as language, cognition, executive functioning, and motor skills may contribute to differences in social cognition outside of prematurity (Bradshaw, 2022) The preliminary evidence of differences identified in other studies, and possible differences in social preference suggested here are important to warrant further investigation. How these differences develop from earlier life, and

the impact of differences in social preference in this group on social cognition, is still not yet well understood.

Tolerability and Feasibility of the Protocol

The current study suggests MLP *and* FT children were able to tolerate the protocol. However, the fewer number of trials viewed by children in the BioMotion (C2) task likely impacted the quality of the data and thus conclusions which can be drawn. The eye tracking task was often the final assessment of the day and as such children may have been tired and thus less able to engage. Reducing the task time to exclude the BioMotion (C2) task *or* changing the timings of eye tracking assessments may have allowed children to engage in these tasks and is an important strategy to support future data collection. The calibration attention grabbers used here were not sufficiently engaging and often calibration took longer than hoped. Ensuring calibration attention grabbers are sufficiently engaging may be important for future eye tracking studies of children within this age range.

Recruitment of the MLP sample was difficult and the sample collected here may not be representative of the broader MLP population. Qualitative research has suggested mothers of children born within this gestational window felt less aligned with the "preterm" label (Laverty et al., 2023). It is possible that social media recruitment using the preterm label with this sub-category of prematurity may therefore not adequately capture the audience of interest. Additionally, from conversations with families participating in this research, the impact of the covid pandemic resulted in many children not having health visitor appointments. The timing of recruitment, in addition to reliance purely on a social media strategy may have encouraged families with concerns about their child's development to be motivated to engage in this research. This has implications for external validity of both MLP and FT samples included in this research.

Strengths and limitations

Lack of Conformity to ESCogS Sequence

In this study we did not assess inter-rater reliability and experimenters were not blind to study group which may have impacted performance. However, a stringent coding method was applied to score the ESCogS and this was completed from video recording which allowed for repeated viewing for any ambiguities during coding.

Measures of Social Preference

Despite the use of more, well validated paradigms of social preference selected for this study, it is possible that these videos still did not represent real life social interactions and thus may have failed to capture the realities of children's social attentional preferences. The inconsistences in associations between implicit and explicit measures of social cognition found in the literature and lack of correlation found here may be supportive of this. Future studies should aim to overcome these limitations.

Measure of Explicit social cognition (ESCogS)

It is possible that other variables (e.g. – motor skills, attention, executive functioning, cognitive ability) may have influenced performance on tasks included in the ESCogS and thereby have an impact on reliability and validity. The MLP sample here presented with significant differences in cognitive ability in comparison to the full-term control group, however only the early learning composite score on the MSEL was reported, and sub-domains, such as the fine motor subscale were not explored. Measuring and reporting sub-domains of general cognitive ability will likely be important in future studies to better understand variables which may impact performance on explicit social cognitive tasks.

Representativeness of our Sample

Our sample may have limited external validity for a number of reasons**.** Participants include children who completed assessments before and after the covid pandemic. There are

emerging studies investigating the impact of the changes in daily life and socialisation of children during the pandemic, which may be an important confound to consider.

Sample Size

Pilot studies can provide us with an estimate of the range of possible associations and differences between groups, in addition to assessment of recruitment rates and protocol feasibility (Lee et al., 2014). However, effect size estimates from pilot studies are unstable and do not provide a usual estimation for power calculations. Therefore, those presented here must be interpreted with caution. The use of small sample pilot data to conduct power calculations for future larger-scale studies, as done here is problematic and not recommended due to the imprecision inherent in small samples (Leon et al., 2011). Lee et al., (2010) recommend pilot studies should focus on descriptive statistics and estimation of protocol feasibility therefore we should remain cautious of the effect size estimates presented. Instead, the value of the pilot data presented here, should focus on feasibility and tolerability of the protocol.

Clinical Implication and Considerations

Experimental means for measuring subdomains of social cognition should be treated with caution. The lack of evidence in the wider literature, and that found here, weakens the assumption that social preference can be useful on its own as a marker of divergent social cognition. The GeoPref in particular is suggested as a promising prognostic tool for autism (Bacon et al., 2020; Moore et al., 2018), yet little evidence exists to support a strong correlation with real-world social behaviours that would likely be different in autistic people. Despite differences in social preference found between autistic and typically developing children, the use of the GeoPref as a prognostic marker for divergent social cognition may not be indicated at this time. Therefore, more traditional observational approaches to measurement of social cognition *in addition to* early measures of social preference may have added value.

. Evidence presented here suggests MLP children may not preferentially attend to social stimuli compared to FT children. This is a feature present in autistic children and therefore may support hypotheses suggesting atypical social developmental trajectory of children born within later gestational windows. Additional neurodevelopmental follow up of this MLP group may be beneficial for this group when supported by further research. This is in line with NICE guidance (NICE, 2017) which explains the increased developmental risks and extra support children born preterm (before 37 weeks GA) may require.

Recommendations for Future Studies

The power analysis conducted following this pilot study suggested a very large sample size would be necessary to evidence a statistically significant effect of the correlation between implicit and explicit measures of social cognition used here (*N*=182). Correlations were very small suggesting minimal associations; therefore, a fully powered study may not add additional value, notwithstanding the lack of practicality in recruiting the required number of participants.

Exploring differences in social preference between MLP and FT children may be beneficial, given that estimated sample sizes for a fully powered study are likely feasible (total *N*=52 for GeoPref; total *N=102* for BioMotion (C1), if comments about recruitment are taken into consideration, particularly for the MLP group.

Longitudinal designs, spanning multiple time points across development, may also be important to explore how early social preference impacts later development. There is preliminary evidence of this with studies finding preference for non-social images on the GeoPref at 23 months predicting greater symptom severity on ADOS and other self-report measures at age 8 (Bacon et al., 2020). In line with a developmental cascades model (Bradshaw, 2022), it is likely that implicit and explicit measures of social cognition may not correlate when measuring cross-sectionally. And perhaps early social preference impacts *later* social-cognitive skills.

Early social attention is important for social development as suggested by the social motivation hypothesis (Chevalier et al., 2012; Dawson, 2008; Kohls et al., 2012). Studies have suggested early differences in children born prematurely as compared to their FT peers, however these differences disappear by the age of five (Dean et al., 2020). Longitudinal designs, spanning multiple time points across development, when exploring differences between MLP and FT children is likely of particular importance to better understand potential pathways to different developmental trajectories between groups. When exploring differences, it may also be important to measure, report and analyse other variables which may impact performance on measures of explicit social cognition.

Conclusions

In summary, this pilot study has shown preliminary evidence of the following: large correlations between implicit and explicit social cognition are unlikely when measured cross-sectionally *and* important differences may exist in social preference between MLP and FT children.

References

Albert, W. B., & Tullis, T. S (2023) in Chapter 7 - Eye Tracking, Editor(s): Albert, W, B., & Tullis, T. S. In Interactive Technologies, Measuring the User Experience (Third Edition), Morgan Kaufmann, 177-193. Doi: 10.1016/B978-0-12-818080-8.00007-8.

Annaz, D., Campbell, R., Coleman, M., Milne, E., & Swettenham, J. (2012). Young children with autism spectrum disorder do not preferentially attend to biological motion, *Journal of Autism & Developmental Disorders,* 42, 401-408. Doi: 10.1007/s10803-011-1256-3.

Bacon, E. C., Moore, A., Lee, Q., Carter Barnes, C., Courchesne, E., & Pierce, K. (2020). Identifying prognostic markers in autism spectrum disorder using eye tracking, *Autism,* 24 (3): 658-669. Doi: 10.1177/1362361319878578.

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game, *Developmental Science, 8*(6), 492–499. Doi: 10.1111/j.1467-7687.2005.00440.

Bellagamba, F., & Tomasello, M. (1999). Re-enacting intended acts: Comparing 12- and 18-month-olds. *Infant Behaviour and Development, 22*(2), 277–282. Doi: /10.1016/S0163-6383(99)00002-8.

Blencowe, H., Cousens, S., Oestergaard, M. Z., Chou, D., Moller, A. B., Narwal, R., Adler, A., Vera Garcia, C., Rohde, S., Say, L., & Lawn, J. E. (2012). National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *Lancet,* 379:2162-2172. Doi: 10.1016/S0140-6736(12)60820-4.

Blencowe, H., Cousens, S., Chou, D., Oestergaard, M., Say, L., Moller, A. B., Kinney, M., & Joy, L. (2013). Born too soon: The global epidemiology of 15 million preterm births,

Reproductive Health, 10 (Suppl 1): S2. www.reproductive-health-journal.com/content/10/S1/S2.

Boyle, E., Johnson, S., Manktelow, B., Seaton, S. E., Draper, E. S., Smith, L. K., Dorling, J., Marlow, N., Petrou, S., & Field, D. J. (2015). Neonatal outcomes and delivery of care for infants born late preterm or moderately preterm: a prospective population-based study, *Archives of Disease in Childhood. Fetal & Neonatal Edition,* 100: F479-F485. Doi: 10.1136/archdischild-2014-307347.

Boyle, E. M., Poulsen, G., Field, D. J., Kurinczuk, J., Wolke, D., Alfirevic, Z., & Quigley, M. A. (2012). Effects of gestational age at birth on health outcomes at 3 and 5 years of age: population based cohort study, *British Medical Journal,* 344, 7848, 17. Doi: 10.1136/bmj.e896.

Bradshaw, J., Schwichtenberg, A. J., & Iverson, J. M. (2022). Capturing the complexity of autism: Applying a developmental cascades framework, *Child Development Perspectives,* 16: 18-26. Doi: 10.1111/cdep.12439.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16* (4), 231–239. Doi: 10.1016/J.TICS.2012.02.007.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*, 37–46. Doi: 10.1177/001316446002000104.

Davis, R., Donati, G., Finnegan, K., Boardman, J., Dean, B., Fletcher-Watson, S., & Forrester, G. S. (2022). Social gaze in preterm infants may act as an early indicator of atypical lateralization, *Child Development,* 93: 869-880. Doi: 10.1111/cdev.13734.

Dawson, G. (2008). Early behavioural intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and Psychopathology*, *20*(03), 775–803. Doi: 10.1017/ S0954579408000370.

Dean, B, M., Ginnell, L., Boardman, J.P., & Fletcher-Watson, S (2021). Social cognition following preterm birth: A systematic review, *Neuroscience & Biobehavioural Reviews,* 124, 151-167. Doi: 10.1016/j.neubiorev.2021.01.006.

Dean, B., Ginnell, L., Ledsham, V., Tsanas, A., Telford, E., Sparrow, S., Fletcher-Watson, S., & Boardman, J. P (2021). Eye-tracking for longitudinal assessment of social cognition in children born preterm, *The Journal of Child Psychology & Psychiatry,* 62: 4, 470-480. Doi:10.1111/jcpp.13304.

De Groote, I., Roeyers, H., & Warreyn, P. (2006). Social-communicative abilities in young high-risk preterm children, *Journal of Developmental & Physical Disabilities,* 18: 2, 183-200. Doi: 10.1007/s10882-006-9011-y.

De Jong, M., Verhoeven, M., & Van Baar, A. L. (2012). School outcome, cognitive functioning, and behaviour problems in moderate and late preterm children and adults: A review, *Seminars in Fetal & Neonatal Medicine,* 17, 163-169. Doi:10.1016/j.siny.2012.02.003.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vision Res.* 36, 1827–1837. Doi: 10.1016/0042-6989(95)00294-4

De Schuymer, L., De Groote, I., Beyers, W., Striano, T., & Roeyers, H. (2011). Preverbal skills as mediators for language outcome in preterm and full-term children, *Early Human Development,* 87, 265-272. Doi:10.1016/j.earlhumdev.2011.01.029.

Dubey, I., Brett, S., Ruta, L., Bishain, R., Chandran, S., Bhavnani, S., Belmonte, M. K., Estrin, G. L., Johnson, M., Gliga, T., & Chakrabarti, B. (2022). Quantifying social preference for social stmilui in young children using two tasks on a mobile platform, *PLOS One,* 17 (6): e0265587. Doi: 10.1371/journal.pone.0265587.

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2017). Beyond eye gaze: What else can eye tracking reveal about cognition and cognitive development? Developmental Cognitive Neuroscience. Doi: 10.1016/j.dcn.2016.11.001

Ellis, K., Lewington, P., Powis, L., Oliver, C., Waite, J., Heald, M., & Crawford, H. (2020). Scaling of Early Social Cognitive Skills in Typically Developing Infants and Children with Autism Spectrum Disorder, *Journal of Autism and Developmental Disorders,* 50 (11), 3988–4000. Doi: 10.1007/s10803-020-04449-9.

Ellis, K., Moss, J., Stefanidou, C., Oliver, C., & Apperly, I. (2021). The development of early social cognitive skills in neurogenetic syndromes associated with autism: Cornelia de Lange, fragile X and Rubinstein–Taybi syndromes. *Orphanet Journal of Rare Diseases,* 16 (1), 1–12. Doi: 10.1186/s13023-021-02117-4.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments & Computers, 28*(1), 1–11. Doi: 10.3758/BF03203630.

Ernst, M., & Paulus, M. P. (2005). Neurobiology of decision making: a selective review from a neurocognitive and clinical perspective. *Biological Psychiatry* 58, 597–604. Doi: 10.1016/j.biopsych.2005.06.004.

Falck-Ytter, T., Fernell, E., Gillberg, C., & von Hofsten, C. (2010). Face scanning distinguishes social from communication impairments in autism. *Developmental Science*, *13*, 864–875. Doi: 10.1111/j.1467-7687-2009.00942.

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

Fletcher-Watson, S., & Hampton, S. (2018). The potential of eye-tracking as a sensitive measure of behavioural change in response to intervention. Scientific Reports, 8(1). Doi: 10.1038/s41598-018-32444-9.

Ford, R. M., Griffiths, S., Neulinger, K., Andrews, G., Shum, D. H. K., & Gray, P. H. (2016). Impaired prospective memory but in tack episodic memory in intellectually average 7- to 9-year-olds born very preterm and/or very low birth weight, *Child Neuropsychology,* 23: 8, 954-979. Doi: 10.1080/09297049.2016.1216091.

Fu, X., Hung, A., De Silva, A. D., Busch, T., Mattson, W. J., Hoskinson, K. R., Hudson, G. T., & Nelson, E. E. (2022). Development of the mentalizing network structures and theory of mind in extremely preterm youth, *Social Cognitive & Affective Neuroscience,* 17, 977-985. Doi:10.1093/scan/nsac027.

Gueron-Sela, N., Atzaba-Poria., Meiri, G., & Marks, K. (2015). The caregiving environment and developmental outcomes of preterm infants: Diathesis stress or differential susceptibility effects? *Child Development,* 86: 4, 1014-1030. Doi: 0009-3920/2015/8604-0003.

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2006). The Autism Diagnostic Observation Schedule (ADOS). *Journal of Autism and Developmental Disorders*. Doi: 10.1007/s10803-006-0280-1.

Green, B. F. (1956). A method of scalogram analysis using summary statistics, *Psychometrika,* 21 (1). Doi: 10.1007/BF02289088.

Guo, B. Q., Li, H. B., Zhai, D. S., & Yang, L. Q. (2022). Prevalence of autism spectrum disorder by birth wright, gestational age, and size for gestational age: a systematic review, meta-analysis, and meta-regression, *European Child & Adolescent Psychiatry.* Doi: 10.1007/s00787-022-02078-4.

Guy, A., Seaton, S. E., Boyle, E. M., Draper, E. S., Field, D. J., Manktelow, B. N., Marlow, N., Smith, L. K., & Johnson, S. (2015). Infants born late/moderately preterm are at

increased risk for a positive autism screen at 2 years of age, *The Journal of Pediatrics,* 269-265.e3. Doi: 10.1016/j.jpeds.2014.10.053.

Hessels, R. S., & Hooge, I. T. C. (2019). Eye tracking in developmental cognitive neuroscience – The good, the bad and the ugly. Developmental Cognitive Neuroscience, 40. Doi: 10.1016/j.dcn.2019.100710.

Hintze, J. M., & Matthews, W. J. (2004). The Generalizability of Systematic Direct Observations Across Time and Setting: A Preliminary Investigation of the Psychometrics of Behavioral Observation, *School Psychology Review,* 33(2), 258–270. Doi: 10.1080/02796015.2004.12086247.

Hornman, J., Winter, A. F., Kerstjens, J. M., Bod, A. F., & Reijneveld, S. A. (2017). Stability to developmental problems after school entry in moderately late preterm and early preterm-born infants, *The Journal of Pediatrics,* 187: 73-79. Doi: 10.1016/j.jpeds.2017.05.022.

Hung, M., Bounsanga, J., & Wright-Vos, M. (2017). Interpretation of correlations in clinical research, *Postgraduate Medicine,* 128 (8): 902-906. Doi: 10.1080/00325481.2017.1382820.

Imafuku, M., Kawai, M., Niwa, F., Shinya, Y., & Myowa, M. (2020). Longitudinal assessment of social attention in preterm and term infants: Its relation to social communication and language outcome, *Infancy,* 26: 617-634. Doi: 10.1111/infa.12402.

Johnson, S., & Marlow, N. (2017). Early and long-term outcome of infants born extremely preterm, *Archives of Disease in Childhood,* 102: 97-102. Doi: 10.1136/archdischild-2015-309581.

Jones, W., Carr, K., & Klin, A. (2008). Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism

spectrum disorder. Archives of General Psychiatry, 65, 946–954. Doi: 10.1001/rchpsyc.65.8.946.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics,* 14 (2), 201–211. Doi.org/10.3758/BF03212378.

Johnson, S.., Waheed, G., Manktelow, B. N., Field, D. J., Marlow, N., Draper, E. S., & Boyle, E. M. (2018). Differentiating the preterm phentotype: Distinct profiles of cognitive and behavioural development following late and moderately preterm birth, *The Journal of Pediatrics,* 193: 85-92. Doi: 10.1016/j.peds.2017.10.002.

Jones, K. M., Champion, P. R., & Woodward, L. J. (2013). Social competence of preschool children born very preterm, *Early Human Development,* 89, 795-802. Doi: 10.1016/j.earlhumdev.2013.06.008.

Kerstjens, J. M., De Winter, A. F., Bocca-Tjeertes, I. F., Vergert, E. M. J., Reijneveld, S. A., & Bos, A. F. (2010). Developmental delay in moderately preterm-born children at school entry, *The Journal of Pediatrics,* 159: 92098. Doi: 10.1016/j.jpeds.2010.12.041.

Kim, S. H., Joseph, R. M., Frazier, J. A., O'Shea, T. M., Chawarska, K., Allred, E. N., & Kuban, K. K. (2016). Predictive Validity of the Modified Checklist for Autism in Toddlers (M-CHAT) Born Very Preterm, *Journal of Pediatrics,* 178, 101-107.e2. Doi: 10.1016/j.jpeds.2016.07.052.

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*, 809–816. Doi: 10.1001/archpsyc.59.9.809.

Kohls, G., Chevallier, C., Troiani, V., & Schultz, R. T. (2012). Social wanting dysfunction in autism: neurobiological underpinnings and treatment implications. *Journal of Neurodevelopmental. Disorders,* 4:1. Doi: 10.1186/1866-1955-4-10.

Laverty. C., Surtees, A., O' Sullivan, R., Sutherland, D., Jones, C., & Richards, C. (2021). The prevalence and profile of autism in individuals born preterm: a systematic review and meta-analysis, *Journal of Neurodevelopmental Disorders,* 13: 41. Doi: 10.1186/s11689-021-09382-1.

Laverty. C., Surtees, A., O' Sullivan, R., Sutherland, D., & Richards, C. (2023). A qualitative interview with mothers of moderately or late preterm infants, where are the care gaps? *In Prep.*

Leon, A.C., Davis, L.L., & Kraemer, H.C. (2011). The role and interpretation of pilot studies in clinical research, *Journal of Psychiatric Research,* 45 (5): 626-629. DOI: 10.1016/j.jpsychires.2010.10.008.

Longa, L. D., Nosarti, C., & Farroni, T. (2022). Emotion recognition in preterm and full-term school-age children, *International Journal of Environmental Research & Public Health,* 19, 6507. Doi:10.3390/ijerph19116507.

MacKay, D.F., Smith, G. C., Dobbie, R., & Pell, J. P. (2010). Gestational age at delivery and special educational need: retrospective cohort study of 407,503 schoolchildren, *PLOS Medicine,* 7 (6): e1000289. Doi: 10.1371/journal.pmed.1000289.

Marleau, I., Vona, M., Gagner, C., Luu, T. M., & Beauchamp, M. H. (2020). Social cognition, adaptive functioning, and behaviour problems in pre-schoolers born extremely preterm, *Child Neuropsychology,* Doi: 10.1080/09297049.2020.1797656.

Marotta, A., Rea, M., Ferri, R., & Casagrande, M. (2020). Investigating socio-emotional cognition in late preterm children: A case-control study, *European Journal of Developmental Psychology,* 17:3, 365-378, Doi: 10.1080/17405629.2019.1637339.

Mason, L., Shic, F., Falck-Ytter, F., Chakrabarti, B., Charman, T., T., Loth, E., Tillman, J., Banaschewski, T., Baren-Cohen, S., Bolte, S., Buitelaar, J., Durston, S., Oranje, B., Persico, A. M., Beckmann, C., Bougeron, T., Dell'Acqua, F., Ecker, C., Moessnang,

C., Murphy, D., Johnson, M. H., Jones, E. J. H., & LEAP team (2021). Preference for biological motion is reduced in ASD: implications for clinical trials and the search for biomarkers, *Molecular Autism,* 12 (74), 1-13. Doi: 10.1186/s13229-021-00476-0.

Mateus, V., Osorio, A., & Martins, C. (2019). Effects of prematurity on infants' joint attention abilities: A meta-analytic study, *Infant & Child Development,* 28: e2145. Doi: 10.1002/icd.2145.

Mateus, V., Vieira, E. D. A. P., Martins, C., Pachi, P. R., & Osorio, A. (2020). Joint attention abilities in Brazilian preterm and full-term infants, *Infant Behaviour & Development,* 60, 101-451. Doi: 10.1016/j.infbeh.2020.101451.

Montagna, A., & Nosarti, C. (2016). Socio-emotional development following very preterm birth: Pathways to pathology, *Frontiers in Psychology,* 7: 80. Doi: 10.3389/fpsyg.2016.00080.

Moore, A., Wozniak, M., Yousef, A., Carter Barnes, C., Cha, D., Courchesne, E., & Pierce, K (2018), *Molecular Autism,* 9 (19), 1-13. Doi: 10.1186/s13229-018-0202-z.

Moll, H., & Tomasello, M. (2007). How 14- and 18-month-olds know what others have experienced. *Developmental Psychology, 43*(2), 309–317. Doi: 10.1037/0012-1649.43.2.309.

Morgan, J. C., & Boyle, E. M. (2017). The late preterm infant, *Paediatrics & Child Health,* 28: 1. Doi: 10.1016/j.paed.2017.10.003.

Moster, D., Wilcox, A.J., Vollset, S. E., Markestad, T., & Lie, R.T. (2010). Cerebral palsy among term and post term births, *The Journal of the Medical Association,* 304 (9): 976–82. Doi: 10.1001/jama.2010.1271.

Moser, K., Macfarlane, A., Chow, Y. H., Hilder, L., & Dattani, N. (2007). Introducing new data on gestation-specific infant mortality among babies born in 2005 in England and Wales, *Health Statistics Quarterly,* (35):13–27. PMID: 17894197.

Mossad, S. I., Smith, M. L., Pang, E. W., & Taylor, M. J. (2017). Neural correlates of "theory of mind" in very preterm born children, *Human Brain Mapping,* 38: 5577-5589. Doi: 10.1002/hbm.23750.

Mossad, S. I., Young, J. M., Wong, S. M., Dunkley, B. T., Hunt, B. A. E., Pang, E. W., & Taylor, M. J. (2021). Very preterm brain at rest: longitudinal social-cognitive network connectivity during childhood, *Social Cognitive & Affective Neuroscience,* 00, 1-10. Doi: 10.1093/scan/nsab110.

Mossad, S. I., Vandewouw, M. M., Smith, M. L., & Taylor, M. J. (2021). The preterm social brain: altered functional networks for theory of mind in very preterm children, *Brain Communications,* 1: 1-13. Doi: 10.1093/braincomms/fcaa237.

Mullen, E. M. (1995). *Mullen Scales of Early Learning* (AGS ed.). Circle Pines, MN: American Guidance Service Inc.

Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., & Matsuzawa T (2003). Preference for human direct gaze in infant chimpanzees (Pan troglodytes*), Cognition,* ; 89: 113–124. Doi: 10.1016/S0010-0277(03)00071-4.

Oktay-Gur, N., Schulz, A., & Rakoczy, H. (2018). Children exhibit different performance patterns in explicit and implicit theory of mind tasks, *Cognition,* 73, 60-74. Doi: 10.1016/j.cognition.2018.01.001.

Olafsen, K. S., Ronning, K. A., Kaaresen, P. I., Ulvund, S. E., Handegard, B. H., & Dahl, L. B. (2006). Joint attention in term and preterm infants at 12 months corrected age: The significance of gender and intervention based on a randomized controlled trial, *Infant Behaviour & Development,* 29, 554-563. Doi:10.1016/j.infbeh.2006.07.004

Pierce, K., Contant, D., Hazin, R., Stoner, S., & Desmond, J (2011). Preference for geometric patterns early in life as a risk factor for autism, *Archives of General Psychiatry,* 68 (1), 101-109. Doi:10.1001/archgenpsychiatry.2010.113.

Pierce, K., Marinero, S., Hazin, R., McKenna, B., Carter Barnes, C., & Malige, A. (2016). Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder subtype associated with increased symptom severity, *Biological Psychiatry,* 79 (8), 657–66. Doi: 10.1016/j.biopsych.2015.03.032.

Pisula, E., & Ziegart-Sadowska, K (2015). Broader autism phenotype in siblings of children with ASD-A review, *International Journal of Molecular Sciences,* 16, 13217-13258; Doi:10.3390/ijms160613217.

Quigley, M. A., Poulsen, G., Boyle, E., Wolke, D., Field, D., Alfirevic, Z., & Kurinczuk, J. J. (2011). Early term and late preterm birth are associated with poorer school performance at age 5 years: a cohort study, *Archives of Disease in Childhood: Fetal & Neonatal Edition,* 97: F167-F173. Doi:10.1136/archdischild-2011-300888F167.

Rijn, A. V., Urbanus, E., & Swaab, H. (2018). Eye tracking measures of social attention in young children: How gaze patterns translate to real-life social behaviours, *Social Development,* 28:564-580. Doi: 10.1111/sode.12350.

Risko, E. F., Laidlaw, K. E. W., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity, *Frontiers in Human Neuroscience,* 6: 14. Doi: 10.3389/fnhum.2012.00143.

Robins, D. L., & Fein, D. (2018). Modified Checklist for Autism in Toddlers, Revised, with Follow- Up (M-CHAT-R/F), also M-CHAT. In Encyclopaedia of Clinical Neuropsychology (pp. 2253– 2258). Doi:10.1007/978-3-319-57111-9_1569.

Roldan-Tapia, M. D., Moreno-Rios, S., & Canovas-Lopez, R. (2017). Thinking about social and non-social alternative possibilities in premature pre-schoolers, Journal of Clinical and Experimental Neuropsychology, 39:8, 725-737. Doi: 10.1080/13803395.2016.1257703.

Rolschau, A. H., Olesen, A. W., Obel, C., Olsen, J., Wu, C. S., & Kofoed, P. E. (2020). Cerebral disorders in the first 7 years of life in children born post-term: a cohort study, *BMC Paediatrics,* 20:51. Doi: 10.1186/s12887-020-1950-4.

Sato, J., Safar, K., Vandewouw, M. M., Bando, N., O'Connor, D. L., Unger, S. L., & Taylor, M. J. (2021). Altered functional connectivity during face processing in children born with very low birth weight, *Social Cognitive & Affective Neuroscience,* 1182-1190. Doi: 10.1093/scan/nsab070.

Schietecatte, I., Roeyers, H., & Warreyn, P. (2012). Exploring the nature of joint attention impairments in young children with autism spectrum disorder: Associated social and cognitive skills, *Journal of Autism Developmental Disorders,* 42:1-12. Doi: 10.1007/s10803-011-1209-x.

Sekigawa-Hosozawa, M., Tanaka, K., Shimizu, T., Nakano, T., & Kitazawa, S. (2017). A group of very preterm children characterised by atypical gaze patterns, *Brain & Development,* 39: 218-224. Doi: 10.1016/j.braindev.2016.10.001.

Shimojo, S., Simion, C., & Changizi M. A (2010). Gaze and Preference-Orienting Behavior as a Somatic Precursor of Preference Decision. The Science of Social Vision. Oxford University Press;pp.151–163. Doi: 10.1093/acprof:0s0/97801953333176.003.0009.

Shimojo, S., Simion, C., Shimojo, E., & Scheier. C (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience,* 6: 1317–1322. Doi: 10.1038/nn1150.

Sifre, R., Olson, L., Gillespie, S., Klin, A., Jones, W., & Shultz, S. (2018). A longitudinal investigation of preferential attention to biological motion in 2-to 24-month-old infants, *Scientific Reports,* 8:2527. Doi: 10.1038/s41598-018-20808-0.

Sperotto, R. G. (2015). The influence of gestational age on social attention and language in the second year of life, Cardiff University, Unpublished thesis.

Speer, L. L., Cook, A. E., McMahon, W. M., & Clark, E. (2007). Face processing in children with autism: effects of stimulus contents and type. *Autism* 11, 265–277. Doi: 10.1177/1362361307076925.

Tammela, O. (2022). Neurodevelopmental impairments in childhood after moderate and late preterm birth in Finland: a summary of national register studies, *Medical Research Archives,* [Online]. 10 (10). Doi: 10.18103/mra.v10i10.3205.

Taylor, G. H. (2022). Neurodevelopmental origins of social competence in very preterm children, *Seminars in Fetal & Neonatal Medicine,* 25. 101108. Doi: 10.1016/j.siny.2020.101108.

Telford, E. J., Fletcher-Watson, S., Gillespie-Smith, K., Pataky, R., Sparrow, S., Murray, I. C., O'Hare, A & Boardman., J, P. (2016). Preterm birth is associated with atypical social orienting in infancy detected using eye tracking, *Journal of Child Psychology & Psychiatry,* 57: 7, 861-868. Doi:10.1111/jcpp.12546.

Tomasello, M. (2014). *A natural history of human thinking*. Cambridge: Harvard University Press.

Twilhaar, S. E., De Kieviet, J. F., Bergwerff, C. E., Finken, M. J. J., Elburg, R. M., & Oosterlaan, J. (2019). Social adjustment in adolescents born very preterm: Evidence for a cognitive basis of social problems, *The Journal of Pediatrics,* 213. Doi: 10.1016/j.jpeds.2019.06.045.

Uljarevic, M., Frazier, T. W., Jo, B., Phillips, J. M., Billingham, W., Cooper, M. N., & Hardan, A. Y. (2021). Relationship between social motivation in children with autisim spectrum disorder and their parents, *Frontiers in Neuroscience,* 15: 660330. Doi: 10.3389/fnins.2021.660330.

Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Development, 77*(3), 640–663. Doi: 10.1111/j.1467-8624.2006.00895.x.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science, 311*(5765), 1301–1303.Doi: 10.1126/science.1121448.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy, 11* (3), 271–294. Doi: 10.1111/j.1532-7078.2007.tb00227.

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

Wen, T. H., Cheng, A., Andreason, C., Zahiri, J., Xiao, Y., Xu, R., Bao, B., Courchesne, E., Barnes, C. C., Arias, S. J., & Pierce, K. (2022). Large scale validation of an early-age eye tracking biomarker of an autism spectrum disorder subtype, *Nature,* 12: 4253. Doi: 10.1038/s41598-022-08102-6.

WHO: WHO: recommended definitions, terminology and format for statistical tables related to the perinatal period and use of a new certificate for cause of perinatal deaths. (1977). Modifications recommended by FIGO as amended October 14, 1976, *Acta Obstetricia et Gynecologica Scandinavica*, 56:247-253. Doi: 10.3109/00016347709162009.

Williamson, K. E., & Jakobson, L. S. (2014). Social attribution skills of children born preterm at very low birth weight, *Developmental & Psychopathology,* 26, 889-900. Doi:10.1017/S0954579414000522.

Williamson, K. E., & Jakobson, L. S. (2014). Social perception in children born at very low birthweight and its relationship with social/behavioural outcomes, *The Journal of Child Psychology & Psychiatry,* 55: 9, 990-998. Doi:10.1111/jcpp.12210.

Williamson, K. E., Jakobson, L. S., Saunders, D. R., & Troje, N. F. (2015). Local and global aspects of biological motion perception in children born at very low birth weight, *Child Neuropsychology*, 21:5, 603-628. Doi: 10.1080/09297049.2014.945407.

Witt, A., Theurel, A., Tolsa, C. B., Lejeune, F., Fernandes, L., De Jonge, L. V. H., Monnier, M., Graz, M. B., Barisnikov, K., Gentaz, E., & Huppi, P. (2014). Emotional and effortful control abilities in 42-month-old very preterm and full-term children, *Early Human Development,* 90: 565-569. Doi: 10.1016/j.earlhumdev.2014.07.008.

Wolf, A., & Ueda, K. (2021). Contribution to eye-tracking to study cognitive impairments among clinical population, *Frontiers in Psychology,* 12, 590986. Doi: 10.3389/fpsyg.2021.590986.

Yamaoka, N., & Takada, S. (2016). Joint attention development in low-risk very low birth weight infants at around 18 months of age, *Kobe Journal of Medical Science,* 62: 4, E89-E98.

Yang, S., Platt, R. W., & Kramer, M.S. (2010). Variation in child cognitive ability by week of gestation among healthy term births, *The American Journal of Epidemiology,* 171(4):399–406. Doi: 10.1093/aje/kwp413.

Zivan, M., Morag, I., Yarmolovsky, J., & Geva, R. (2021). Hyper-reactivity to salience limits interaction among infants born preterm and infant siblings of children with ASD, *Frontiers in Psychiatry,* 12: 646838. Doi: 10.3389/fpsyt.2021.646838.

Zmyj, N., Witt, S., Weitkamper, A., Neumann, H., & Lucke, T. (2017). Social cognition in children born preterm: a perspective on future research directions. *Frontiers in psychology,* 8. 455. Doi: 10.3389/fpsyg.2017.00455.

Press Release: Meta-Analysis

Children born early find it more difficult to understand the emotions and mental state of others compared to children born on time!

Children who are born early (known as preterm) can suffer from difficulties in their development due to being born before the typical 37-week gestational period (known as full-term). In particular, research has suggested higher rates of autism and Attention Deficit Hyperactivity Disorder (ADHD) when compared to children born on time. Social and communication differences are important features of Autism and ADHD and children with these diagnoses often show differences in developing important social skills.

Research teams have begun to explore skills which are important for social development in children born preterm, to try to better understand any differences in development of these important social skills compared to children born at term. An increased interest in understanding social development in children born preterm, means there are many more studies looking at these important social skills than ever before.

The research team at the University of Birmingham used gold standard methodology to review *all* studies trying to understand differences in skills important for social development in children born early. Bringing all studies together in this way can help us get a clearer picture about true differences in social development in children born early compared to children born on time.

There were two important findings: Children born before 32 weeks gestational age (known as very preterm) have difficulty understanding another person's belief's thoughts and intentions and difficulty in identifying the emotions of other people when compared to children born full-term.

Recognising emotions and understanding beliefs and intentions are important skills for successful social development. Of particular interest in this research was the mixed findings of

research exploring earlier developing social skills important for children's social understanding. Skills such as the ability to follow an adult's gaze and jointly coordinating attention with another person are important skills that are essential for later developing social skills. Some studies suggested preterm children performed better than children born on time in these skills, whilst others suggested reduced abilities or no differences at all in comparison to children born on time.

We therefore can identify differences in these later developing skills such as theory of mind and emotion recognition, but we do not yet understand how the early development of these skills contributes to later skills important in developing social competence.

Lead researcher Emma Covernton puts forward two key recommendations for researchers in the field. She suggests we need to better understand earlier skills important for social understanding and studies which follow children over a period of time to understand how these skills develop in children born early. She also highlights that we know very little about children born closer to the normal gestational period. There can be many differences in the experiences of children born at different stages of prematurity. Those born very early tend to experience more medical complications as a result of less time in the womb. Whilst those born closer to term present with less medical complications, however some research has suggested that these children may also present with the same differences in social development as those born much earlier.

The results of this research offer interesting insights into differences in social understanding of preterm children when compared to full-term children. Identifying that children born earlier show differences in skills important for social understanding is an important discovery so we can better identify key avenues of support for optimal social development.

Press Release: Empirical Paper

Part One

Eye tracking technology and more traditional observational methods my not give us the information we think they are!

Human beings show a preference for social aspects of the world which is suggested to occur from the moment of birth. This idea is known as *social preference*, and it is suggested to play a key role in a child's social development. If a child finds social content more rewarding, they may be more likely to interact with other children and adults more, which offers lots of opportunity for developing important social skills. A reduced preference to social content is suggested to play a role in differences in social development and may be an early indicator of a child being on a different pathway of development than what is typical. This has been demonstrated in autistic children, where such children show a reduced preference for looking at social images. This knowledge is helpful and means research looking at whether children show a preference for social images over non-social images may be helpful in identifying children who may be at risk for differences in social development.

A research team at the University of Birmingham, in collaboration with Birmingham Baby Lab measured social preference with eye tracking technology and early social skills through direct play-based assessments in children ages 18 to 36 months. The research team were interested in better understanding how early social preference might be associated with real-life social skills. This research will help us to better understand the contribution of social preference to the development of social skills important for later social competence.

Findings showed little overlap in social preference measured by eye tracking and behavioural assessment. This is interesting as it may suggest that a preference for social aspects of our world may not be underlying development of real-life social skills which are measured in more traditional observational methods. This may mean that other characteristics are

important to consider in children who show different social developmental. Using measures of early social preference is important to understand these differences in social attention, however, continuing to use other assessment approaches in important to allow for a clearer picture of children's individual strengths and weaknesses.

Part Two

Eye tracking technology suggests children born early show differences in social attention compared to children born at term!

A research team at the University of Birmingham explored differences in social understanding in children born early, but closer to the normal 37-week gestational period and children born on time. This group are known as moderate to late preterm and includes children born between 32 weeks and 37 weeks gestational age. In collaboration with Birmingham Baby Lab, a novel pilot study explored differences in early social preference measured using eye tracking technology and early skills important for social development using play-based assessments.

This study was the first of its kind investigating children born at later gestational ages of prematurity in isolation from those born earlier. This is an important development to be able to better understand the social development of these children, outside of the medical complexity often experiences in children born at earlier gestational ages. Emma Covernton, a member of the research team says "these findings are the first to suggest important differences in earlier skills important for social understanding in this cohort of children. Exploring this further in a larger scale study are important to support the development of children born at later gestational ages of prematurity".

Appendix 1

Appendices showing PRISMA checklist.

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	yes
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	yes
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	yes
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	yes
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	yes
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	yes
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	yes
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	yes
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	yes
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	yes
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	yes
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	yes
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	yes
Synthesis	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and	yes

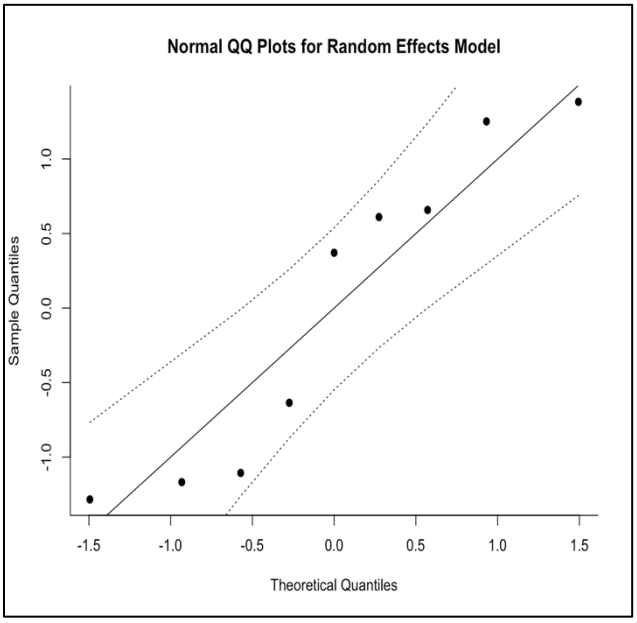
Section and Topic	Item #	Checklist item	Location where item is reported
methods		comparing against the planned groups for each synthesis (item #5)).	
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	yes
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	yes
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	yes
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	yes
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	yes
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	yes
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Check with A/C
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	yes
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	n/a
Study characteristics	17	Cite each included study and present its characteristics.	yes
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	yes
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	yes
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	yes
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	yes
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Yes
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Yes N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	yes
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Yes/N/A
DISCUSSION			

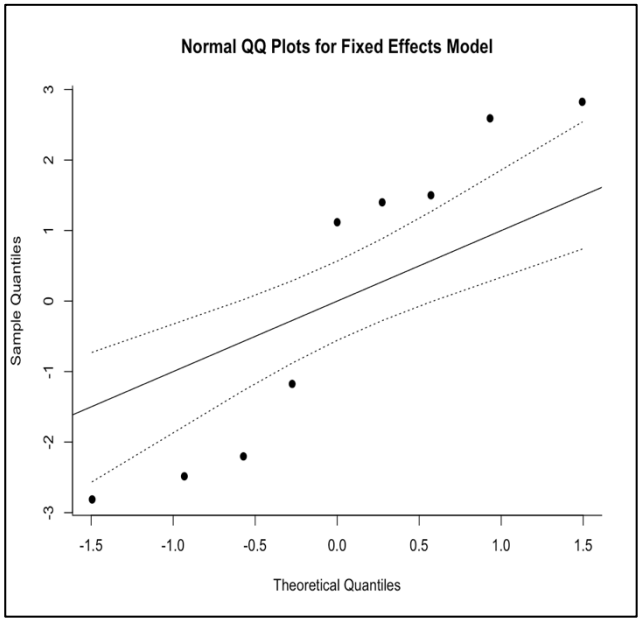
Section and Topic	Item #	Checklist item	Location where item is reported
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Yes
	23b	Discuss any limitations of the evidence included in the review.	Yes
	23c	Discuss any limitations of the review processes used.	yes
	23d	Discuss implications of the results for practice, policy, and future research.	yes
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	yes
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	yes
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	yes
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	To compete prior to publication if accepted by a journal
Competing interests	26	Declare any competing interests of review authors.	To compete prior to publication if accepted by a journal
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	To compete prior to publication if accepted by a journal

Appendix 2

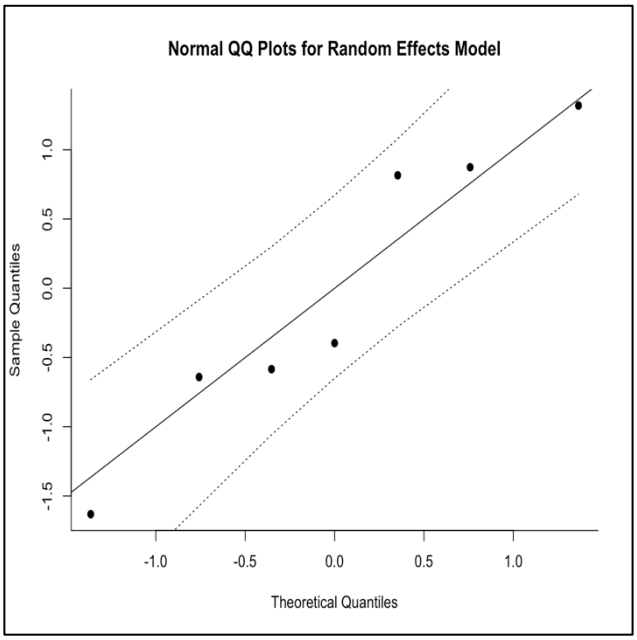
Appendix showing QQ plots for the distribution of standardised mean differences for both random and fixed effect models for studies measuring ***gaze following, RJA (point + gaze), initiating joint attention, theory of mind (after removal of Marleau et al., 2020) and affect recognition.***

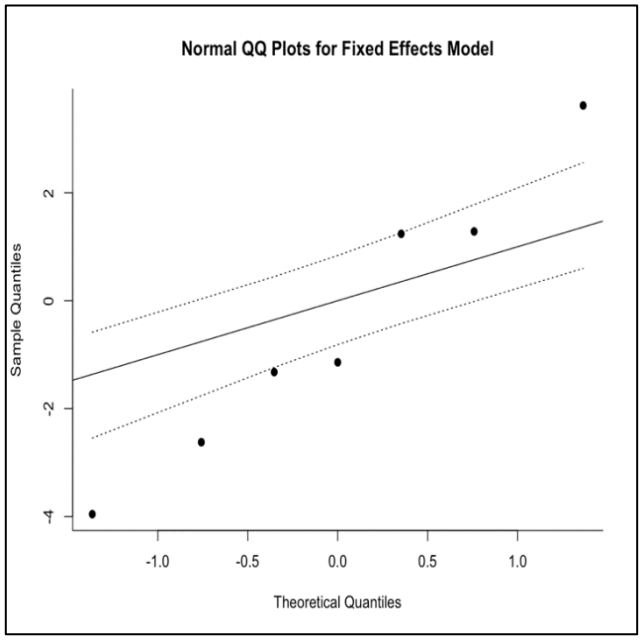
A. Gaze following



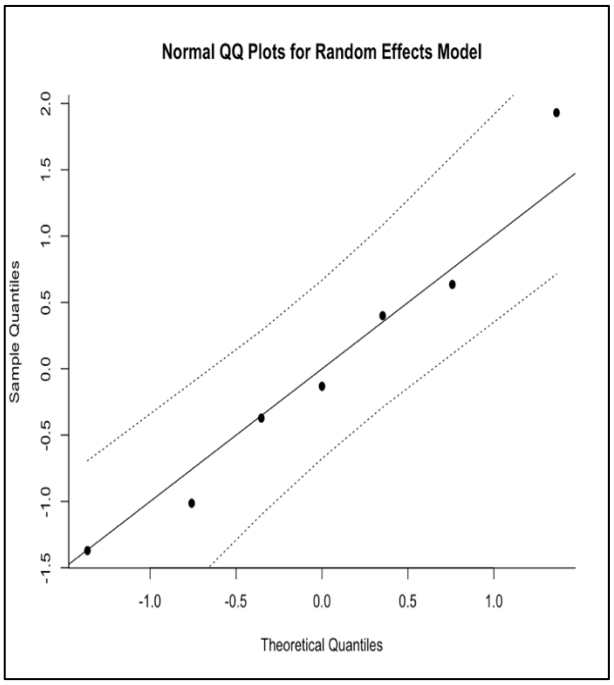


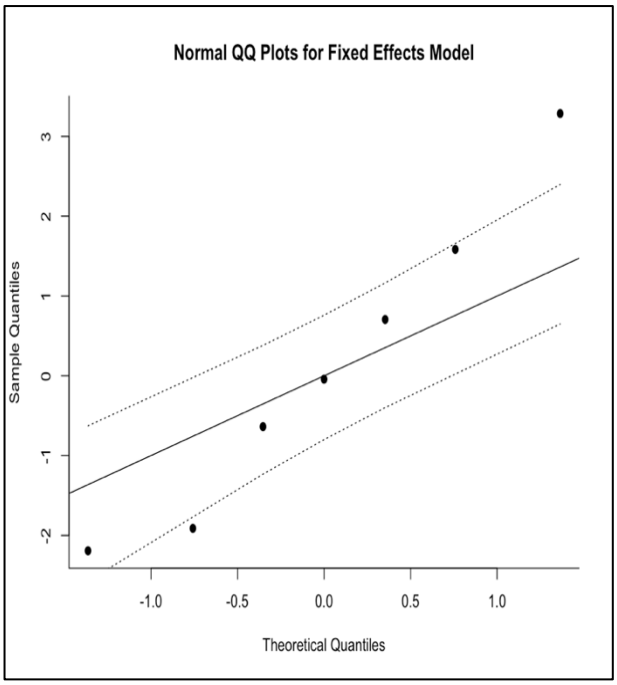
B. RJA (point + gaze)



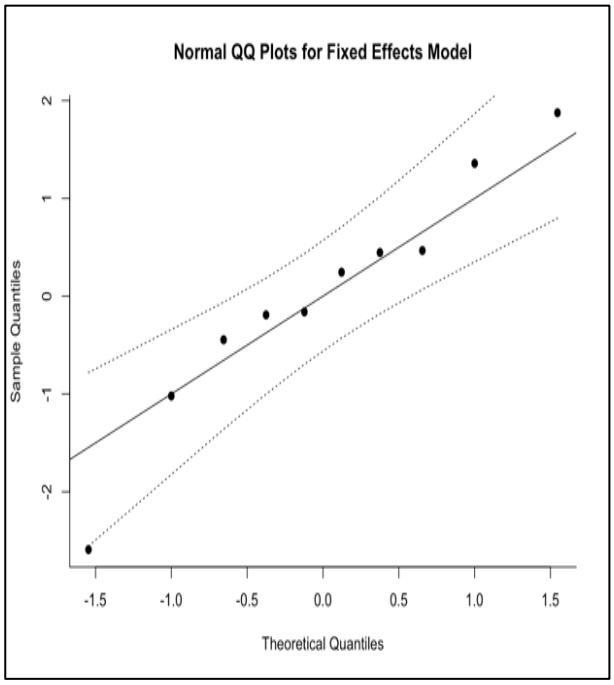


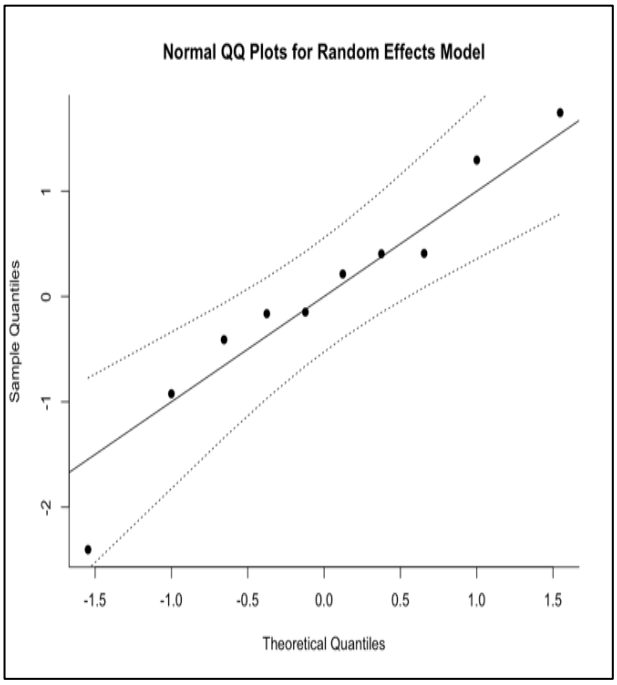
C. Initiating joint attention



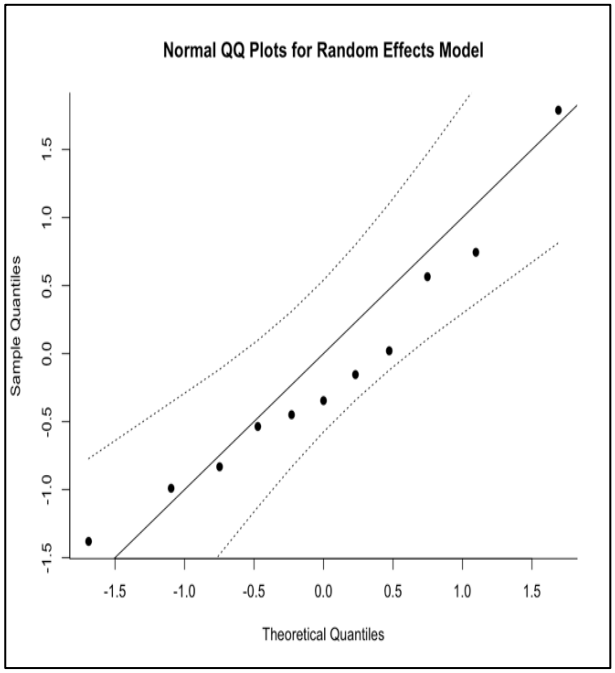


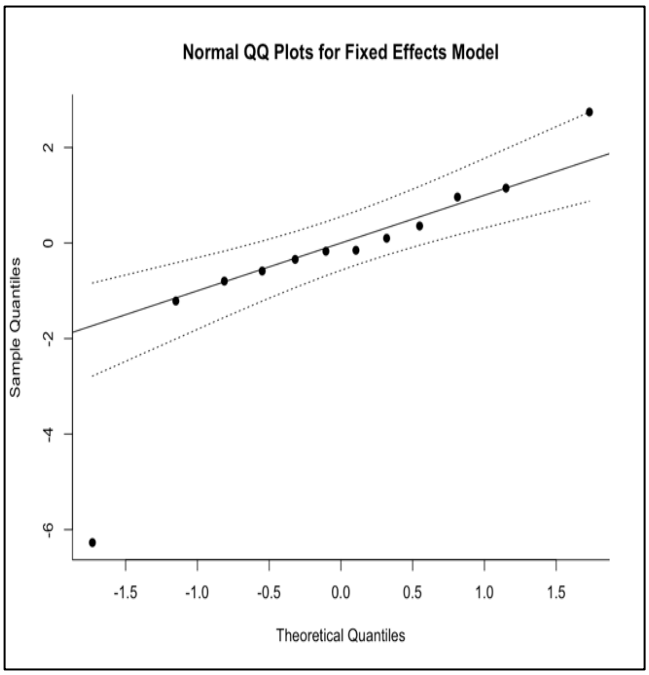
D. Affect recognition.





E. Theory of Mind

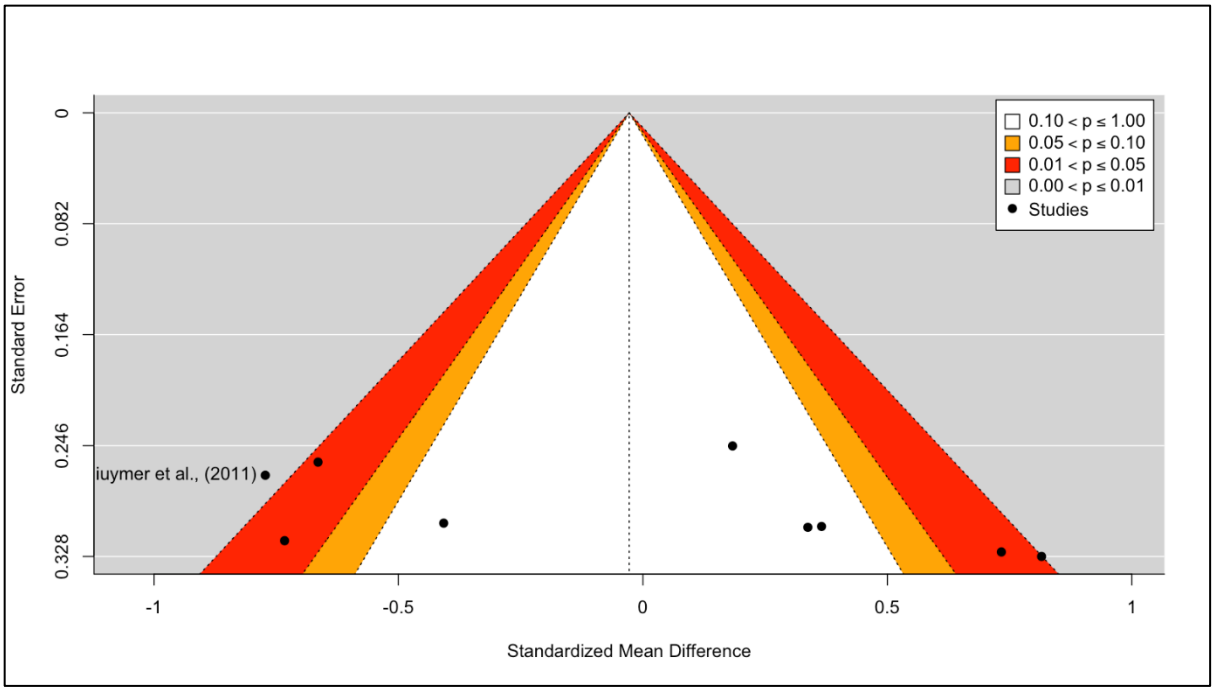




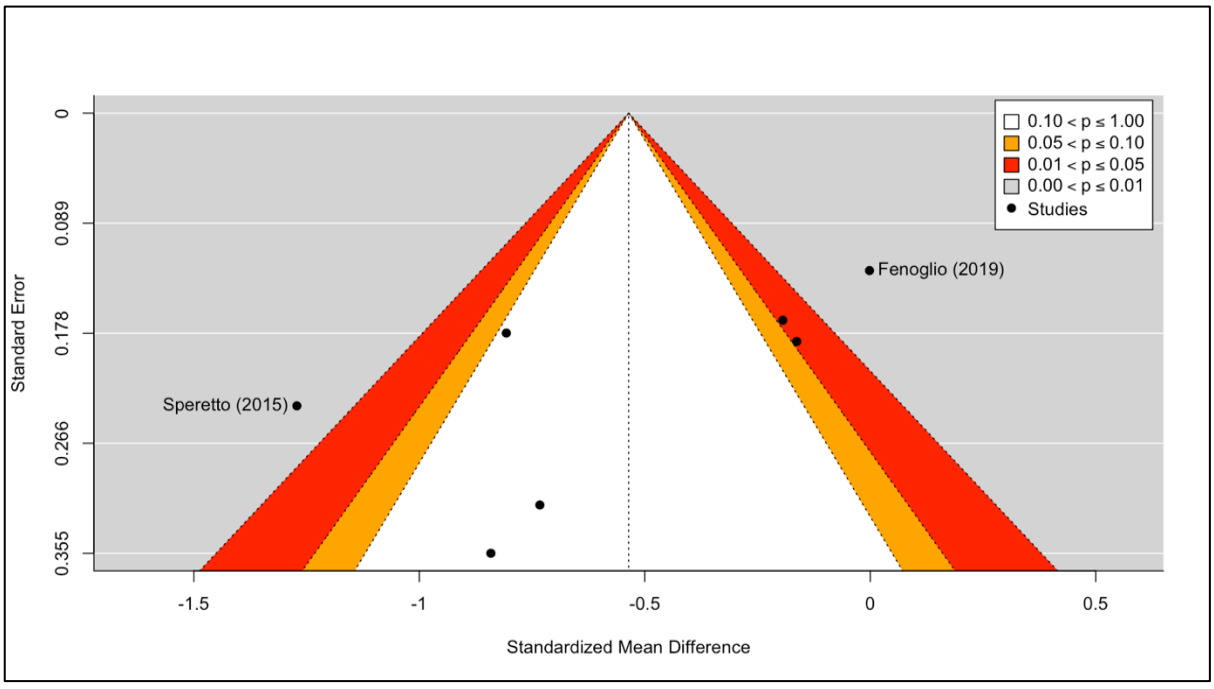
Appendix 3

Figure showing funnel plot of the SMD in gaze following, RJA (gaze + point), IJA, and affect recognition between preterm and full-term children. The 95% confidence interval of the expected distribution of effect is shown as an inverted "funnel".

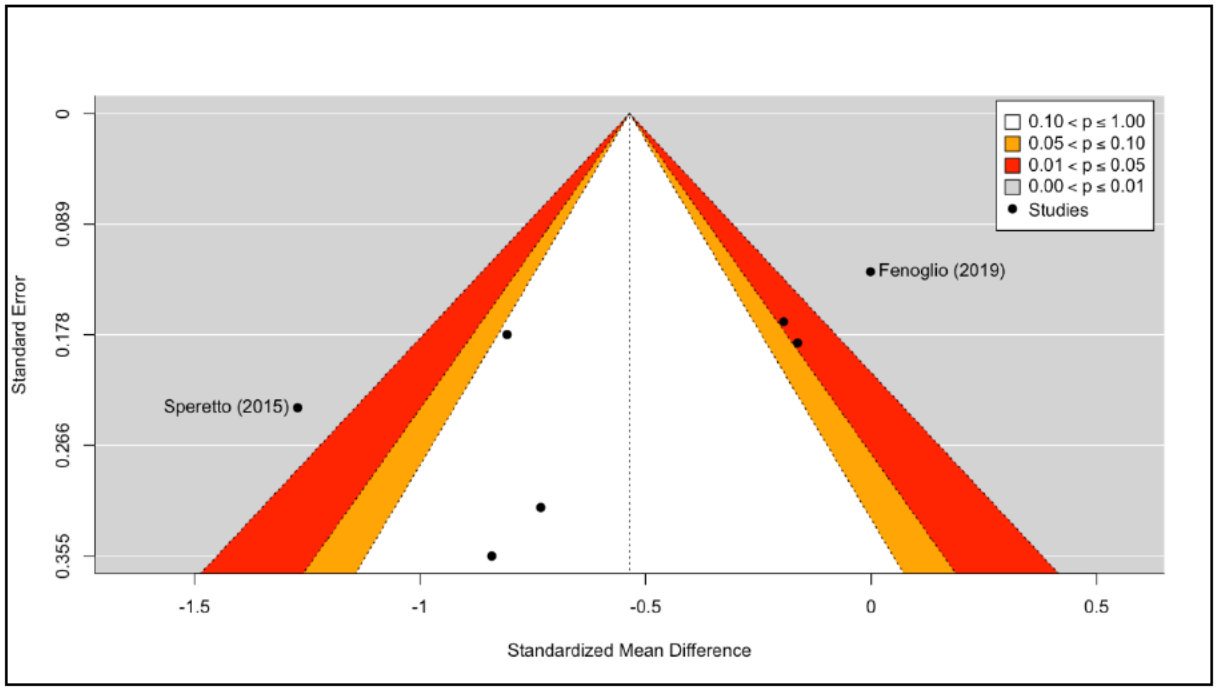
A. Gaze following



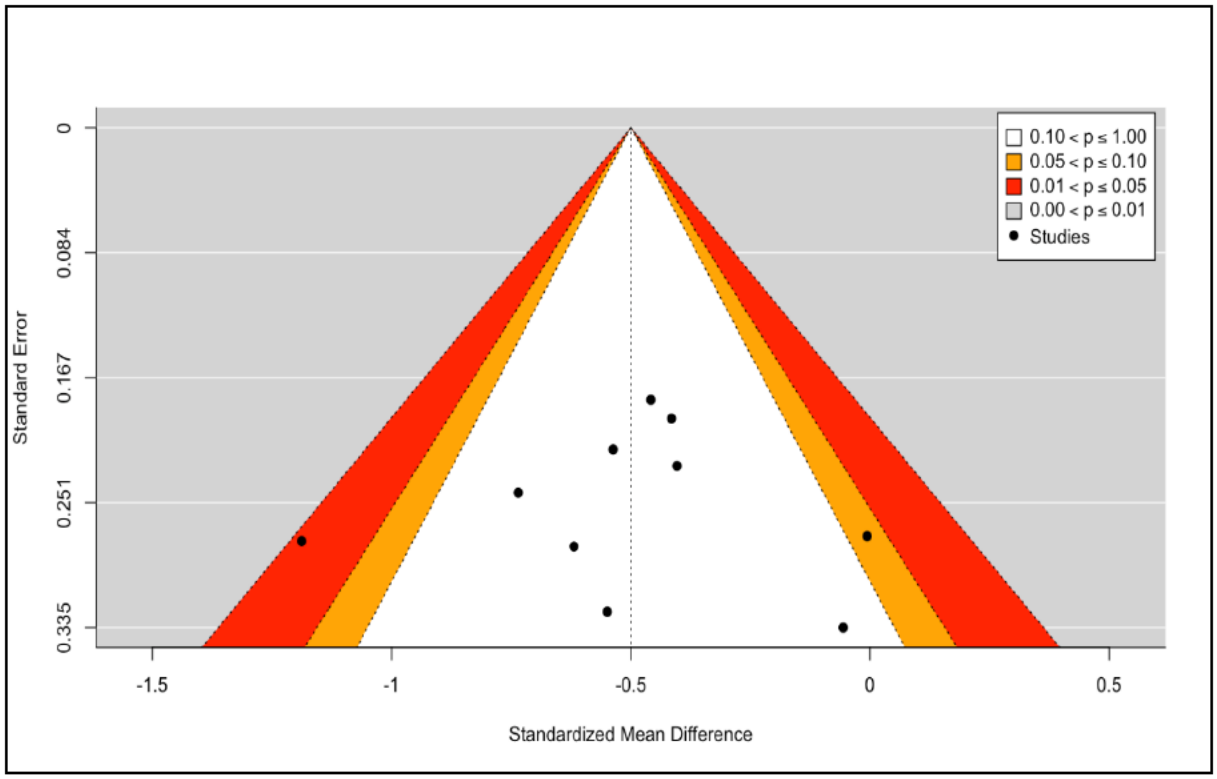
B. RJA (gaze + point)



C. Initiating joint attention



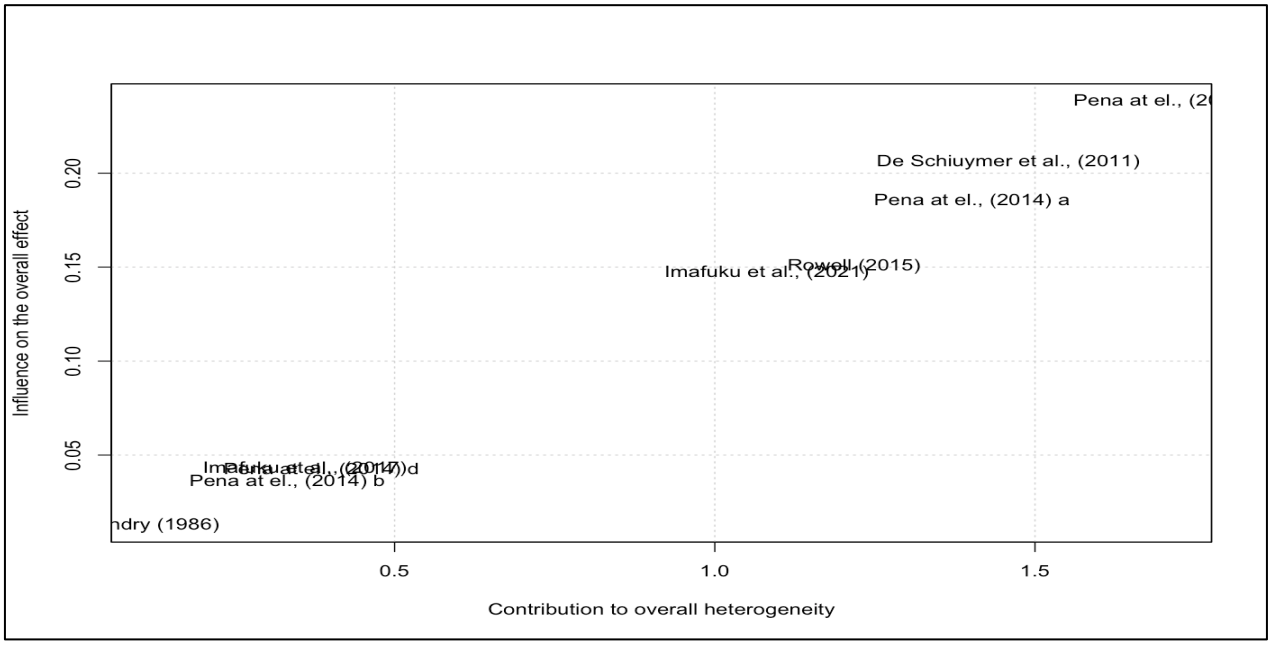
D. Affect recognition



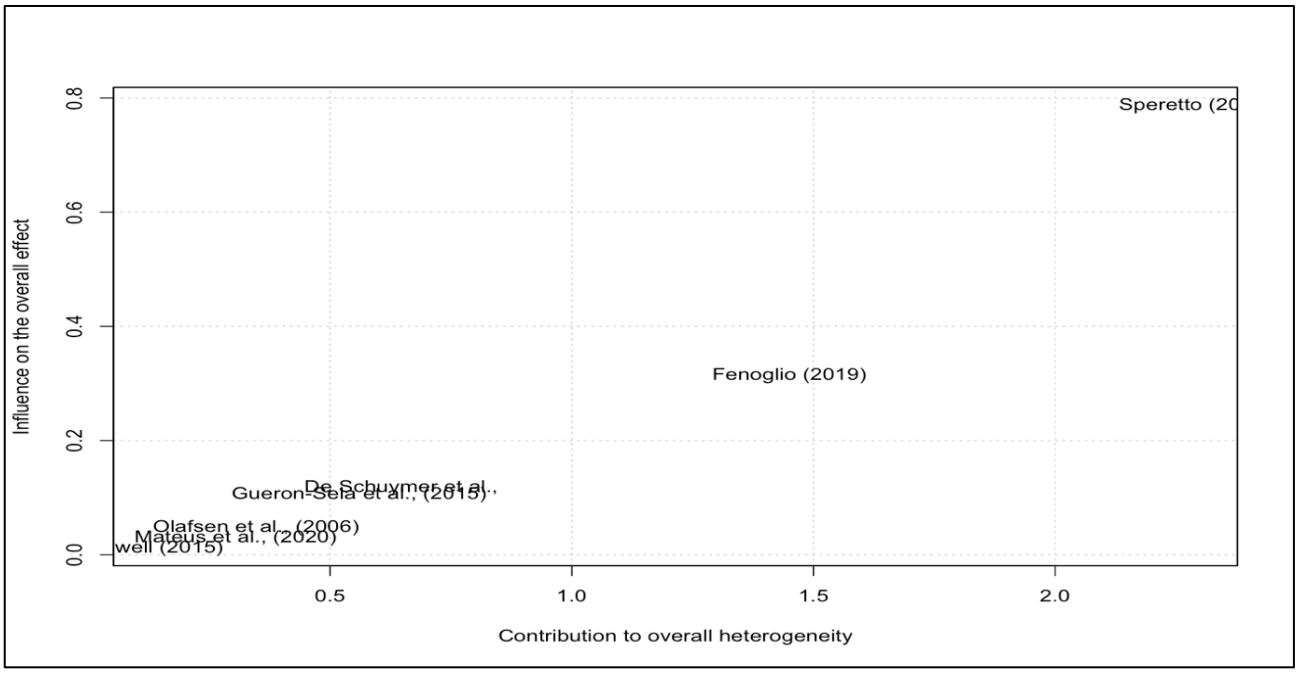
Appendix 4

Figure showing Baujat scatter plots for studies measuring ***gaze following, RJA (point + gaze) and initiating joint attention.***

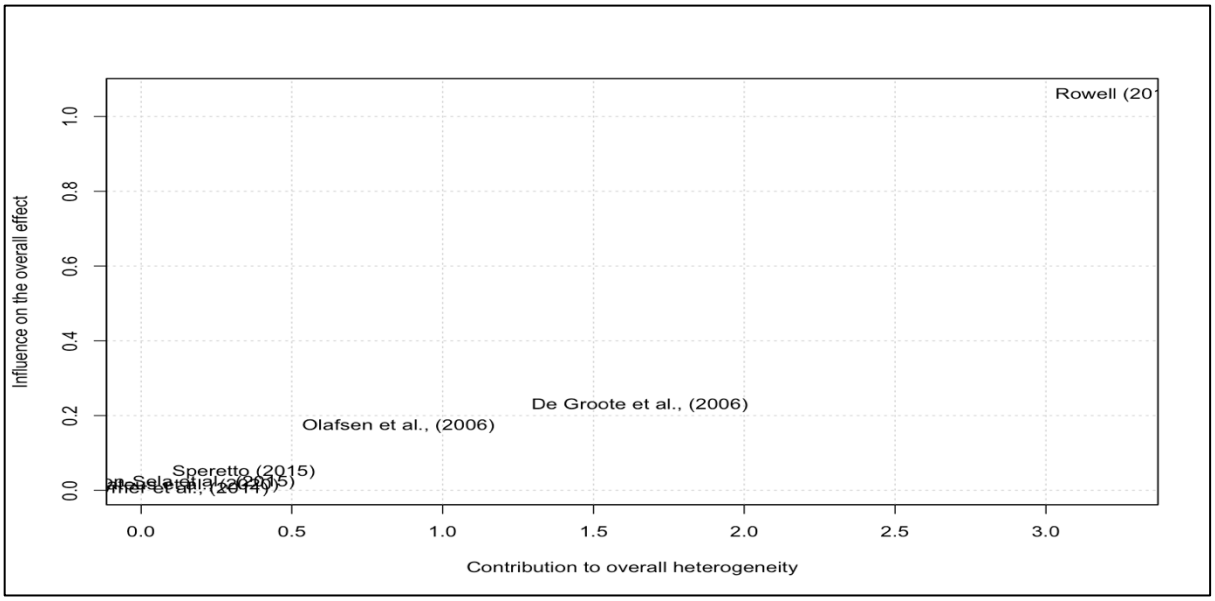
A. Gaze following



B. RJA (point + gaze)



C. Initiating joint attention



Appendix 5

Figures showing leave one out analysis for studies measuring gaze following, RJA (gaze + point), and initiating joint attention.

A. Gaze following

```
Influential analysis (random effects)

                                        SMD            95%-CI p-value   tau^2     tau    I^2
Omitting De Schiuymer et al., (2011)    0.0683 [-0.3476; 0.4842]  0.7475  0.2711  0.5206  75.1%
Omitting Imafuku et al., (2021)         0.0562 [-0.3743; 0.4867]  0.7981  0.2960  0.5440  76.6%
Omitting Imafuku et al., (2017)         0.0199 [-0.4344; 0.4742]  0.9315  0.3431  0.5858  79.8%
Omitting Landry (1986)                 -0.0548 [-0.5200; 0.4105]  0.8176  0.3599  0.5999  79.9%
Omitting Pena at el., (2014) a         -0.1193 [-0.5350; 0.2964]  0.5737  0.2748  0.5242  76.1%
Omitting Pena at el., (2014) b         -0.0718 [-0.5256; 0.3819]  0.7564  0.3423  0.5851  79.5%
Omitting Pena at el., (2014) c         -0.1289 [-0.5337; 0.2759]  0.5326  0.2564  0.5064  75.1%
Omitting Pena at el., (2014) d         -0.0754 [-0.5274; 0.3765]  0.7436  0.3388  0.5821  79.3%
Omitting Rowell (2015)                  0.0562 [-0.3692; 0.4817]  0.7956  0.2913  0.5398  77.6%

Pooled estimate                        -0.0280 [-0.4387; 0.3827]  0.8936  0.3079  0.5549  77.8%

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
```

B. RJA (point + gaze)

```
Influential analysis (random effects)

                                     SMD            95%-CI p-value   tau^2     tau    I^2
Omitting De Schuymer et al.,        -0.6060 [-1.0063; -0.2057]  0.0030  0.1965  0.4433  84.8%
Omitting Fenoglio (2019)            -0.6396 [-1.0022; -0.2769]  0.0005  0.1489  0.3859  76.6%
Omitting Gueron-Sela et al., (2015) -0.6030 [-1.0078; -0.1982]  0.0035  0.2009  0.4482  84.8%
Omitting Mateus et al., (2020)      -0.5013 [-0.8943; -0.1082]  0.0124  0.1999  0.4471  84.7%
Omitting Olafsen et al., (2006)     -0.4901 [-0.8967; -0.0836]  0.0181  0.2038  0.4514  81.9%
Omitting Rowell (2015)              -0.5135 [-0.9163; -0.1106]  0.0125  0.2092  0.4574  84.9%
Omitting Speretto (2015)            -0.3974 [-0.7030; -0.0918]  0.0108  0.0993  0.3151  73.4%

Pooled estimate                     -0.5358 [-0.8918; -0.1798]  0.0032  0.1806  0.4249  82.6%

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
```

C. Initiating joint attention

```
Influential analysis (random effects)

                                      SMD            95%-CI p-value   tau^2     tau    I^2
Omitting De Groote et al., (2006)    0.0222 [-0.3041; 0.3486]  0.8938  0.1132  0.3365  69.0%
Omitting De Schuymer et al., (2011) -0.0456 [-0.4368; 0.3457]  0.8195  0.1789  0.4230  74.7%
Omitting Gueron-Sela et al., (2015) -0.0275 [-0.4238; 0.3688]  0.8918  0.1793  0.4235  74.2%
Omitting Mateus et al., (2020)      -0.0798 [-0.4518; 0.2923]  0.6743  0.1639  0.4049  74.0%
Omitting Olafsen et al., (2006)      0.0185 [-0.3383; 0.3752]  0.9192  0.1340  0.3660  66.6%
Omitting Rowell (2015)              -0.1820 [-0.4195; 0.0555]  0.1330  0.0367  0.1916  44.2%
Omitting Speretto (2015)            -0.1007 [-0.4807; 0.2793]  0.6034  0.1624  0.4030  71.0%

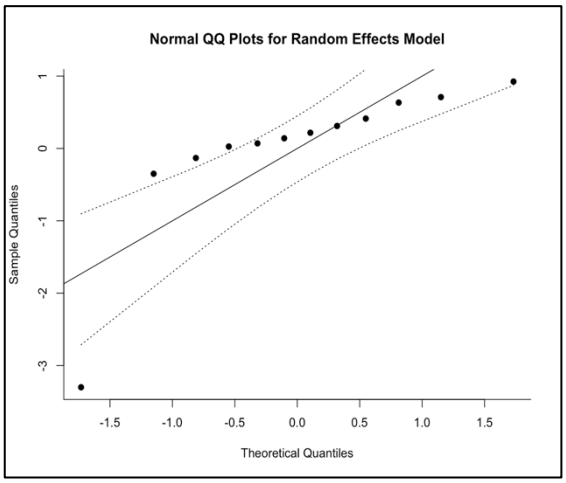
Pooled estimate                     -0.0573 [-0.3835; 0.2689]  0.7306  0.1350  0.3674  69.6%

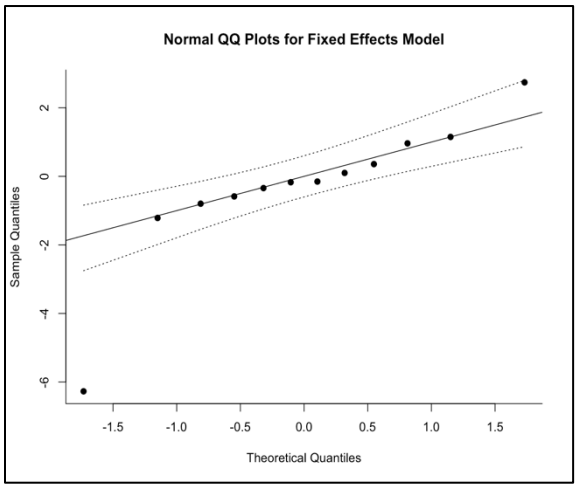
Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
```

Appendix 6

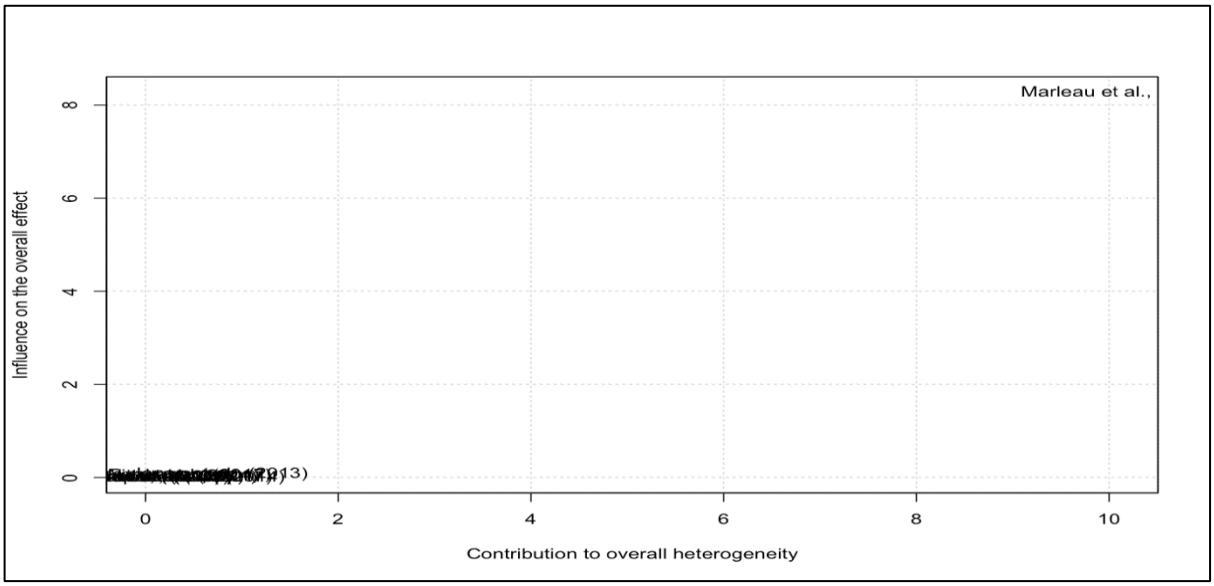
Appendices showing results from preliminary analysis for theory of Mind (prior to Marleau et al., (2020) study was omitted from analysis.

A. *Figure showing QQ plots for both fixed and random effects model for* ***theory of mind (TOM)*** *meta-analysis (prior to removal of Marleau et al., (2020).*





B. *Figure showing Baujat scatterplot for* ***theory of mind (TOM)*** *meta-analysis (prior to removal of Marleau et al., 2020).*



C. *Figure showing outlier and sensitivity analysis via leave one out method for* ***theory of mind (TOM)*** *meta-analysis (prior to removal of Marleau et al., 2020)*

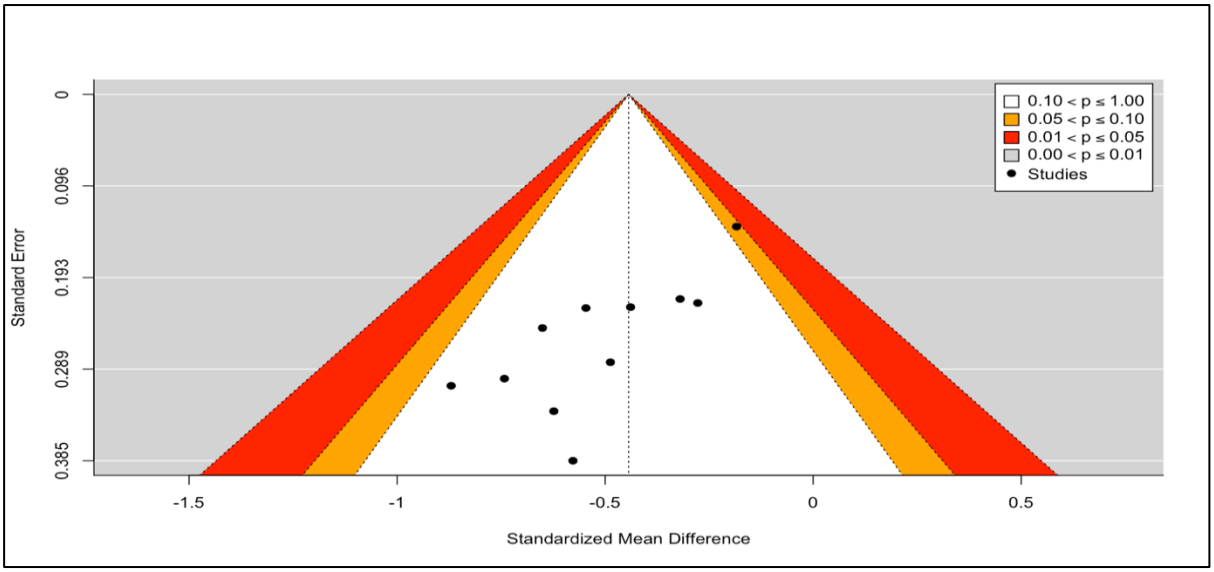
```
Influential analysis (random effects)

                                        SMD            95%-CI  p-value   tau^2     tau    I^2
Omitting Ford et al., (2017)          -0.7061 [-1.0712; -0.3409]   0.0002  0.3079  0.5549  78.7%
Omitting Fu et al., (2022)            -0.6512 [-1.0185; -0.2839]   0.0005  0.3164  0.5625  78.6%
Omitting Jones et al., (2013)         -0.7186 [-1.0780; -0.3591] < 0.0001  0.2930  0.5413  75.4%
Omitting Marleau et al., (2020)       -0.4459 [-0.5955; -0.2964] < 0.0001  0.0080  0.0893   0.0%
Omitting Mossad et al., (2017)        -0.6627 [-1.0328; -0.2925]   0.0005  0.3220  0.5675  79.0%
Omitting Mossad et al., (2021)        -0.6913 [-1.0628; -0.3198]   0.0003  0.3215  0.5670  79.2%
Omitting Roldan-Tapia et al., (2017)  -0.6813 [-1.0550; -0.3076]   0.0004  0.3262  0.5711  79.2%
Omitting Williamson & Jakobson (2014) -0.6711 [-1.0445; -0.2978]   0.0004  0.3263  0.5713  79.1%
Omitting Witt et al., (2014)          -0.7025 [-1.0698; -0.3351]   0.0002  0.3123  0.5589  78.8%
Omitting Witt et al., (2018)          -0.6853 [-1.0555; -0.3150]   0.0003  0.3215  0.5670  79.2%
Omitting Wolfe et al., (2015)         -0.6728 [-1.0419; -0.3037]   0.0004  0.3214  0.5669  79.2%
Omitting Ye et al., (2015)            -0.6757 [-1.0421; -0.3093]   0.0003  0.3182  0.5641  79.2%

Pooled estimate                       -0.6655 [-1.0015; -0.3295]   0.0001  0.2814  0.5305  77.2%

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
```

D. Figure showing funnel plot of the standardised mean differences in ***theory of mind (TOM)*** *between preterm and full-term children* ***prior*** *to trim and fill procedure. The 95% confidence interval of the expected distribution of differences in TOM is shown as an inverted "funnel".*



Appendix 7

Tables showing the impact of study level effects for studies measuring performance on gaze following, responding to joint attention (gaze + point). Initiating joint attention, affect recognition and theory of mind between preterm and full-term children.

A. Social attention

	Study name	Effect	Std Error	CI Lower	CI Upper	Weight (fixed)	Weight (random)
1	Dean et al., (2021)	-0.3965	0.1666	-0.7230	-0.06999	36.04	36.04
2	Imafuku et al., (2017)	-0.0586	0.2928	-0.6325	0.51530	11.66	11.66
3	Imafuku et al., (2021)	-0.4449	0.2462	-0.9049	0.01507	18.16	18.16

B. RJA (Gaze following)

	Study name	Effect	Std Error	CI Lower	CI Upper	Weight (fixed)	Weight (random)
1	De Schuymer et al., (2011)	-0.7722	0.2679	-1.2973	-0.2471	13.933	2.634
2	Imafuku et al., (2021)	-0.6644	0.2582	-1.1704	-0.1584	15.001	2.669
3	Imafuku et al., (2017)	-0.4074	0.3033	-1.0019	-0.1871	10.870	2.500
4	Landry (1986)	0.1834	0.2462	-0.2992	-0.6660	16.492	2.713
5	Pena et al., (2014) a	0.7331	0.3246	0.0969	1.3694	9.490	2.419
6	Pena et al., (2014) b	0.3373	0.3064	-0.2632	0.9378	10.653	2.489
7	Pena et al., (2014) c	0.8154	0.3279	0.1727	1.4581	9.300	2.407
8	Pena et al., (2014) d	0.3654	0.3057	-0.2338	0.9647	10.698	2.491
9	Rowell (2015)	-0.7328	0.3163	-1.3526	-0.1129	9.998	2.451

C. RJA (gaze + point)

	Study name	Effect	Std Error	CI Lower	CI Upper	Weight (fixed)	Weight (random)
1	De Schuymer et al., (2011)	-0.16321	0.1843	-0.5244	0.1980	29.445	4.661
2	Fenoglio (2019)	-0.00171	0.1270	-0.2507	0.2473	61.963	5.084
3	Gueron-Sela et al., (2015)	-0.19400	0.1672	-0.5217	0.1337	35.775	4.796
4	Mateus et al., (2020)	-0.84144	0.3553	-1.5377	-0.1452	7.924	3.260
5	Olafsen et al., (2006)	-0.80710	0.1775	-1.1550	-0.4592	31.743	4.715
6	Rowell (2015)	-0.73277	0.3163	-1.3526	-0.1129	9.998	3.564
7	Speretto (2015)	-1.27162	0.2362	-1.7346	-0.8086	17.921	4.231

D. Initiating joint attention (IJA)

	Study name	Effect	Std Error	CI Lower	CI Upper	Weight (fixed)	Weight (random)
1	De Groote et al., (2006)	-0.6780	0.3130	-1.2915	-0.06445	10.21	4.293
2	De Schuymer et al., (2011)	-0.1105	0.2390	-0.5790	0.35801	17.50	5.206
3	Gueron-Sela et al., (2015)	-0.1940	0.1681	-0.5235	0.13549	35.38	6.127
4	Mateus et al., (2020)	0.1319	0.3402	-0.5349	0.79867	8.64	3.989
5	Olafsen et al., (2006)	-0.4323	0.1726	-0.7706	-0.09399	33.56	6.070
6	Rowell (2015)	0.7730	0.2787	0.2268	1.31911	12.88	4.704
7	Speretto (2015)	0.1867	0.1999	-0.2051	0.57849	25.02	5.717

E. Affect recognition.

	Study name	Effect	Std Error	CI Lower	CI Upper	Weight (fixed)	Weight (random)
1	Longa et al., (2022)	-0.00581	0.2736	-0.5421	0.53043	13.359	10.024
2	Marleau et al., (2020)	-1.188215	0.2770	-0.7311	-0.64537	13.036	9.841
3	Marotta et al., (2020)	-0.414763	0.1949	-0.7968	-0.03268	26.313	15.897
4	Mossad et al., (2021)	-0.403350	0.2266	-0.8475	0.04078	19.475	13.115
5	Sato et al., (2021)	-0.618873	0.2806	-1.1689	-0.06887	12.699	9.648
6	Twilhaar et al., (2019)	-0.457946	0.1823	-0.8153	-0.10062	30.087	17.201
7	Williamson & Jakobson (2014)	-0.735520	0.2445	-1.2148	-0.25625	16.724	11.807
8	Witt et al., (2014)	-0.537081	0.2157	-0.9598	-0.11439	21.501	14.004
9	Wolfe et al., (2015)	-0.549203	0.3243	-1.1847	0.08633	9.511	7.690
10	Ye et al., (2015)	-0.055907	0.3350	-0.7124	0.60061	8.913	7.294

F. Theory of mind (ToM)

	Study name	Effect	Std Error	CI Lower	CI Upper	Weight (fixed)	Weight (random)
1	Ford et al., (2017)	-0.2775	0.2169	-0.7027	0.147647	21.252	18.173
2	Fu et al., (2022)	-0.8702	0.3011	-1.4604	-0.280048	11.029	10.138
3	Jones et al., (2013)	-0.1842	0.1385	-0.4556	0.087178	52.153	36.840
4	Mossad et al., (2020)	-0.7422	0.2939	-1.3182	-0.166230	11.579	10.601
5	Mossad et l., (2021)	-0.4389	0.2215	-0.8731	-0.004774	20.379	17.531
6	Roldan-Tapia et al., (2017)	-0.5463	0.2221	0.1109	0.981651	20.267	17.448
7	Williamson & Jakobson (2014)	-0.6508	0.2428	-1.1267	-0.174873	16.962	14.942
8	Witt et al., (2014)	-0.3199	0.2132	-0.7377	0.098000	22.001	18.718
9	Witt et al., (2018)	-0.4875	0.2773	-1.0310	0.056049	13.003	11.782
10	Wolfe at al., (2015)	-0.6235	0.3260	-1.2624	0.015519	9.408	8.752
11	Ye et al., (2015)	-0.5775	0.3742	-1.3110	0.155928	7.140	6.756

CHAPTER TWO APPENDICES

Appendix 1

Appendices showing ethical approval for amended IPSU protocol.



AERamendments_School_Eth...
85 KB

Show all 4 attachments (2 MB)

Save all to OneDrive - University of Birmingham Download all

Dear AER_Ethics,

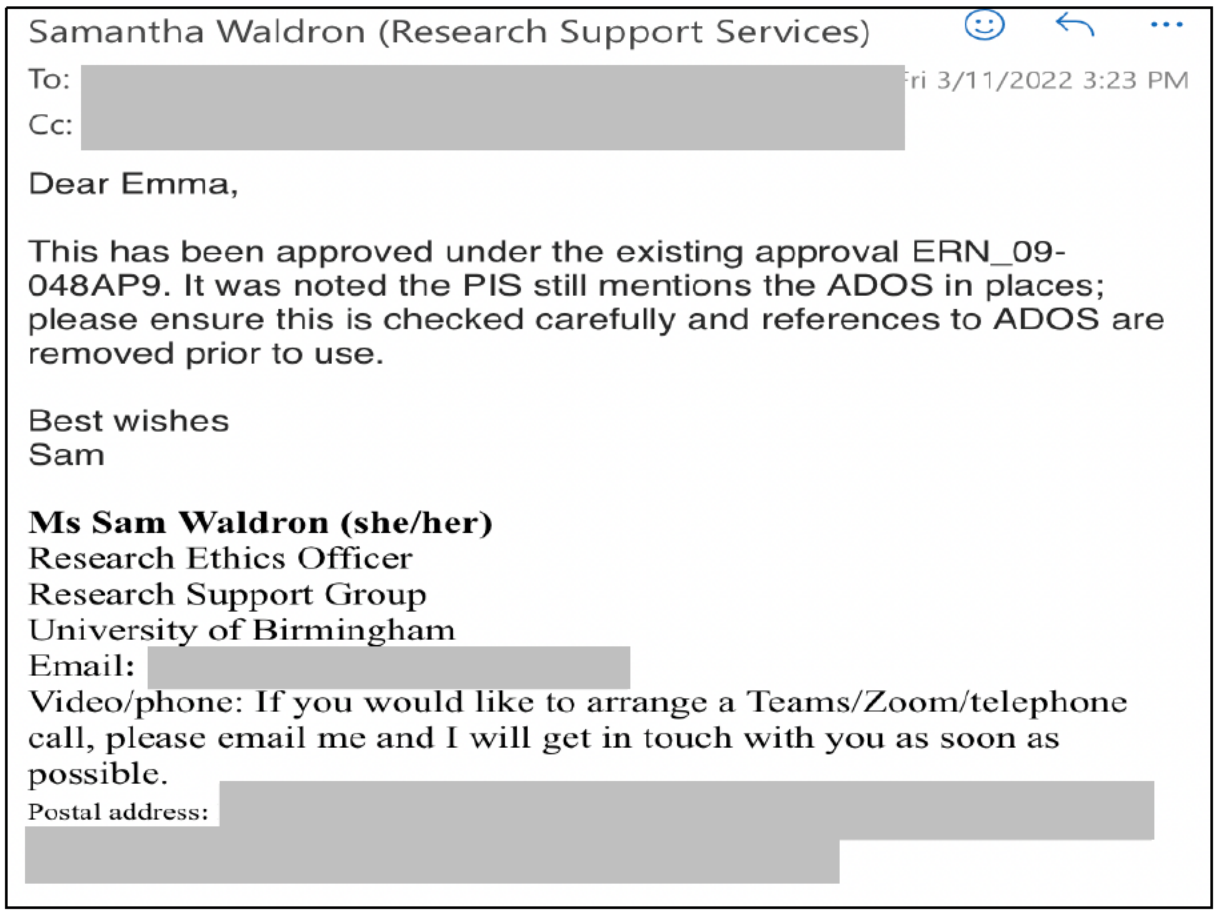
Please find attached an application for an amendment to our currently approved study. The amendment is to remove the ADOS from our already approved study protocol. I have attached an amended participant information sheet, poster and feedback form. I have included tracked changes on the documents so it is easy for you to see the amended information.

We understand how busy you must be, but would hugely appreciate if this request could be considered promptly. Please do let me know if there is anything I can do to help with the processing of this amendment.

Best wishes,

Emma Covernton

Trainee Clinical Psychologist



Samantha Waldron (Research Support Services)

To: [redacted] Fri 3/11/2022 3:23 PM

Cc: [redacted]

Dear Emma,

This has been approved under the existing approval ERN_09-048AP9. It was noted the PIS still mentions the ADOS in places; please ensure this is checked carefully and references to ADOS are removed prior to use.

Best wishes
Sam

Ms Sam Waldron (she/her)
Research Ethics Officer
Research Support Group
University of Birmingham
Email: [redacted]
Video/phone: If you would like to arrange a Teams/Zoom/telephone call, please email me and I will get in touch with you as soon as possible.
Postal address: [redacted]

Appendix 2

Appendices showing health and background questionnaire.



If you're unsure about what any of the questions mean, please ask a member of the study team. If you don't know the answer to any of the questions, please leave blank.

Maternal History

Did the mother smoke during pregnancy? Yes No

Did any members of the household smoke during pregnancy? Yes No

Did the mother drink during pregnancy? Yes No

Did any members of the household drink during pregnancy? Yes No

Did the mother exercise during pregnancy? Yes No

What was the mother's age at birth? _____

Was the mother admitted to hospital during her pregnancy? Yes No

If yes, please detail

_____

_____

_____

Did the mother experience any significant periods of illness during pregnancy? Yes No

If yes, please detail

_____

_____

_____

Was this child part of a multiple pregnancy? Yes No

If yes how many babies? _____ babies

Were any of the multiples identical? Yes No

Birth

When was your child born? ___/___/___ Approximate time: _____

Place of birth (please circle) Hospital / Birthing Centre / Home

What type of birth was this? Vaginal / Caesarean (Emergency/Scheduled/Repeat)

Did labour start on its own? Yes No

At how many weeks gestation? _____

Did your water break on its own? Yes No

Did you have an epidural? Yes No
If YES, how far along were you during labour? _____

Did you have any pain medication? _____

Yes No If YES, what kind? _____

Please detail any known birth complications:

_____

_____

_____

Was a NICU stay required? Yes No

If yes, for how long? Weeks_____ Days_____ Hours_____

Did your infant require the use of a ventilator? Yes No

Did your infant experience any Intracranial haemorrhages (also known as brain bleeds)? Yes No

Did your infant require any surgeries during their first year of life? Yes No

If yes, please detail

_____

_____

Baby

Birth weight _____

Birth height _____

Birth length _____ Head circumference _____

APGAR score (at 1 minute)_____ (at 5 minutes)_____ (at 10 minutes)_____

Age at discharge from hospital (days/hours) _____ _____

Discharge weight _____

Did your baby have the heel prick test? Yes No

Did the results return clear? Yes No

Was your baby breastfed? Yes No

If no, please circle the formula details Milk / Soy / Organic / Homemade

Does your baby take any supplements directly? Yes No

If yes, please detail

_____

_____

The following questions ask for background information <u>about you and your family</u>. Please tick the appropriate boxes or write in the spaces provided.

1. Are you male or female? Male ☐ Female ☐ **Postcode**:_____

2. What was your age in years on your last birthday? _____years

3. Please tick the highest level of your educational qualifications.

No formal educational qualifications..... ☐
Fewer than 5 GCSE's or O Level's (grades A-C), NVQ 1, or BTEC First Diploma……. …. ☐
5 or more GCSE's or O Level's (grades A-C), NVQ 2, or equivalent……..…….. ☐
3 or more 'A' Levels, NVQ 3, BTEC National, or equivalent..... ☐
Polytechnic/University degree, NVQ 4, or equivalent..... ☐
Masters/Doctoral degree, NVQ 5, or equivalent……..... ☐

4. What is your relationship to your infant participating (e.g. mother, father, stepmother, grandmother, adoptive parent)? _____

5. In total how many people currently live in your home? _____ Adults _____ Children

6. What is your current marital status?

Married, and living with spouse..... ☐

Living with partner..... ☐

Divorced/Separated/Widowed/Single and NOT living with a partner..... ☐

If living with partner/spouse, please answer the following questions, if not, please go to question 11.

7. Is your partner male or female? Male ☐ Female ☐

8. What was their age in years on their last birthday? _____ years

9. Please tick the highest level of your partner/spouse's educational qualifications.

No formal educational qualifications..... ☐

Fewer than 5 GCSE or O Level (grades A-C), NVQ 1, or BTEC First Diploma…… ☐

5 or more GCSE or O Level (grades A-C), NVQ 2, or equivalent……..… ☐

3 or more 'A' Levels, NVQ 3, BTEC National, or equivalent..... ☐

Polytechnic/University degree, NVQ 4, or equivalent..... ☐

Masters/Doctoral degree, NVQ 5, or equivalent……..... ☐

10. What is your partner/spouse's relationship to your infant participating (e.g., mother, father, stepmother, adoptive parent)? _____

11. What is your current total annual family income? Please include a rough estimate of total salaries and other income (including benefits) before tax and national insurance/pensions.

Please tick one box only:

Less than £15,000……..……...… ☐

£15,001 to £25,000……...……. ☐

£25,001 to £35,000……...……..……. ☐

£35,001 to £45,000……...…… ☐

£45,001 to £55,000……..…… ☐

£55,001 to £65,000……..…….. ☐

£65,001 or more……..……..……. ☐

12.Which of the following best describes the ethnic group of your child?

☐ White

- ☐ Mixed or multiple ethnic groups
- ☐ Asian or Asian British
- ☐ Black, African, Caribbean, or Black British
- ☐ Not listed: _____

Appendix 3

Appendices showing the Modified Checklist for Autism in Toddlers – Revised (MCHAT-R). Please note the follow up questions are not included in the appendices but can be accessed via the following link https://drexel.edu/~/media/Files/autismInstitute/EDI/M-CHAT-R_F.ashx.

M CHAT www.m-chat.org

Child's name _____ Date _____

Age _____ Relationship to child _____

M-CHAT-R™ (*Modified Checklist for Autism in Toddlers Revised*)

Please answer these questions about your child. Keep in mind how your child usually behaves. If you have seen your child do the behavior a few times, but he or she does not usually do it, then please answer **no**. Please circle **yes** or **no** for every question. Thank you very much.

1. If you point at something across the room, does your child look at it? (**FOR EXAMPLE**, if you point at a toy or an animal, does your child look at the toy or animal?)	Yes	No
2. Have you ever wondered if your child might be deaf?	Yes	No
3. Does your child play pretend or make-believe? (**FOR EXAMPLE**, pretend to drink from an empty cup, pretend to talk on a phone, or pretend to feed a doll or stuffed animal?)	Yes	No
4. Does your child like climbing on things? (**FOR EXAMPLE**, furniture, playground equipment, or stairs)	Yes	No
5. Does your child make unusual finger movements near his or her eyes? (**FOR EXAMPLE**, does your child wiggle his or her fingers close to his or her eyes?)	Yes	No
6. Does your child point with one finger to ask for something or to get help? (**FOR EXAMPLE**, pointing to a snack or toy that is out of reach)	Yes	No
7. Does your child point with one finger to show you something interesting? (**FOR EXAMPLE**, pointing to an airplane in the sky or a big truck in the road)	Yes	No
8. Is your child interested in other children? (**FOR EXAMPLE**, does your child watch other children, smile at them, or go to them?)	Yes	No
9. Does your child show you things by bringing them to you or holding them up for you to see – not to get help, but just to share? (**FOR EXAMPLE**, showing you a flower, a stuffed animal, or a toy truck)	Yes	No
10. Does your child respond when you call his or her name? (**FOR EXAMPLE**, does he or she look up, talk or babble, or stop what he or she is doing when you call his or her name?)	Yes	No
11. When you smile at your child, does he or she smile back at you?	Yes	No
12. Does your child get upset by everyday noises? (**FOR EXAMPLE**, does your child scream or cry to noise such as a vacuum cleaner or loud music?)	Yes	No
13. Does your child walk?	Yes	No
14. Does your child look you in the eye when you are talking to him or her, playing with him or her, or dressing him or her?	Yes	No
15. Does your child try to copy what you do? (**FOR EXAMPLE**, wave bye-bye, clap, or make a funny noise when you do)	Yes	No
16. If you turn your head to look at something, does your child look around to see what you are looking at?	Yes	No
17. Does your child try to get you to watch him or her? (**FOR EXAMPLE**, does your child look at you for praise, or say "look" or "watch me"?)	Yes	No
18. Does your child understand when you tell him or her to do something? (**FOR EXAMPLE**, if you don't point, can your child understand "put the book on the chair" or "bring me the blanket"?)	Yes	No
19. If something new happens, does your child look at your face to see how you feel about it? (**FOR EXAMPLE**, if he or she hears a strange or funny noise, or sees a new toy, will he or she look at your face?)	Yes	No
20. Does your child like movement activities? (**FOR EXAMPLE**, being swung or bounced on your knee)	Yes	No

© 2009 Diana Robins, Deborah Fein, & Marianne Barton

Appendix 4

Appendices describing the Early Social Cognition Scale individual task discrepancies.

Name of task/*Original Author*	**Task as described in original paper**	**Task as shown in original video footage**[115]	**Scoring sheet**[116] **instructions and pre covid protocol**[117]	**Decision/ Rationale**
Helping task *Warneken & Tomasello (2006;2007)*	E1 does ***all*** trials in both control and experimental condition. Additionally, there are ***three*** trials for paper bell task and two trials for pen drop task.	**Ellis original footage videos**: show E1 does ***one*** experimental trial and ***one*** control trial. E2 does ***one*** experimental and ***one*** control for both paper bells and pen task	E1 does ***all*** control conditions (paper bells and pen drop). E2 does ***all*** experimental conditions (paper bells and pen drop).	Decision to remain with scoring sheet and precovid protocol for consistency. The discrepancy is related to the role of E1 and E2/number of trials. How the task is performed remains consistent.
Gestures task ***Behne,*** *Carpenter & Tomasello (2005)*	No description of control condition. Experimental condition - states that E1 does *both* gaze and point and says to the child "I can see" whilst hiding the toy. E2 (described as "assistant") asks child "where is the toy". Assistant does not say "I can see it" which makes sense if its E1 who always hides the toys *and* conducts the control/experimental cues.	**Ellis original footage videos:** show E1 does *control* cues (gaze and point) E2 does *experimental* cues (gaze and point). E1 always asks the child "where is it" on *all* control *and* experimental trials. E2 Says "I can see it" on experimental trials only. E1 turns away from the table whilst E2 does either experimental gaze or point, E1 then turns back to face infant and then E1 asks infant "where is the toy?" **Supplementary material from Ellis**: E1 does control cues (point and gaze), E2 asks infant "where is the toy?" (on control conditions) E2 does experimental cues (point and gaze)	E1 does *all* control and experimental cues. E2 "I can see it" on experimental conditions only. E1 always says "where is it?" to the child- across control and experimental conditions.	Decision to remain with scoring sheet and precovid protocol for consistency. The discrepancy is predominantly related to the role of E1 and E2. How point and gaze cues are delivered remains consistent.

[115] Footage available at the Richards lab which was conducted as part of ESCogS author thesis project (Kat Ellis).
[116] Scoring sheet was kindly provided to the Richards Lab by ESCogS author (Kat Ellis).
[117] Pre-covid protocol refers to those participants included in this study, whose data was collected prior to national lockdown.

Name of task/*Original Author*	**Task as described in original paper**	**Task as shown in original video footage[115]**	**Scoring sheet[116] instructions and pre covid protocol[117]**	**Decision/ Rationale**
		E1 turns away from the table whilst E2 does either experimental gaze or point. E1 then turns back to face infant and then E1 asks infant "where is the toy?"		

Please note – *E1 = Experimenter one; E2 = Experimenter two.*

Appendix 5

Appendices showing contractual agreement for the use of the Geometric Preference task.

Re: Eye Tracking Stimuli

Amanda Cheng [redacted]

Mon 04/11/2019 23:01

To: Catherine Laverty (PhD Psychology Lab FT) <[redacted]>

Cc: Pierce, Karen <[redacted]>; Donna Shaw [redacted]

Hi Catherine,

Thank you for your interest in our GeoPref research! The technology transfer has been completed by UCSD and I have attached the GeoPref stimuli to this email. As a reminder, this test is only validated for use between the ages of 12 and 36 months, and can only be used for research purposes only. Please let me know if you have any questions!

Original GeoPref.mp4

Sincerely,
Amanda Cheng
Clinical Research Coordinator
UCSD Autism Center of Excellence
[redacted]

On Tue, Oct 1, 2019 at 5:49 AM Catherine Laverty <[redacted]> wrote:

> Hi Karen,
>
> Thank you so much for your response! We look forward to receiving the technology transfer agreement and any other advice regarding the tasks.
>
> The GeoPref tests sounds perfect for our current study. We are interested in the social cognitive development of infants born preterm, and as part of this will be recruiting both term and moderately-late preterm infants aged between 18-36 months (corrected for gestational age in the preterm group). We will also be conducting a range of behavioural assessments alongside the eye-tracking tasks which would complement the GeoPref test really well!
>
> Many thanks for your help with this, we are really looking forward to incorporating the tasks within our research.
>
> Best,
>
> Catherine
>
> **Catherine Laverty**
>
> Doctoral Researcher
>
> University of Birmingham
>
> School of Psychology
>
> 304, 52 Pritchatts Road

Appendix 6

Appendices showing contractual agreement for use of biological motion preference task.

UNIVERSITY OF CALIFORNIA, SAN DIEGO UCSD

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO SANTA BARBARA • SANTA CRUZ

OFFICE OF INNOVATION AND COMMERCIALIZATION

DIRECT DIAL:
FAX:
http

October 7, 2019

University of Birmingham
52 Pritchatts Road
School of Psychology
Edgbaston, Birmingham
B15 2TT

Re: Non-Exclusive Copyright License to "New Early Diagnostic Test for Autism", Case No. SD2010-252 ("COPYRIGHT MATERIALS")

We are pleased to learn that University of Birmingham ("LICENSEE") has an interest in obtaining from The University of California San Diego ("UNIVERSITY") a copy of the COPYRIGHT MATERIALS and a license to use, copy and modify the COPYRIGHT MATERIALS for academic purposes. The research study "The Impact of Prematurity on Social Understanding" will be conducted by Drs. Caroline Richards, Andrew Surtees and Ms. Catherine Laverty. We also understand that LICENSEE would use the COPYRIGHT MATERIALS and practice such license for such research purposes at LICENSEE's facilities as addressed above ("SITE") during the period beginning on the date LICENSEE signs this Agreement and ending on August, 2021 ("TERM").

In order to avoid any misunderstanding, UNIVERSITY would like to memorialize the terms and conditions of such a license by way of this Agreement. Accordingly, LICENSEE and UNIVERSITY agree that:

1. **Grant.** Subject to the terms and conditions of this Agreement, UNIVERSITY hereby grants to LICENSEE, and LICENSEE accepts, a non-exclusive license under the copyrights subsisting in the COPYRIGIIT MATERIALS to use, copy and modify the COPYRIGIIT MATERIALS for research and academic purposes only, at the SITE and for the TERM.

2. **Consideration.** In consideration for the license granted in paragraph 1 herein, LICENSEE promises that **it will not**:

(a) install the COPYRIGHT MATERAIALS in a single CPU or in a network, multiple CPU or multiple-user arrangement, except within the confines of the SITE;

Appendix 7

Appendices showing table of The Early Social Cognition Scale task counterbalancing – from Ellis et al., (2020) supplementary materials – open access journal with permission to use with citation.

Order 1	Order 2	Order 3	Order 4	Order 5	Order 6
Re-enactment of Intended Acts	Helping (control)	Helping (control)	Re-enactment of intended acts	Gestures	Gestures
Helping Task (control)	Re-enactment of intended acts	Gestures	Gestures	Re-enactment of Intended Acts	Helping (control)
Gestures	Gestures	Re-enactment of intended acts	Helping (control)	Helping (control)	Re-enactment of intended acts
Tubes	Trampoline	Helping (experimental)	Tubes	Trampoline	Helping (experimental)
Trampoline	Tubes	Trampoline	Helping (experimental)	Helping (experimental)	Tubes
Helping (experimental)	Helping (experimental)	Tubes	Trampoline	Tubes	Trampoline

Appendix 8

Appendixes showing table showing group scaling and participant characteristics according to the Early Social Cognition Scale compared to Mullen composite score and age for both moderate to late preterm children and full-term children.

A. Moderate to late preterm children.

	Patterns					
Pattern	**0**	**1**	**2**	**3**	**4**	**Other patterns**
Helping						-
REI[118] or communication point						-
Communication: Gaze or Co-operation: Tubes						-
Trampoline						-
Number of cases	0	0	1	2	0	5
Age (mean)[119]	-	-	23.23	26.6	-	
Age (SD)	-	-	N/A	1.3	-	
Age (Range)	-	-	N/A	25.7-27.5	-	
Mullen composite score (mean)	-	-	77	70.5	-	
Mullen composite score (SD)	-	-	N/A	3.5	-	
Mullen composite (Range)	-	-	N/A	68-73	-	

Note – Green = passed task, red = task failed

[118] Re-enactment of intended acts task
[119] Adjusted age in months; NA = not applicable

B. Full-term children

	Patterns					
Pattern	**0**	**1**	**2**	**3**	**4**	**Other patterns**
Helping						-
REI[120] or communication point						-
Communication: Gaze or Co-operation: Tubes						-
Trampoline						-
Number of cases	1	0	0	1	1	5
Age (mean)[121]	18.4	-	-	21.7	24.0	
Age (SD)	N/A	-	-	N/A	N/A	
Age (Range)	N/A	-	-	N/A	N/A	
Mullen composite score (mean)	122	-	-	76	- 105	
Mullen composite score (SD)	N/A	-	-	N/A	N/A	
Mullen composite (Range)	N/A	-	-	N/A	N/A	

Note *– Green = passed task, red = task failed*

[120] Re-enactment of intended acts task
[121] Chronological age in months; NA = not applicable

Appendix 9

*Appendices of table showing group scaling and participant characteristics according to the Early Social Cognition Scale compared to Mullen composite score and age, for both moderate to late preterm children and full-term children demonstrating **other** patterns of performance on the Early Social Cognition Scale.*

A. Moderate to late preterm children

Pattern	Other observed patterns			
	5	6	7	8
Helping				
REI[122] or communication point				
Communication: Gaze or Co-operation: Tubes				
Trampoline				
Number of cases	1	1	1	2
Age (mean)[123]	33.6	18.1	31.7	25.2
Age (SD)	N/A	N/A	N/A	0.2
Age (Range)	N/A	N/A	N/A	25.4 – 25.1
Mullen composite score (mean)	69	94	124	84.5
Mullen composite score (SD)	N/A	N/A	N/A	30.4
Mullen composite (Range)	N/A	N/A	N/A	63-106

[122] Re-enactment of intended acts task
[123] Corrected age in months; NA = not applicable

B. Full-term children

Pattern	Other observed patterns			
	5	**6**	**7**	**8**
Helping				
REI[124] or communication point				
Communication: Gaze or Co-operation: Tubes				
Trampoline				
Number of cases	1	1	1	2
Age (mean)[125]	33.3	22.9	28.8	27.2
Age (SD)	N/A	N/A	N/A	1.2
Age (Range)	N/A	N/A	N/A	26.3-27.2
Mullen composite score (mean)	117	77	141	127
Mullen composite score (SD)	N/A	N/A	N/A	8.5
Mullen composite (Range)	N/A	N/A	N/A	121-133

Note *– Green = passed task, red = task failed*

[124] Re-enactment of intended acts task
[125] Chronological age in months; NA = not applicable