Lab and Lounge

Relating Measures of Visual Working Memory from Naturalistic to Laboratory Setting in Rural India

Prerna Aneja

100184077

School of Psychology

University of East Anglia

A thesis submitted for the degree of

Doctor of Philosophy

September 2022

Presentations Arising From This Thesis

- Aneja, P., Forbes, S., & Spencer, JP. (2021) Bridging the lab and lounge to understand the early development of visual working memory in rural India. In P. Aneja, (Chair), *Parent-Infant Interactions Around the World: What Mother Behaviours Across Socio-Cultural Groups Promote Cognitive and Language Outcomes?* [Presentation] Biennial Meeting Society for Research in Child Development.

- Aneja, P., Carr, T., Sami, S., & Spencer JP. (2021) *Automatic Coding of Parent-Infant Dyadic Play Videos Using Convolutional Neural Networks.* [Poster] Biennial Meeting Society for Research in Child Development.

- Aneja, P., Carr, T., Sami, S., Forbes, S., Maurya, V., & Spencer, JP (2020). *Leveraging technological advances to bridge methodological limitation in low-resource setting.* [Poster] virtual International Congress of Infant Studies.

- Aneja, P., Forbes, S., & Spencer, JP (2020). *Influence of caregiver's visual exploration on infant's cognitive outcome in India and UK.* [Poster] Budapest CEU Conference on Cognitive Development, Budapest.

- , Aneja, P., Forbes, S., & Spencer, JP. (2019) Understanding caregivers' visual experience and infants' cognitive outcomes: A cross cultural study. S.H. Forbes (Chair), *Global Brain Health and Development.* [Presentation] XVI European Congress of Psychology, Moscow.

- Aneja, P., Forbes, S., & Spencer, JP. (2019). *Visual exploration during parent-infant interaction in India and UK.* [Presentation] The 4th Lancaster Conference on Infant and Early Child Development, Lancaster.

- Aneja, P., Forbes, S., & Spencer, JP. (2018). *Cultural differences in dyadic interaction using head-mounted cameras.* [Presentation] From research to theory and practise: changing lives, BPS East of England Annual Conference.

- Aneja, P., Forbes, S., & Spencer, JP. (2018). *Cultural differences in dyadic interaction using head-mounted cameras.* [Presentation] European Social Cognition Network, Cologne, Germany.

- Aneja, P., Forbes, S., & Spencer, JP. (2018). *Cultural differences in dyadic interaction using head-mounted cameras.* [Poster] Social Communication Across the Lifespan, Canterbury, UK.

- Aneja, P., Forbes, S., & Spencer, JP. (2018). *Cultural differences in dyadic interaction using head-mounted cameras.* [Presentation] The 48th Annual Meeting of the Jean Piaget Society, Amsterdam, The Netherlands.

Abstract

Research from laboratory and naturalistic settings, has argued for the importance of visual cognition for infants' early development and noted its predictive value for later social and cognitive developmental outcomes. To date, however, few research studies have explored the relationship between measures of visual cognition in the lab and the real world. The aim of this thesis is to understand whether individual differences in visual attention measured during caregiver-infant interaction relate to individual differences in measures from a visual working memory (VWM) laboratory task. The thesis adopts an eco-cultural perspective by embedding this question in the cultural and socio-emotional context of the caregivers and its connections to their infants. Studies 1 and 2 focus on laboratory measures of caregivers and infants respectively, and on their connection to the socio-emotional context of the caregiver in a low-resource setting. Findings suggest that caregivers' socio-emotional context influences their own VWM performance but not that of their infants. In addition, caregiver and infant VWM performances were inversely related to each other. Study 3 uses a machine learning pipeline developed in Chapter 3 to assess dyadic interactions from participants in India, as well as from a UK group of dyads (high resource setting). Interactions were recorded by using head-mounted eye trackers, with the machine learning pipeline being used to extract measures of visual attention in a naturalistic context. Findings revealed similarities and differences across the cultural groups in line with culturally normative parenting styles. The final study (Chapter 5) links measures of infants' visual cognition from the lounge and

the lab, exploring the associations between parental and infant visual cognition as measured across settings. In Chapter 6, findings are integrated and the contributions to the literature are discussed. We further consider how the insights from our research can help inform future parent-focused interventions.

Acknowledgements

Just like infants do not grow in isolation, the development and completion of this thesis involved support and contributions from many people.

I feel immense gratitude towards my primary supervisor, Professor John P. Spencer, for the constant support, guidance, and mentorship he has provided. Thank you, John, for giving me the opportunity to be a part of this wonderful project, and for giving me the space to develop my confidence and grow as a researcher. I would also like to thank Ghost for acting as honorary cat supervisor during covid – Ghost's presence always made our online meetings more fun!

As well as John, I extend my thanks to Larissa for making DDLab, creating a wonderful community in it, and taking me into it, thus making me one of their academic children. I would also like to acknowledge the support and input from every member of DDLab (past and present): Jo, Jordan, Joe, Laura, Laia, Lourdes, Kate, Kiara, Ellie, Melina, Sushila, and Barbara. Some of you contributed to this project through data collection and data processing, others helped me with fruitful conversations about the thesis, and most importantly many of you have patiently listened to me complain (over coffee, of course!). Among every colleague in DDLab, Sam deserves a special acknowledgement as my friend and mentor, and for giving me advice on R, teaching me LATEX, and encouraging me to voice my thoughts and delve into open science!

I would also like to acknowledge the hard work that went into recruiting and collecting data in India by the CEL team. I would particularly like to thank Vineeta and Babita for all their hard work. Vineeta has been central to collecting and organising data, taking me into the community to better understand them and their context, and contributing to solving many of the challenges of researching parent-infant interaction in Shivgarh.

Tom, Saber, and Jake also deserve being acknowledged for their invaluable contributions to setting up and running the machine learning algorithm. I particularly would like to thank Tom and Jake for the time and care they put into teaching me and patiently answering all my questions. I am very proud of my Machine Learning chapter, and I am indebted to you both for that.

A special mention must also go to my best friend Ramiz Ilham Khan - you know I wouldn't have been here without you. My thanks also extend to Maitrey for all the love and support in this journey, to Vivek Sharma for constantly asking me "karna kya hai?", and to Ginni, Neha, and Bon for always being there despite the distance and many special moments that I missed.

Among my people in England, Iona deserves special recognition. From the very first day, you have been a such great support for me. You have been there at each and every step of this thesis. You have known all my highs and lows during this journey and have been there to support me at all times. For making me dinner when I was working late, and especially for surprising me with a bike so I could happily ride to uni. Similarly, I must thank my friends Eesha and Rand not only for being supportive but also for jumping on board to learn R together, although I was the only one who needed it. To Yuqing, Carys and Melina (again), thank you for always reminding me "you got this!". This thesis would not have been completed without their constant support.

I would also like to extend my thanks to my family, by blood and beyond it. Thanks to Mumma and papa for their love and support as well as for the constant supply of *'pinnis', khakhra, banana chips', badam, walnuts* in order to ensure I could keep my cognitive resources running. My thanks also go to Da, for the most basic yet necessary IT support, a laptop, but most importantly, for ensuring that I had my space to work on this PhD project. To my other family - Isa, Nati, abuela and tortilla; gracias por darme tanto amor (y comida!) y apoyo incondicional. Gracias por todo y más!

And last but not least, I must thank my rock, my anchor, my partner – Alvaro. This thesis would not have been possible without you, verdad? For being patient and listening to my thoughts and academic arguments, for asking questions to expand my perspective and spark critical thinking, for encouraging me to rest, for making me coffee and breakfast, for everything! Por todo! Gracias mi amor.

This thesis is dedicated to Medha aunty, Alvi, and my younger self.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Visual cognition is one of the key tools for infants to learn and acquire knowledge from their surroundings in the early years of their life. How infants distribute their looks not only tells us about what they are learning but also how they are learning. Indeed, their deployment of attention informs us of their learning strategies for constructing knowledge from their surroundings. Measures of visual cognition are important indicators of infant attention and speed of processing (Rose, Feldman, & Jankowski, 2002). Studies on infants' visual cognition from both lab and real-world settings have been associated with cognitive and behavioural outcomes such as memory formation (Rose, Gottfried, Melloy-Carminar, & Bridger, 1982), word learning (C. Yu & Smith, 2012; Markus, Mundy, Morales, Delgado, & Yale, 2000), and social interactions (Dawson et al., 2004).

Individual differences in visual cognition are typically measured in lab-based tasks. Despite being predictive of long-term outcomes, an open question is how these lab-based measures relate to the development of visual cognition outside the laboratory, for instance, during caregiver-infant interactions. Given that infants do not grow in a vacuum, we are specifically interested in understanding the role of caregivers in the development and deployment of their infants' visual cognitive abilities in their naturalistic context. This is an important issue to understand in the global health context, particularly in low and middle-income countries, where children's developmental status has been primarily characterised by gestational age and nutritional status rather than direct measures of cognitive function. Moreover, parent-infant interaction has been used as a tool for intervention as it is low-cost and scalable and has shown some promising outcomes (Aboud, Moore, & Akhter, 2008; Aboud & Yousafzai, 2015; Attanasio et al., 2014; Britto, Ponguta, Reyes, & Karnati, 2015; Jeong, Franchett, Ramos de Oliveira, Rehmani, & Yousafzai, 2021).

The core goal of the project entails exploring whether individual differences in visual cognition (specifically Visual Working Memory; VWM) in the laboratory are related to individual differences in visual cognition measured during caregiver-infant interaction in a low-resource context (i.e., rural region in the north of India). To examine this question, we measure visual cognition in the laboratory by using a preferential-looking paradigm. In addition, we develop innovative ways to measure visual cognition as a marker of early cognitive development outside the laboratory by using head-mounted eye-trackers and machine learning. We show how both of these approaches can be deployed in a low-resource setting.

1.1 Visual Cognition in the Lab

Research on visual cognition has noted the importance of measuring individual differences in laboratory-based tasks (Jankowski, Rose, & Feldman, 2001). Behavioural measures of visual information processing, such as fixation duration (sustained visual attention to the target) and shift rate (shifting attention between two stimuli), have shown to be reliable and stable indicators for assessing individual differences in visual cognition (Colombo, Mitchell, & Horowitz, 1988; Jankowski et al., 2001; Rose et al., 2002). The general consensus is that the length of fixation duration and disengaging attention to scan stimuli inform us of the time an individual takes to encode and process the information. In terms of infants' developmental trajectories, these measures have demonstrated a robust age-related change (Colombo, Mitchell, Coldren, & Freeseman, 1991). For example, younger infants tend to look longer at a target stimulus than older infants, suggesting that young infants are slower processors than older infants and that slower processors require a longer time to encode information (Colombo et al., 1991). Similarly, the frequency of shifting looks between stimuli is faster in older infants than in younger infants, suggesting faster visual information processing (Colombo et al., 1988, 1991; Frick, Colombo, & Allen, 2000; Rose, Feldman, & Jankowski, 2001).

Importantly, individual differences in the laboratory-based measures of visual cognition (e.g., fixation duration) during infancy are predictive of attentional and behavioural outcomes in childhood (Papageorgiou et al., 2014; White, Heck, Jubran, Chroust, & Bhatt, 2022). For instance, in a longitudinal study by Papageorgiou et al. (2014), results revealed that individual differences in infants' fixation duration were positively related to parent-reported measures of infants' effortful control, and negatively related to infants' surgency and hyperactivity-inattention in childhood. These findings show how individual differences in infants' visual cognition are predictive of later developmental outcomes. Thus, understanding the individual differences in visual cognition in infancy has broad implications.

A central part of visual cognition that develops early in life is Visual Working Memory (VWM). VWM comprises the mechanism of actively maintaining, storing, updating and manipulating information (Baddeley, 2012; Baddeley & Hitch, 1994) that underpins complex cognitive functions and behaviour (Moser et al., 2018). As adults, we use this system roughly 10,000 times a day, either to compare objects that cannot be viewed simultaneously or to detect changes in the environment (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). For example, while crossing the road, we look to one side and then the other in order to ensure safety. Doing this, requires holding the initial information (e.g., no vehicles are approaching on the right-hand side) in our mind for a short period of time, even while (and after) we shift our visual attention to look to the other side. Therefore, it plays an essential role in our daily functioning.

Individual differences in children's VWM abilities have been shown to predict both concurrent and future academic achievements such as reading, mathematics, science understanding and general intelligence (Bull, Espy, & Wiebe, 2008; Bull & Scerif, 2001; Holmes & Adams, 2006; Gathercole, Tiffany, Briscoe, Thorn, & The ALSPAC team2, 2005; Gathercole, Pickering, Knight, & Stegmann, 2004; Jarvis & Gathercole, 2003; St Clair-Thompson & Gathercole, 2006). Holmes and Adams (2006) examined, systematically, the contribution of working memory (WM) components on 8- to 10-year-old school children's performance in mathematical skills (e.g. mental arithmetic) as outlined by the National Curriculum England, while controlling for general number fluency. While all components of WM (central executive function, phonological loop, and visuo-spatial sketchpad) predicted unique variance in children's performance-related skills on curriculum-based mathematics abilities, results specifically indicated a strong role of visuo-spatial working memory in younger children's overall performance as well as older children's performance on difficult questions. Similarly, Bull et al. (2008) conducted a longitudinal study to examine primary school children's (4.5 years old) performance on a battery of cognitive tasks and relations to children's maths and general learning capacities. Notably, results revealed that across the first three years of primary school, children's VWM (measured using the Corsi-blocks task) was a significant predictor of their performance in maths. These results suggest that VWM predicts wider aspects of scholastic learning, particularly in young children.

On the other hand, deficits in VWM are associated with learning difficulties such as reading difficulty (Kudo, Lussier, & Swanson, 2015) and dyscalculia (Szucs, Devine, Soltesz, Nobes, & Gabriel, 2013), which further indicates the importance of VWM for broader cognitive and intellectual abilities. Critically, VWM is open to intervention (Holmes, Gathercole, & Dunning, 2009; Klingberg et al., 2005) and can be assessed as early as four months of age (Wijeakumar, Kumar, Delgado Reyes, Tiwari, & Spencer, 2019), making it a good candidate for early assessment and intervention. Given the predictive value of VWM (e.g. in scholastic achievement), understanding how the measures of VWM assessed in a laboratory setting relate to the real world can have important implications.

VWM has been reliably assessed and quantified in laboratory settings using the change detection task (Luck & Vogel, 1997). In this task, participants are presented with two arrays of items and are asked to identify if the items are identical across both arrays. Using the change detection task, research has shown that children's VWM capacity continues to develop from 1.5 items at 3 years of age to adult-like capacity by the age of 7 (Riggs, Simpson, & Potts, 2011; Simmering, 2012). In the case of laboratory research with infants, in which the explicit question of item equivalence cannot be asked, a modified version of the change detection task is used to assess VWM capacity, namely, preferential looking change detection (”VWM-PL” henceforth) (Ross-sheehy, Oakes, & Luck, 2003). VWM-PL is a simple task that looks at infant VWM in isolation from other influences. A recent review by Buss, Ross-Sheehy, and Reynolds (2018) noted that studies in infancy show a substantial increase in the VWM capacity from 6- to 12- months using the VWM-PL tasks (Oakes, Ross-Sheehy, & Luck, 2006; Ross-sheehy et al., 2003). Moreover, the task provides us with reliable behavioural measures of visual cognition that have been noted to underlie the individual differences seen in working memory capacity (Rose, Feldman, & Jankowski, 2011). For instance, increased speed of processing during childhood, measured using fixation duration, and shift rate, have been found to explain the age-related increase in working memory capacity (Fry & Hale, 1996; Kail & Salthouse, 1994). Given the suitability of the VWM-PL task for research with infants and the track record of successful research linking it to actual working memory capacity, we use the VWM-PL task to obtain the laboratory-based measures used in this project.

Experimental lab-based research is helpful to recreate the phenomenon of interest in a research context and allows us to control for multiple variables in order to establish causal relationships. In the context of VWM, it helps us isolate this specific cognitive mechanism. However, infants’ naturalistic environment is complex. In a real-world setting, infants often explore their surroundings along with their social partners, who may follow the infants’ focus but may also influence the process of exploration. Neither the dyadic engagement nor the busy environment are present in the highly controlled laboratory setting. Therefore, understanding whether the measures from laboratory research play the same causal role in the real-world context is a key point for the development of literature in the field.

A central goal of this project is to connect the research conducted in laboratories, which is expected to have real-world outcomes, *with* said real-world outcomes. For example, from lab-based studies, we know that the infant holds information and discriminates between two stimuli (e.g., VWM-PL task). These processes are interpreted as being connected to real-life scenarios such as differentiating between, say, a toy duck and a cup. While experiments in the laboratory and naturalistic setting vary in their limitation and strengths and answer different questions (McCall, 1977), relating the two contexts allows the development of a more holistic and deeper understanding of development (Dahl, 2017). Therefore, in this thesis, we take a step to connect the same laboratory-based performance to the actual behaviour of the infant along with their caregiver in a real-world context.

1.2 Visual Cognition in the Lounge

Parallel to laboratory research, an increasingly large number of studies have been conducted in infants' naturalistic settings, such as the home environment, to understand what elements of parent-infant interaction aid infants' learning and development (e.g. Fausey, Jayaraman, & Smith, 2016; Karasik, Tamis-LeMonda, Adolph, & Bornstein, 2015; Lamm et al., 2014; Tamis-LeMonda, Custode, Kuchirko, Escobar, & Lo, 2019; West & Iverson, 2017). Caregiver-infant interaction has been shown to fuel infant's early learning experience such as in word learning (Markus et al., 2000; Mundy et al., 2007; C. Yu, Suanda, & Smith, 2019), attentional control (Niedźwiecka, Ramotowska, & Tomalski, 2018), social skills (Landry, Smith, & Swank, 2006), and cognitive development (Mundy & Newell, 2007; Tamis-LeMonda, Shannon, Cabrera, & Lamb, 2004).

A well-known aspect of early caregiver-infant social interaction that has been of interest to developmental psychologists is Joint Attention. Joint attention refers to the length of joint fixation by caregivers and infants on the same target in between attentional switches (Abney, Suanda, Smith, & Yu, 2020; C. Yu et al., 2019). Parent-infant dyads use joint attention (also labelled as ”joint triadic interaction” when objects are involved in the dyadic interaction; Little, Carver, & Legare, 2016) during their daily interactions to attend to the same objects or events. Such coordinated attention with caregivers forms the basis for infants' understanding of goals and intentions (Charman et al., 2000) and underpins word learning (C. Yu & Smith, 2016) by providing key 'object-word' mapping opportunities necessary for learning new words (Baldwin & Markman, 1989; Mundy et al., 2007). For instance, joint attention supports word learning by extending infants' attention to the shared object of interest (C. Yu & Smith, 2016; C. Yu et al., 2019).

The influence of contingent responses to infants extends beyond the development of language (e.g. development of social skills; Mundy & Newell, 2007). Early intervention studies have shown that parents who were trained to notice and "follow in" on their infants' attention and hold their infants' attention to the infants' object of interest had a positive impact on their infant's language, cognitive and social development outcomes (Landry et al., 2006; Landry, Smith, Swank, & Guttentag, 2008).

It is important to note that in the VWM-PL task, the measures of length of fixation and frequency of shifting are representative of individual attentional focus. Thus, when connecting lab and lounge measures, it is important to consider the relationship between each lab measure (e.g., fixation duration) and its lounge equivalent during bouts of joint attention as well as outside of them.

Even though there is a long research tradition focusing on the study of visual cognition in naturalistic settings, to date, most of our understanding of joint attention and its systematic relationship to infants' cognitive outcomes almost exclusively comes from highly educated, high-income countries (Bard et al., 2021). Given that joint attention is interactive, and thus social in nature, considering socio-cultural influences on it, such as those stemming from variations on caregiving norms (e.g. Abels & Hutman, 2015; Bard et al., 2021; Little et al., 2016), is of particular importance.

Episodes of joint attention differ qualitatively and quantitatively across cultures (see Bard et al., 2021). For instance, Salomo and Liszkowski (2013) studied the frequency and occurrence of joint triadic attention in natural unstructured interaction between caregivers and their infants in Chinese, Mayan, and Dutch communities. They found that Mayan infants spent significantly less time in joint attention than dutch infants and that Chinese infants spent the most time in triadic joint attention with their caregivers.

Another study by Little et al. (2016) compared triadic object exploration between parent-infant dyads from the rural subsistence farming community in Vanuatu and infants from urban California in the US. They did not find a cultural difference in time spent in triadic object exploration or contingent responsiveness, but they did find significant cultural differences in the modality of engagement. Dyads from the US were more likely to engage in visual triadic engagement characterised by caregivers placing themselves in a way that allowed face-to-face interaction and mutual eye contact with both caregivers and infants being able to alternate attention towards their social partner and the objects. Conversely, caregivers in Vanatanu were more likely to engage in physical triadic interaction, characterised by physical contact with the infant and the object at the same time as well as tactile stimulation. In terms of infants' attention in these dynamic interactions, the results revealed that caregivers' actions on a target object were preceded by their infant looking at the target significantly more often for the dyads from the U.S. compared to the Vanuatu dyads. Clearly, there are cultural influences in some aspects of shared attention that may reflect the caregiving practices that shape how the parent-infant dyads engage with each other and their surroundings. Therefore, there is a need to understand the social and cultural context in order to understand the interactions between parents and their infants.

1.3 Cultural Elements in Caregiver-Infant interaction

An important point to consider when addressing cultural variations in parenting practices and dyadic interactions is that caregiver-infant interactions underscore the process of transmission of socio-cultural characteristics such as values, beliefs and goals from the adult environment to the child's (Kagitcibasi, 2005; Keller, Borke, Chaudhary, Lamm, & Kleis, 2010). Parenting styles are often influenced by the cultural expectations linked to social norms (i.e., what is typically done as well as what should be done). For instance, mothers in most Indian communities fear the dangers of the *"Evil Eye" or "Najar"*. The fear of the evil eye stems from the belief that there is a form of supernatural power within the gaze of others (potentially seen as a malicious gaze with ill intention toward the individual). The evil eye is social in nature and is connected to expected feelings of envy and jealousy among others. Therefore, mothers may not display affection, or praise their children, in public. By doing so, there is an expectation that they will minimise envy and the risk of being harmed by the evil eye. Additional practices to protect children from the evil eye include blacking infants' eyes with kohl (also known as *"kajal"*), tying black thread around the waist or left leg, and/or wearing a bracelet (Spiro, 2005). Therefore, how the caregiver interacts with the infant would be dependent on what they think is the best for their child with reference to the society in which they live.

With their extensive work on parent-infant interactions across cultures, Keller and colleagues have shown two distinctive types of parenting: proximal and distal style (Keller et al., 2009, 2004). The distal style is characterised by more face-to-face interactions with extensive verbal input, and is primarily a system of the Western, Educated, Industrialised, Rich and Democratic (WEIRD) societies (Henrich, Heine, & Norenzayan, 2010). In contrast, the proximal style focuses heavily on body contact and motor simulation and is typically seen in non-WEIRD, traditional societies e.g. rural India and sub-Saharan Africa (Keller, Borke, Lamm, Lohaus, & Dzeaye Yovsi, 2011; Lamm et al., 2015). A key point is that the different parenting styles are considered to be adaptive to their surroundings (see Kagitcibasi, 2005). Taking into account proximal and distal parenting styles allows a more nuanced understanding of caregiver-infant dynamics observed in research.

Although there is a general consensus that, overall, caregivers contingently respond to their infants across cultures, there are differences in the said responses (Broesch, Rochat, Olah, Broesch, & Henrich, 2016; Kärtner et al., 2008). In their research, Kärtner, Keller, and Yovsi (2010) analysed the emergence of cultural-specific contingency patterns during mother-infant interaction in German and Nso communities. They measured the mother's contingent response (auditory, visual and proximal) to their infants' non-distress signals at five time points with two weeks gap. Results indicated that mothers from both cultural groups responded contingently to their infant's non-distress signals. However, variations in the form of responses across both groups were seen from 2 months of age. Infants from Germany tended to experience significantly more visually contingent responses from their caregivers compared to the Nso infants. In contrast, infants from Nso experienced more proximal (tactile) contingent responses from their caregivers. Interestingly, over time, caregivers in Germany showed a linear increase in visually contingent response and a decrease in proximal contingent response towards their infants. However, Nso caregivers showed consistency in the modality of response (primarily proximal and auditory) throughout infants' development. These culturally specific interactions between the infant and their caregivers embody goals and beliefs that reflect cultural values, e.g., ethnographic work with Nso mothers notes the importance of bodily closeness and contact as central to sensitive parenting and to child development (Goncu, 1999; Kärtner et al., 2010; Keller et al., 2005; Yovsi & Keller, 2007).

Further research by Chavajay and Rogoff (1999) found that caregivers and their 14-to-20-month-old toddlers managed their attention in different ways in Guatemalan Mayan and U.S. European-descent families during a time-sharing activity which was recorded during a home visit. U.S European-decent caregivers and toddlers tended to alternate their attention between events (e.g. stop a conversation with an adult to attend to their toddler and then get back to the conversation with the adult), whereas parents and toddlers from the Mayan community simultaneously attended to multiple events (e.g talk to the interviewer while playing with the child). The authors speculated that parents attending to several simultaneous events may rely more on the verbal channel to communicate with adults and non-verbal channel (e.g. body simulation) to communicate with toddlers (see also Rogoff et al., 1993). Their research suggested that parents may deploy their attention in ways that adhere to what the community sees as desirable or valuable during social interaction. Given that parent-infant interactions embody goals and values that are reflective of socio-cultural demands (Keller, 2018; Keller & Demuth, 2007), not only should there be social outcomes to dyadic interactions (e.g., infants learn social values), but the social context and values should be expected to shape the interactions themselves. Critically, through them, the social context should be linked to the outcomes of the dyadic interaction.

Although research can be found outside WEIRD populations, its focus is often ethnographic (Yovsi & Keller, 2007; Keller et al., 2003; Rogoff et al., 1993) and anthropological (Lancy, 2007, 2014), with psychological research lacking the comprehensive and systematic focus it has had in the west. Theory and research rooted in western views and values risk remaining detached from the socio-cultural realities for which interventions are devised. Given that cultural variation in parent-infant interactions is embedded in cultural values and expectations of what is best for their child, simply applying western research in a non-western setting that differs in cultural values can result in varied, weakened, and even poor outcomes. Critically, although there are variations in wealth and access to resources within nations, socio-economic differences across nations must be acknowledged. That is, research carried out in developing countries will not only reveal cultural dimensions but also the influences of available resources (e.g. access to electricity, running water, etc.).

When it comes to interventions in developing countries, psychological research can be critiqued for diverting attention from the socio-structural elements (e.g., poor access to education and poverty) which underlie developmental outcomes and fall outside of the reach of individual and family-based interventions. Thus, ethical concerns have been raised in relation to research and interventions in low- and middle-income countries that fail to understand and engage with the communities, contexts, and values which shape their needs. In the worst-case scenario, implementing interventions devoid of contextual understanding can be harmful to the community (Morelli et al., 2018; Scheidecker, Oppong, Chaudhary, & Keller, 2021). Thus, researching parent-infant interactions requires situating them within the context in which they take place, considering both the available resources and the norms of the society (Rogoff, Dahl, & Callanan, 2018).

Despite limitations, parenting intervention programmes in low and middle income countries have resulted in positive effects on direct measures of children's cognitive and language outcomes (Aboud & Yousafzai, 2015; Britto et al., 2015). For example, Andrew et al. (2020) conducted a cluster randomised controlled intervention in urban-slum India with the aim of increasing and improving caregiver-child interactions in order to support child development. The primary caregivers were trained using a structured curriculum which encouraged the caregiver to respond to their child's actions and verbalisation. Outcomes of the intervention included cognitive, language, and motor development measured through scores in the ASQ-3 and Bayley-III. Overall, children performed better on the cognitive tasks (e.g exploration and early memory) and showed marginal improvements on expressive and receptive language scores. However, there was no significant impact on motor skills.

Additionally, a recent review on the effect of a parenting intervention on infants' developmental outcomes in both low and middle-income as well as high-income countries found that not only did parenting interventions have a significant effect on infants' cognitive, language, and motor development but that the effects on cognitive outcomes were three times higher in the low and middle-income countries compared to the high-income countries (Jeong et al., 2021). This was particularly the case for interventions that included responsive caregiving content. Moreover, the benefits of interventions extended beyond outcomes on parent-infant interaction to parent-level outcomes (e.g., parenting knowledge and practices; also see Jeong, Pitchik, & Yousafzai, 2018). Given that parenting interventions aimed at enhancing parent-infant interaction are not just low-cost but also scalable (Grantham-McGregor et al., 2020) and have yielded benefits to infant cognitive development, it is crucial to understand the mechanisms through which infants' social context impacts their cognitive development.

The literature consistently indicates the importance of parent-infant interaction by highlighting the multiple positive outcomes of interventions, but it also portrays the need for caution when designing and carrying out interventions across settings and invites further systematic research outside western nations to inform said interventions. Understanding parent-infant interactions and their outcomes in a more systematic way (e.g., accounting for individual differences in caregiver visual cognitive abilities as well as socio-cultural contexts) can help inform interventions that better adhere to the needs of specific families within concrete contexts (see Morelli et al., 2018). The present project was carried out in a low-resource rural population (Shivgargh, Uttar Pradesh, India), so an overlap between cultural and wealth dimensions is inevitable. Moreover, given that this is an understudied population, the goal of gaining a comprehensive understanding of the relationship between maternal and infant visual cognition requires that we are cautious and adopt an exploratory approach (i.e., not assuming the universality of findings from research on other socio-cultural settings).

1.4 Connecting Lab-and-Lounge

Having addressed the lab and lounge aspects of studying visual cognition separately, the natural step involves addressing how the two settings can be connected. Dynamic Field Theory (DFT) can provide powerful insights into this link (Spencer, Perone, & Johnson, 2009). The theory provides a framework for embodied cognition and suggests that changes in basic neural function are related to developmental capacity growth. Specifically, we can look at the DF model of autonomous visual exploration by Perone and Spencer (2013a) to understand the relationship between VWM in the lab and lounge (also see Spencer & Schoner, 2003).

The DF model of visuospatial cognition consists of layers of feature-related excitatory neurons coupled with inhibitory neurons. The model consists of two components of the neurocognitive system, a perceptual field and a working memory field. In response to an input (e.g. stimuli), neurons in the perceptual field interact with one another, thus encoding the information. Encoding within the perceptual field leads to continuous fixation and the formation of working memory. That is, activation in the perceptual field leads to excitation in the reciprocally coupled neurons in the working memory field. Once the activation of the working memory reaches the required threshold, it suppresses the activation in the perceptual field and results in releasing fixation from the currently attended item. Importantly, activation peaks in the working memory field are stronger than in the perceptual field and can self-sustain even when the input is removed following the shift in attention (see also Perone & Spencer, 2013b). A final element in the model enables learning. In particular, when working memory peaks form, they boost the strength of a localised memory trace (akin to strengthening weights in a connectionist network). The memory trace, in turn, boosts the level of activation locally in the field, effectively 'priming' previously visited feature values. This has the consequence of leading to more robust working memory peaks over time as the model (or the infant) learns about visual object features.

Using these mechanisms, Perone and Spencer (2013a) modelled infants' visual experience in the real world and connected it with infants' performance in the laboratory-based task developed by Rose et al. (2002). Term and preterm infant models were placed in a virtual world which included multiple items to which the infants were exposed over months. The process mimicked how parents and infants play together with multiple toys on repeated occasions over time, thus accumulating visual experience "outside the lab". Over the simulated time period, and at repeated points, virtual infants were modelled to perform the preferential-looking task (specifically, a speed of processing task). Much like infants in the real world (see Rose et al., 2002), results captured the infants' looking behaviour and replicated the developmental shifts in visual processing speed in real-world infants. They also replicated the developmental delays found in preterm infants (Rose et al., 2002). The authors explained that the developmental shifts in looking behaviour and speed of processing (as well as developmental lag for preterm infants) resulted from the activation peaks formed in the working memory field, supported by the build-up of memory traces over longer-term learning. Over the DFT developmental trajectory, infants tended to form working memory peaks increasingly fast as memory traces became stronger, thus leading to shorter looking time and more switches in the preferential looking task (Perone & Spencer, 2013a).

Having successfully replicated Rose et al. (2002) research, the authors created an intervention model for their preterm infant based on the intervention study by Landry et al. (2006). Landry et al. (2006) found that, caregivers who were trained to follow in on their preterm infant's object of attention and hold the infant's attention to the object (e.g. by manipulating the object) enhance cognitive and social outcomes for their infants. Perone and Spencer (2013a) modelled the preterm infant to fixate on an item and introduced a "caregiver" to sustain the infant's attention to that item. Sustained attention in the model was achieved by increasing the input from the object (e.g., hands manipulating the object that holds the infant's attention). By 12 months, the looking behaviour in the preterm model was comparable to the term model, with the outcome being attributed to robust peaks in the working memory field following the intervention, supported by stronger memory traces.

Although the work of Perone and Spencer (2013a) did not explicitly model data from real-world settings, it provides insights into the processes that underpin the development of working memory and enables us to connect visual exploration across lab and lounge. Critically, by modelling the caregiver as following in on their infant, and linking this to the development of the neural network and cognitive outcomes, they invite considering elements of joint attention (particularly when led by infants) in the development of VWM and deployment of visual attention in the real world and laboratory settings.

1.5 Leveraging technological advances in a low resource setting

More than 200 million children, under the age of five, in low- and middle-income countries, fail to reach their expected level of cognitive outcomes due to environmental risks such as poverty, stunting, sanitation and lack of early learning opportunities (Grantham-McGregor et al., 2007; McCoy et al., 2016; Black et al., 2017). India accounts for 65 million children (Grantham-McGregor et al., 2007) failing to reach their potential cognitive outcomes. To address this, the 2030 Agenda for Sustainable Development by the United Nations (2015), calls for urgent actions by developed and developing countries, and donors to form global partnerships to achieve the Sustainable Development Goals (SDG). The goals are considered central to inclusive and sustainable growth in all countries. One of the key issues put forward in this global agenda entails tackling and improving early child development. Thus, developmental science has been encouraged to create objective assessments to understand child development in the local and global contexts in order to generate knowledge for diverse contexts. Using experimental and longitudinal studies to track progress has become of great importance to help create evidence-based programs and inform policies linked to SDG.

Studies on infancy or early child development in low-and-middle-income countries have often been impeded by a lack of suitable methods that can be adapted to low infrastructure and resource settings (e.g. Milosavljevic et al., 2019). For instance, most observations of caregiver-infant interactions have been made by recording free-play sessions in the infants' home environment and are largely based on manual coding or qualitative assessment. Not only does manual coding demand a high investment of time but recordings from a third-person perspective preclude identifying key information which is only available from a first-person view (see L. B. Smith, Yu, & Pereira, 2011). Thus, when considering the deployment of visual cognition in naturalistic settings, a limiting factor for research in low-and-middle-income countries has been the limited availability of tools that can offer objective insight into infant cognitive development (Katus et al., 2019).

Despite traditional limitations, recent technological advances have made it possible to research infants' cognitive development in their naturalistic setting by using head-mounted eye-trackers (e.g. Fausey et al., 2016). Eye trackers can be useful to capture infants' low-level responses. Moreover, combined with a caregiver-infant play session, they can provide a unique and dynamic insight into infants' own visual experiences while interacting with their caregivers, and choosing their visual scene (L. B. Smith, Yu, Yoshida, & Fausey, 2015).

The data for the current thesis comes from a rural region in the state of Uttar Pradesh (UP) in India, a state with 75% of population living in rural area. In addition to cultural dimensions (already discussed), it is essential to note that UP is the most populous state in India and scores among the worst human development index (Dettrick, Jimenez-Soto, & Hodge, 2014). There are significant caste based disparities in the region such that essential services like healthcare often does not reach to the poor and the lower caste (Dettrick et al., 2014; Subramanian et al., 2006). Although healthcare is free India, they are often seen of "low quality" and thus low-cost private school and unqualified private "doctors" are preferred (Willis et al., 2011). Therefore, our research not only taps into socio-cultural elements in parenting but is inseparable from the available socio-economic resources in the region. The setting also informs the goals of the project. Participants in the current project were recruited from Shivgarh, Raebareli District in UP. In recent years, there has been community level interventions by community workers and members aimed at promoting new-born care practices such as skin-to-skin care and breastfeeding from first day of birth leading to a decrease in neonatal mortality rate by 54% (Kumar et al., 2008; Tinker, Parker, Lord, & Grear, 2010). The area has one of the lowest girl-to-boy ratio in India. Although there is a push toward improving the nutritional status of children, the aspect of cognitive development is poorly understood. In our research not only do we use mobile head-mounted eyetrackers and cameras to capture infants and their caregivers' visual fields but also develop and validate a machine-learning pipeline to extract meaningful measures of visual cognition from caregiver-infant interaction. This not only enables us to connect measures from lab and lounge settings but allows us to gain insights into development within context.

1.6 Thesis Overview

Given the massive brain plasticity in infancy, the first 1000 days of infants' development are of great importance for their cognitive and social development. During this period, the rapid development of the brain can be vulnerable to harmful exposure such as poverty as well as receptive to positive stimulation (Jensen, Berens, & Nelson, 2017). This makes the first two years of life a critical period to understand how the infant's surrounding contributes to their development. As argued in the literature review, a key part of cognition that rapidly develops during this time is Visual Working Memory (VWM). Behavioural studies have shown that individual differences in VWM tasks are predictive of later achievements (Rose, Feldman, & Jankowski, 2012). Moreover, it can be assessed as early as the first 3-4 months of an infant's life (Ross-sheehy et al., 2003; Wijeakumar et al., 2019), making it a good candidate for early assessment. Work by Ross-sheehy et al. (2003) has noted that infants' show an improvement in their VWM capacity between the ages of 6-and 8-months with 6.5 months infants performing above chance when the memory load condition has one item. On the other hand, 10- to 13- month old infants show above chance performance in the VWM capacity for two and three item memory load (for similar age trends on visual working memory linked to spatial location and changes in configuration of stimuli also see Oakes, Hurley, Ross-Sheehy, & Luck, 2011). Therefore, the current project examined the visual cognition with 6- and 9-month old age cohorts in our sample. Henceforth, we use a modified version of the preferential-looking task developed by (Ross-sheehy et al., 2003), to isolate and measure VWM in infants in the lab, and head-mounted eye-tackers and our machine learning pipeline for constructing measures in the lounge.

Chapter 2 includes two studies which aim to understand whether parents' and infants' visual cognition in the laboratory are related to one another. Moreover, the studies explore the connections between the VWM capacities and socio-economic and socio-emotional measures. Knowing parents' abilities can have important implications as they might modulate the effectiveness of, for example, parent-based intervention. Studies 1 and 2 serve to tackle the parents' side of the equation by tapping into their visual cognition and understanding what characteristics of the parent are linked to their infants' VWM. Chapter 2 further provides laboratory measures of VWM-PL, enabling the exploration of relationships between them and measures from dyadic interactions in later stages of the thesis.

Chapter 3 focuses on establishing how recent technological advances can be used to quantify caregiver-infant interaction through low-cost and less labour-intensive methodologies in a low-resource contexts (i.e., a rural region of northern India). Previous research in western, high-resource settings has successfully used head-mounted eye-trackers for studying caregiver-infant social interaction in their naturalistic setting. However, to our knowledge, no previous study has used these tools in a non-WEIRD context. Although some researchers have made an effort to capture caregiver-infant interaction beyond the WEIRD context using hand-held or tripod-mounted cameras (Abels, Papaligoura, Lamm, & Yovsi, 2017) or observations (Roopnarine, Ahmeduzzaman, Hossain, & Riegraf, 1992), the problems of coding frame by frame in an objective manner have remained unresolved. Indeed, the laborious task of frame-by-frame coding, has inevitably limited sample sizes in research. Thus, chapter 3 addresses the use of low-cost head-mounted eye-trackers and cameras for collecting first-person dyadic data in both high (urban UK) and low (rural India) resource settings. The chapter further develops a methodological pipeline by using freely and openly available machine learning algorithms to quantify parent-infant interaction. The machine learning pipeline is trained and validated with data from both UK and India. We also describe the process of extracting meaningful measures of visual cognition that can be used to examine the deployment of attention in dyadic interactions (Chapter 4) and the relationship between infants' visual cognition from lab and lounge in chapter 5.

The inclusion of the high resource sample in the UK served as a point of reference for testing the use of the chosen technology within a known and controlled setting (e.g., ensuring consistent access to electricity, and awareness of technology amongst participants) in parallel to deploying the use of technology in the low resource setting. In addition to resources, including both an Indian and a UK sample allowed testing the machine learning pipeline a greater variety of facial features and objects (toys). Regarding the chosen age for UK infants, even though having both 6-month and 9-month-old UK participants would have served to establish clearer comparisons across settings (of relevance to Chapter 4), practicalities in the project precluded doing so. It must be noted that this project is a part of a larger project with multiple lines of enquiry. The age group of 6 months old was selected as a developmental baseline of value for all research streams in the overall project. A single age group was sufficient for testing and developing the machine learning pipeline in a high resource setting and for enhancing variety of imagery. In terms evaluating our data across settings (Chapter 4), this baseline was considered sufficient (though not ideal) for initial explorations of trends, which provided a point of comparison at the earliest age.

Following the description of the methodological pipeline, in Chapter 4, we use our pipeline to extract measures of deployment of attention from a larger dataset including three groups of participants – dyads with 6-months old infants in the UK, and dyads with 6-months infants in India, and dyads with 9-month-old infants in India. Including a larger data-set serves to 1) validate the proposed machine learning pipeline from Chapter 3, and 2) understand the similarities and differences in the deployment of visual cognition across socio-cultural context during caregiver-infant interaction. Previous research suggests that parenting strategies and behaviour are reflective of the cultural models of the society (Goncu, 1999) as well as what is deemed appropriate for infant development within that society (Harkness & Super, 2020; Keller, 2007; Rogoff et al., 2018). Therefore, chapter 4 uses the eco-cultural model of parenting (Keller, 2007) to understand the visual dynamics of parent-infant interaction in each cultural context.

In the final empirical chapter (Chapter 5), we integrate the data from previous chapters to explore a holistic picture of infants' development of visual cognition in India. In it, we explore whether and how caregivers' and infants' visual exploratory behaviour in the lounge relates to their visual cognition measures assessed by using a standardised VWM-PL task in the lab.

It must be noted in regards to sample sizes across our empirical studies, that decisions were made to ensure, to the greatest degree possible, an appropriate statistical power across research projects. We expected some loss of data as a result from participants engaging in the numerous tasks required for the different studies, so large recruitment targets were set with the goal of offsetting the impact of potential data loss. Ultimately, analyses were carried out with the available data. (IS THERE AN INITIAL RATIONALE FOR A TARGET NUMBER?).

In Chapter 6, we synthesise the findings of our research and discuss its contributions to the literature. We place our findings and interpretations in the context of low-resource settings and in connection to cultural elements. Last, we highlight the real-life implications of our findings and consider how the insights provided by our research can contribute to the development of interventions in the context of our population.

Chapter 2

Examining the relationship between mothers' and infants' visual cognition

2.1 Introduction

Throughout Chapter 1, we have established the importance of VWM and the development of research that connects laboratory and real-life outcomes. The role of caregivers was highlighted as an important influence on early infant development, as the latter does not happen in a vacuum. Chapter 2 serves to establish the empirical base for this project, which was carried out in a rural region of northern India (Shivgarh, Uttar Pradesh). Studies 1 and 2 explore the role of caregivers' visual cognitive abilities and whether (and, if so, how) these abilities predict infants' early visual cognitive development. Given the variability in the effectiveness of interventions in low- and middle-income countries (Aboud & Yousafzai, 2015; L. Zhang et al., 2021), and the potential for harm when implementing interventions without understanding the context of the community (see Morelli et al., 2018), a central approach to this thesis entails considering individual differences and contextual characteristics of caregivers and infants. Thus, Study 1 places the visual cognitive abilities of caregivers in relation to the caregivers' socio-emotional characteristics (i.e., depressive thoughts, intimate partner violence, and sense of empowerment) and their social context (i.e., socio-economic status). Study 2 builds on Study 1 by exploring the links between parental visual cognition, as well as their socio-emotional characteristics and social context, and infants' visual cognition.

2.2 Study 1: The Socio-emotional and Visual Cognitive Characteristics of Mothers in Rural India

The present study sets the base of later research by contextualising maternal VWM and considering its links with depressive thoughts and intimate partner violence (IPV), including elements of control and empowerment. Research has shown that individuals with depressive thoughts spend a significant amount of time processing negative information relative to other information (Levens & Gotlib, 2010; Joormann & Gotlib, 2008; Harvey et al., 2004). This limits the individuals' ability to rehearse goal-relevant information in working memory as they may find it hard to expel negative thoughts that arise from a negative mood state – thereby, leading to difficulty in attending to and processing new information (Gotlib & Joormann, 2010). Given that caregiving requires multi-tasking between several environmental stimuli (e.g. responding to the infants' cues while integrating them with environmental demands), it is important to address how maternal depressive thoughts may relate to her working memory. However, care should be taken to not look only at depressive thoughts, as it risks leading to reductionism.

The Indian state of Uttar Pradesh (UP) is characterised by high levels of IPV (Ackerson & Subramanian, 2008) against women. Not only does IPV have a severe negative impact on women's physical health (Krug, Mercy, Dahlberg, & Zwi, 2002) but also on mental health, leading to heightened stress, depression, anxiety and PTSD (Malik, Munir, Ghani, & Ahmad, 2020; Chandra, Satyanarayana, & Carey, 2009). These factors, in turn, can alter cognition. For instance, a study by Gonzalez, Jenkins, Steiner, and Fleming (2012) addressed the association between maternal early adversity experience (i.e., stress) and maternal sensitivity during parent-infant interaction. Results indicated that poor performance of spatial working memory mediated the association between maternal stress (measured using salivary cortisol samples) and lower maternal sensitivity during mother-infant interaction. The prevalence of IPV within the population, together with the complex overlap between IPV, stress, depression (Malik et al., 2020), and visual cognition, require exploring the unique influence of IPV on VWM.

Within current western perspectives in both research, (e.g., Dutton & Goodman, 2005; Hamberger, Larsen, & Lehrner, 2017) and policy (Home Office, 2018), understanding of IPV has been extended to account for elements of coercion, control of aggressors over victims, and victim's lack of empowerment. However, within the cultural setting of our research, empowerment among women is not necessarily the norm (Cunningham, Ruel, Ferguson, & Uauy, 2015; Sethuraman, Lansdown, & Sullivan, 2006). Spousal control over income, household decisions, and child-rearing practices (Cunningham et al., 2015) can be compounded by other (sometimes coercive) influences from the broader family, including in-laws and extended family (Sethuraman et al., 2006). Thus, within the context of our research, the study of IPV cannot be separated from elements of maternal empowerment such as physical mobility, contributing to household decisions, having access to money, and feeling respected within the household. Thus, Study 1 extends the literature by moving beyond depression and mental health to also investigate the predictive influence of related, yet distinct, variables (i.e., IPV and sense of empowerment).

The goal of a holistic understanding of maternal VWM also invites addressing the influence of socio-economic status. There is an awareness that infants in SES-deprived contexts are at risk of not fulfilling their developmental potential. In relation to VWM, research has shown the positive impact of higher SES on cognitive development. An essential point to bear in mind is that adults within these deprived settings have already gone through an adverse developmental trajectory. Indeed, research with older adults has shown the benefit of education (access to which is central to the socio-economic status) for working memory, particularly for females (Pliatsikas et al., 2019). Thus, considering the impact of their socio-economic status (SES) on individuals' visual cognitive capabilities is imperative.

A further key factor that has been associated with VWM is the individual's age (Brockmole & Logie, 2013). VWM abilities peak in the early twenties and then show a steady decline (Brockmole & Logie, 2013). The negative association between age and working memory has been found to be moderated by gender, with male participants experiencing a greater decline over the years (Pliatsikas et al., 2019). Thus, Study 1 looks at the influence of age on VWM in a confirmatory manner (i.e., greater age, lower VWM) and at its interactions with key variables of interest (i.e., depressive thoughts, IPV, and sense of empowerment), in an exploratory manner.

We begin our empirical work by describing our sample through key demographic variables. We then deepen this understanding by exploring the demographic influences (SES, maternal age, and infants' gender) on the socio-emotional variables of interest (i.e., depressive thoughts, IPV, and sense of empowerment). In other settings, evidence has suggested links between lower SES and higher postnatal depression (Shivalli & Gururaj, 2015) as well as between lower SES and experiencing more domestic violence (LaBore, Ahmed, Rizwan-ur-Rashid, & Ahmed, 2021). However, our population of interest lives in an already deprived context, with differences in SES being relative to one another (rather than comparable with high or low SES status in other regions of India or countries). Thus, we explore whether relative SES within this context is associated with our socio-emotional variables of interest in an inductive manner (with no a-priori hypotheses). The same rationale guides the inclusion of the gender of the participants' infant. Research in India has linked the infant's gender to experiences of depression and IPV (Nongrum, Thomas, Lionel, & Jacob, 2014; Patel, Rodrigues, & deSouza, 2002; Savarimuthu et al., 2010) with mothers of female infants experiencing more depression and IPV, which we expect will replicate in our research. Based on the overlap between empowerment and IPV, infant's gender is expected to predict empowerment as well (mothers of female infants feeling less empowered). However, the research that we have addressed focuses on Southern states in India, which have their own socio-cultural characteristics. Thus, although mothers of female infants can be expected to experience more IPV and depression and feel less empowered, cultural variability across India requires caution.

Following this, Study 1 characterises mothers' visual cognition relative to the maternal demographic and socio-emotional characteristics. Evidence of the developmental trajectory of VWM supports the prediction that the older that mothers are, the lower that their VWM will be. Similarly, the literature previously described suggests that depressive symptoms and experiences of domestic violence will predict maternal VWM negatively, whereas SES and feelings of empowerment will predict it positively. Such predictions, however, are tentative due to the lack of research in this region. It also must be noted that the VWM-PL task includes a variable of task difficulty (i.e., load, see further in Method) which was expected affect scores (higher difficulty leading to lower scores; Simmering, 2016b). Task load was also considered in interaction with variables of interest (age and socio-emotional characteristics) given that expected trends might only appear at specific difficulties (e.g., when the task is too simple, age-related variations in VWM may not show in the VWM-PL scores). We explored further interactions between predictors of interest in a purely inductive exploratory manner (e.g., socio-emotional measures interacting with age). Through this approach, we develop a comprehensive and holistic understanding of our sample of mothers and their cognitive abilities. This work sets the stage for Study 2, which examines whether (and how) maternal visual cognition is associated with individual differences in infants' emerging VWM abilities.

2.3 Methods

2.3.1 Participants

Study 1 was part of a larger longitudinal research project on early brain development carried out in Shivgarh, UP, India. The project was carried out in collaboration with a local not-for-profit organisation, the Community Empowerment Lab (CEL), who carried out the recruitment. Two hundred and forty families who were from, and lived in, the Shivgarh block were enrolled in the project. Families were initially screened as belonging to either 'high' or 'low' SES based on their Educational status. Those with both parents having greater than 10 years of education were classified as high SES, and those with both parents having less than or equal to 5 years of education were classified as low SES. This was based on a concordance between years of education of both parents and SES in the same area in previous studies, where parental education was shown to be concordant with the first and last wealth quintiles generated by using a principal component analysis conducted on a DHS SES questionnaire (I. Ahmed et al., 2018). Note that 'high' and 'low' are relative terms referring to our sample.

Additional demographic information such as household income, caste, religion, and resources including access to electricity, type of toilet, cooking fuel, etc was also collected. Table 2.1 shows the sample demographics. This information is essential for the understanding of the sample in our research.

Families completed two visits to the laboratory. The first one (Year 1), when the enrolled infants were 6 or 9 months of age, and the second (Year 2), when infants were 18 or 21 months of age. Data for the present study were collected in Year 2 (N = 236) and distributed across 3 rounds of data collection (in August and November 2018 and February 2019). The final sample for this study was determined by the tasks completed by the caregivers (all being mothers of the infants). A hundred and eighty-seven mothers completed the Edinburgh Postnatal Depression Scale (EPDS), and 138 completed the IPV and sense of empowerment interviews. Additionally, 136 mothers participated in the VWM task (72 were mothers of 18-month-old infants; 64 were caregivers of 21-month-old infants). Eleven mothers were excluded from the VWM analyses due to technical issues, electricity cuts, and distractions from their infants. Thus, the final VWM analyses included a sample of 125 mothers. In total, 120 mothers completed all the interviews and the VWM-PL task. Figure 2.1 shows a breakdown of participants for each stage of the study.

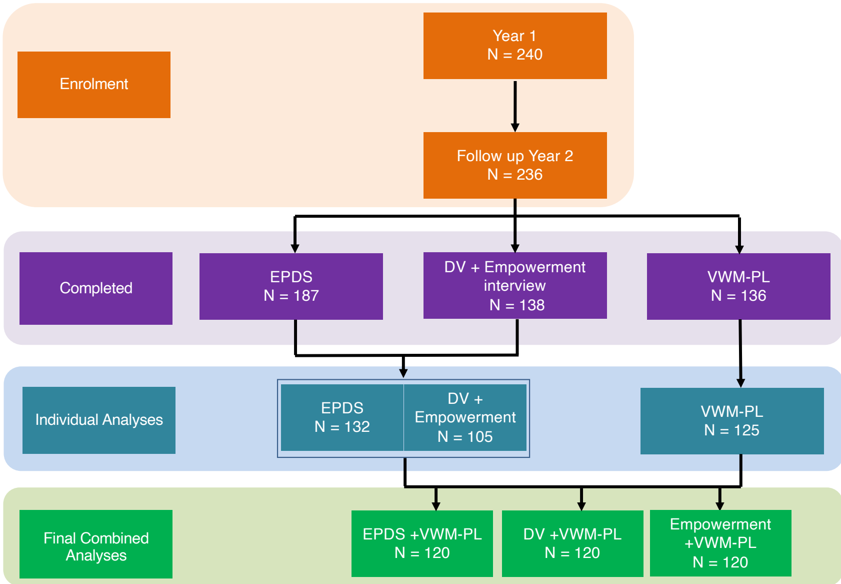


Figure 2.1: Breakdown of the number of participants for each step of the study: enrolment, individual and final combined analyses.

This work was supported by Grant No. OPP1164153 from the Bill & Melinda Gates Foundation was awarded to Prof. John Spencer. The study was approved by the Community Empowerment Lab Institutional Ethics Committee (Ref. No: CEL/2018005) in compliance with ethical regulations and standards. All participants provided written informed consent; where caregivers were illiterate, a witness gave signed consent accompanied by a thumb impression of the caregiver in place of a signature. All caregivers had normal or corrected to normal vision. Caregivers were checked for colour-blindness during recruitment using the Colour Blind Check app (Check, 2016) on a Samsung Galaxy 4 tablet.

2.3.2 Materials

2.3.2.1 Demographic information

A set of demographic questions was used to gain a better understanding of the sample and its context. Examples of items addressing individual characteristics include age, education, occupation, and religion. Information about their spouses (age, education, and occupation) was also gathered. Other items targeted family characteristics, such as the caste, number of people the participants lived with, the participants' number of children under the age of five years, and the family income. To ensure a holistic understanding of SES, information was collected for the household's access to electricity, type of toilet used, cooking fuel used, and type and number of cattle (for personal and economic use). For the full demographic questionnaire, please see Appendix A.1, A.2 and A.3.

2.3.2.2 Maternal socioemotional measures

Edinburgh Postnatal Depression Scale (EPDS). The ten-item EPDS is a commonly used postnatal depression screening tool during the perinatal period (Cox, Holden, & Sagovsky, 1987). Due to feasibility constraints in the project, this assessment was carried out in Year 2. Given that the questionnaire asks participants to think about their thoughts and feelings in the last seven days, we used the EPDS as a way to assess current depressive thoughts or feelings instead of postnatal depression per se (see Appendix for EPDS questionnaire in English in Figure B.1, and Hindi in Figure B.2). The items were assessed on a Likert scale from 0-3 resulting in an overall score from 0-30. Higher scores represent greater severity. For example, "I have been able to laugh and see the funny side of things" is scored on a scale of 0 = "as much as I could", 1 = "Not quite so much now", 2 = "definitely not so much now" and 3 = "not at all". Some items are reversed scored. For example, "I have felt sad or miserable" is scored on a reverse scale of 0 = "yes, most of the time", 1 = "yes, quite often", 2 = "not very often" and 3 = "no, not at all". EPDS has been translated and validated in several Indian regional languages including Hindi and has been used in urban as well as rural India (Benjamin, Chandramohan, Annie, Prasad, & Jacob, 2005; Patel et al., 2002; Werrett & Clifford, 2006). For the sake of our study, we used the Hindi version of EPDS scale (Joshi, Lyngdoh, & Shidhaye, 2020).

IPV & Sense of Empowerment. The IPV and sense of empowerment questionnaire was constructed by the partner organisation, CEL, based on the indicators used by the Demographic and Health Survey (DHS; DHS Program, 2020). The questionnaire was translated into Hindi. It includes 23 items across four sub-topics: sense of empowerment (five items), verbal abuse (nine items), physical abuse (seven items), and sexual abuse (two items). All items were assessed using a Likert Scale with responses being coded as 0 for never, 1 for sometimes, and 2 for often. The sense of empowerment measure included the items in the empowerment subsection, such as "Your husband (respects/respected) you and your wishes?" and "Does your husband consult you when making household decisions?". The IPV measure included the 18 abuse items, such as "your husband has insulted or humiliated you in front of others" for verbal abuse, "does (or did) your husband slap you?" for physical abuse, and "does or did your husband physically force you to have sexual intercourse with him even when you did not want to?" for sexual abuse.

2.3.2.3 VWM-PL Task

Apparatus. A 24-inch BenQ Zowie XL2411P monitor screen was used to display the stimuli, which were presented using SR Research Experiment Builder. The monitor was connected to a Gigabyte mini-computer and a Lenovo laptop that interfaced with the eye-tracking software. An Eyelink Portable Duo (SR Research, Ontario, Canada) eye-tracker was used in remote mode to collect looking data. Eye-tracking data was collected using the binocular mode, at 500Hz. The screen, eye-tracker, Gigabyte mini-computer and laptop were all placed on a table. A large sofa was used for participants to sit, and a target sticker was used for the eye-tracker to track their head movement and eye position. All pieces of equipment other than the screen were portable including a foldable silicone keyboard, a mini Xmi Pte Ltd portable speaker, and a standard computer mouse. Figure 2.2 shows the setup for the experiment. The set-up also included a third-person camera, set up on a tripod stand in a way that it recorded the experiment as it was presented on the monitor as well as to keep a record of participants doing the task.

Stimuli. We used the Preferential looking task developed by Ross-sheehy et al. (2003) to assess visual working memory. This task has been used with adults previously (Simmering, 2016b) and was ideal in our setting given that no language is involved. The stimuli consisted of two side-by-side blinking displays, each including an array of coloured squares (set size = two, four, or six squares). The set size of the array was the same on each side of the display. Each array was 21cm (h) by 29.5cm (w) in projected size, with a separation of 21 cm between them. All squares measured 5cm (h) by 5cm (w).

The colours of the squares were randomly selected from a set of nine colours: black, blue, brown, cyan, green, red, violet, white and yellow. All squares simultaneously appeared for 500ms and disappeared for 250ms per trial. The

Figure 2.2: Portable setup for caregiver's VWM task: 1) participant 2) eye tracker 3) screen 4) laptop interfacing eye-tracking software.

array on one side displayed no change in the colour of the squares throughout the trial. The other side contained an array with the change display. That is, a random square changed colour after each trial. The colour of the changing square was selected from the set of colours not included in that display so that no two squares would have the same colour. Each trial lasted 10s, and the change side was randomly re-selected in between trials. The combination of set size (2, 4, 6 squares) and change side (left, right) resulted in 6 unique trial types. There were a total of 18 trials divided into blocks of six trials. Figure 2.3 shows a schematic trial of set size 2. Loads were classified as low, medium, and high for 2, 4 and 6 coloured squares, respectively.

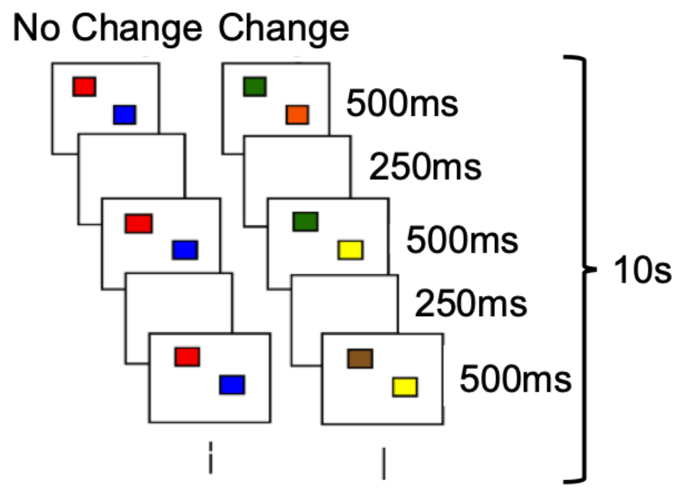


Figure 2.3: Schematic representation of set size 2 from a trial in the VWM-PL task. Trials for this study consisted of 2, 4 or 6 coloured changing (right panel) or non-changing (left panel) squares.

2.3.3 Procedure

As noted above, data for this Study mainly belongs to the second year of data collection. Demographic information was collected in Year 1 and, updated by CEL when relevant, so it is not included as part of our Procedure. At the start of each session, families toured the laboratory while all procedures were explained to them. Families were shown the equipment, it was explained what each piece of equipment does, and they were given the opportunity to ask any questions. They were, then, seated in a common playroom where consent was given.

Task order was randomly assigned: half of the participants completed the VWM task first, while the other half completed the socio-emotional measures first. Socio-emotional measures were completed through an interview. To ensure that the interviews adhered to the National Family Health Survey-3 (NFHS-3) guidelines, the following protocol was observed:

1. Only the caregiver (in this case mother of the infant) was allowed in the

room. Interviews took place in a quiet room by a trained female interviewer from the same state in India who understood the cultural nuances. Having a trained female interviewer was essential to participants feeling safe as well as respecting the cultural values of the community.

2. Participants were welcomed in the room, made comfortable and a rapport was built.

3. The objective of the session was explained to participants. That is, they were informed that they were going to be asked questions that were of a private nature, with questions exploring various aspects of the relationship of the couple that can influence the goal of the project.

4. Participants were informed of the process of maintaining confidentiality and anonymity of their personal information so that informed consent could be obtained before the interview. The interview was recorded only after obtaining individual consent and assuring privacy.

5. If, for any reason, privacy could not be ensured, the interviewer would skip the IPV questionnaire to ensure the safety of the participant.

The interviews took place in Hindi (language), Awadhi (regional dialect) or a mix of both. For the EPDS, participants were asked to consider the response closest to their feelings in the previous seven days. For IPV and empowerment questionnaires, participants were asked to consider the response closest to how they felt in the previous 12 months. The interviewer asked the questions verbally and filled the appropriate response on behalf of the participants. The interviews lasted between forty minutes to an hour, in total, depending on the need for engagement with each participant.

For the VWM task, the participant and experimenter were seated in the same room. Mothers were tested without infants present to prevent distraction. A member of the CEL team or a relative of the family (e.g., grandmother, sister, father) cared for the infant in a nearby playroom. If infants showed distress due to being separated from their mothers, the experiment was stopped so that they could be reunited. Instructions for the task were provided both in Hindi and Awadhi: "For the next few minutes, we will be showing you a set of videos with displays on the right and left sides of the monitor. We would like you to watch these displays carefully. We will ask you a few questions about the displays after the videos are complete. As you watch the videos, we will be monitoring where you look using this camera. To help us track where you are looking, we will have you wear this sticker on your forehead. Do you have any questions?" (see Appendix C.1 for the task instructions in Hindi). The target sticker was then placed on the participants' foreheads and the experimenter adjusted the camera so that the distance from the target to the camera was approximately 50cm. Adjustments were also made to ensure that the eyes of participants were in line with the top part of the screen.

The experimenter began the calibration after checking that the corneal reflection and pupil were visible in the camera. During calibration, participants were shown a geometric white and black target shape in five locations of the screen (middle, top, bottom, left, right). Calibration was used to map raw eye position data onto the camera image data, allowing mapping of gaze position to stimulus presentation. Following a successful calibration, the task began. The calibration took place between each block of trials for caregivers, that is, three times.

Following the VWM task, participants were asked a set of six questions such as "what was your impression of the task?", "did you notice any differences between the displays? please explain", and "any other comments?" (see Figure C.1). The questions were asked to understand whether the participants paid attention to the stimuli presented. Experimenters were trained to not distract, give feedback, or engage with participants during the task. The VWM session took less than eight minutes.

2.3.4 Measure Construction and Methods of analysis

2.3.4.1 Socioeconomic Status Scores

Selection of demographic information for the sake of calculating SES scores was carried out based on the updated Kuppuswamy Scale (Saleem, 2020). Kuppuswamy's scale has been used in both rural and urban contexts in India (Mohan & Bhat, 2022; Pandith, John, Bellon-Harn, & Manchaiah, 2021). The scale uses demographic information including family income, families' education, and occupation. Figure 2.4 shows spearman's correlation between parental education (used to categorise SES Educational status during recruitment), family income, and SES score (calculated based on Kuppuswamy scale), suggesting that the recruitment criteria fit the SES scores.

2.3.4.2 Maternal socioemotional context

EPDS. The total score for depressive thoughts was calculated by adding up the scores of each item in the scale. Given that the score of depressive thoughts was heavily skewed, we excluded participants who scored 0 and carried out a logarithmic transformation. We assumed zero scores to mean "no depressive symptoms". Out of 187 participants, 55 (29.41%) scored 0 on EPDS and were excluded from the analyses (Figure 2.1 in Participants subsection; also see Appendix D.1 for a histogram of non-transformed and transformed depression scores).

IPV and Sense of empowerment. Verbal, physical, and sexual abuse scores were aggregated into a single Abuse Score. Given that the aggregated Abuse

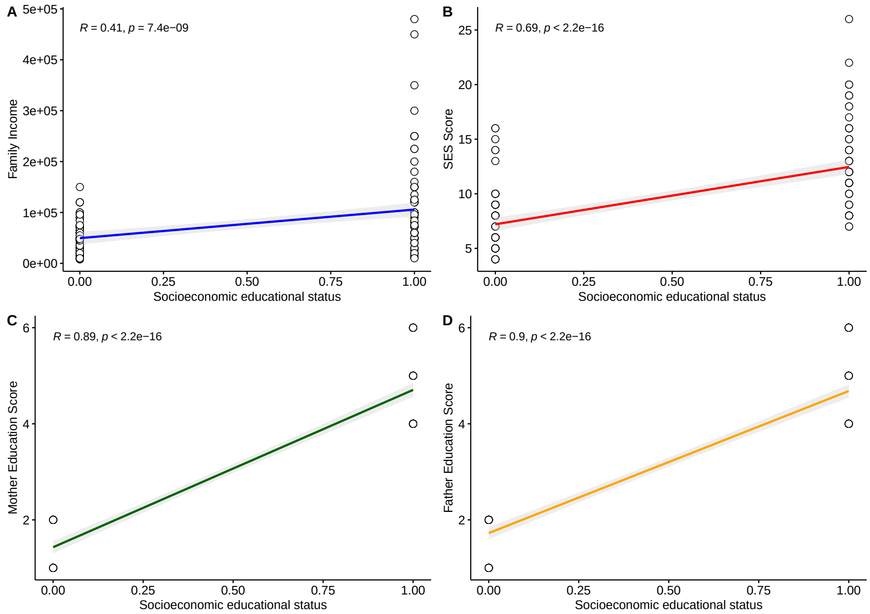


Figure 2.4: Scatter plot for SES educational status and family income, parental education and SES score. Figure A represents the correlation between SES educational status and Family Income. Figure B represents the correlation between SES educational status and SES Scores. Figure C represents the correlation between SES educational status and Mothers' education scores. Figure D represents the correlation between SES educational status and Fathers' Education Score.

Scores were heavily skewed, we excluded participants who scored 0, and we carried out a logarithmic transformation. Out of 138 participants, 33 (23.91%) scored 0 and were excluded from the analyses (Figure 2.1; also see Appendix E.1 for a histogram of non-transformed and transformed abuse scores). Sense of empowerment subsection of the questionnaire items were aggregated into a single composite Empowerment Score.

2.3.4.3 VWM-PL Task Caregiver

The eye-tracking data was pre-processed using Data Viewer (SR-Research, Ontario, Canada). Fixations with a duration under 100ms were merged with the adjacent fixation as long as the latter was within 1 degree. If neighbouring fixations did not meet these criteria or were not temporally adjacent, the short fixation (<100ms) was discarded. Trials were divided into periods of interest (IP) using message-based events. To account for calibration errors and drifts in the eye tracker, the area of interest (AOI) was set to be 50% of the screen size, such that the looking behaviour resulted in looking to the left, right, or away from the screen. Sample reports were exported and the raw gaze position was processed using the statistical package R (R. C. Team, 2017) and eye-tracking R, a statistical package designed for the analysis of eye-tracking data (Dink & Ferguson, 2016).

Looking to the target (change side) and distractor (non-change side) at each point in time during the trial was aggregated into 100ms time bins that allowed calculating the proportion of looks to the target (change side). This was done using a growth curve model (GCA) to measure how the probability of looking to the target (change side) changed over time (Mirman, 2014). To allow for the best possible modelling of the time series data, we trimmed the data to a five-second window. In each trial of the task, the first opportunity for the participants to distinguish between the change side versus the no change side is at 750ms (each display is on for 500ms and off for 250ms). Thus, we focused on the time window from 1500ms to 6500ms. This gave participants one 'on' period after 750ms to start to notice the change. The last 3500ms of each trial were removed because of their tendency to be noisy, particularly for the high load conditions.

2.3.4.4 Analytic Strategy

The analytic strategy for Study 1 was developed to place maternal VWM within the socio-emotional context and characteristics of participants. Although our sample size was sufficiently large relative to other research in developmental psychology, we deemed it to be insufficient for running complex models with multiple variables. Moreover, many participants did not complete all measures and all tasks. That means that the greater the number of variables included in any analysis, the lower the number of participants who completed all of them, the smaller the sample size, and the lower the statistical power. Thus, we took a step-by-step approach to complete our analyses. Our goal was divided into three key stages: 1) understanding the socio-emotional characteristics of mothers in our sample; 2) understanding their VWM performance over time and developing a streamlined model that allowed us to focus on key areas of interest; 3) placing VWM performance in relation to the mothers' socio-emotional context. At each stage, the maximum number of participants possible was included in the analyses.

In the first stage, descriptive statistics and bivariate correlations were calculated for the three socio-emotional measures (EPDS, IPV, and sense of empowerment). Next, we examined the relationships between each of the socio-emotional measures and contextual variables. Three multiple regression models were carried out, one per socio-emotional measure, with the participants' age, SES Scores, and their infants' gender as predictors. Infants' gender was added as a predictor because previous research has indicated the impact of a spouse's insistence on a male child, and/or the birth of a female child, are associated with mothers experiencing domestic violence (Nongrum et al., 2014) and depression (Savarimuthu et al., 2010; Patel et al., 2002) in India. Two-way interactions between the three predictors, and the three-way interaction, were also explored.

In Stage 2, two steps served for an initial exploration of VWM and the development of a 'base' model. First, a timecourse analysis was carried out with Load (low, medium, and high), scaled SES Scores, and first look side as predictors of the probability of looking at the target (change side). Interactions between the predictors, and between predictors and time terms were modelled to evaluate the combined relationships between predictors and the outcome over time. The time-course analysis served to simplify later models. In the next step, we carried out analyses on the aggregated scores for the proportion of time looking at the change side. Having simplified the model, the mother's age was added as a predictor. Analyses served to further polish the VWM model, thus informing the next analyses.

The final stages aimed at understanding the relationships between maternal socio-emotional factors and maternal VWM performance, with the streamlined model for VWM from the previous step serving as a base. Three separate models were constructed, each one adding one of the three socio-emotional variables (EPDS, IPV, and sense of empowerment) to the maternal VWM 'base' model. A total of 120 participants completed all socio-emotional measures and the VWM task (Figure 2.1 in the Participants subsection).

2.4 Results

2.4.1 Maternal socioemotional measures

To understand our sample population, we first looked at the mean scores and bivariate correlations among the three maternal socioemotional measures – the EPDS, IPV and the sense of empowerment scales. The measures were included without transformation in order to avoid losing participants as a result of removing the scores of 0 (i.e., the sample size would be reduced when removing 0s from EPDS, and then reduced again when removing 0s from IPV). However, given that measures had a skewed distribution, all correlations were conducted using Spearman's *rho*. Table 2.2 reports the means, standard deviations and correlations between the measures for the 138 mothers who completed all three questionnaires.

All correlations between socio-emotional variables were significant, with depressive thoughts being positively associated with IPV and negatively with empowerment. The strongest correlation was found between IPV and empowerment, with mothers feeling less empowered the greater the IPV that they experienced. Findings serve to corroborate the expected overlap between measures, with lack of empowerment and control being closely linked to the construct of IPV conceptually and both measures being connected to depressive thoughts. Correlations also supported the inclusion of empowerment as a variable separate from IPV, being distinct despite the conceptual and empirical overlap. That is, medium to strong correlations reveal a clear overlap, but the amount of shared variance (maximum of 28% between IPV and empowerment) indicates the distinctiveness of the variables.

Next, the three multiple regression models on the socio-emotional variables were run (see Appendix F.1 for Scatterplots between each outcome variable and its predictors), using the lm function of the R package (R. C. Team, 2017). Fixed effects were tested with a Wald Chi-squared test to assess the contribution of each parameter in reducing residual deviance of the models. For each model, the effect of parameters was further assessed with an F test using the ANOVA function from the car ANOVA package (Fox & Weisberg, 2019), which tests whether the model terms are significant. Residuals were checked for normality using Q-Q plots and the DHARMa R package (Hartig, 2021).

EPDS The model on EPDS scores was not significant overall (F (7, 124) = 1.73, p = .11, R^2 = .04). No significant main effects or interaction were found (see Table 2.3).

IPV The model on the logged abuse scores also was not significant (F(7, 97) = 1.382, p = .22, R^2 = .02). Nevertheless, results indicated a main effect of Mothers' age on the abuse scores (see table 2.4). As can be seen in Figure 2.5, the older that the mothers were in our sample, the lower the abuse scores. No other main effect and no interactions were found to be significant.

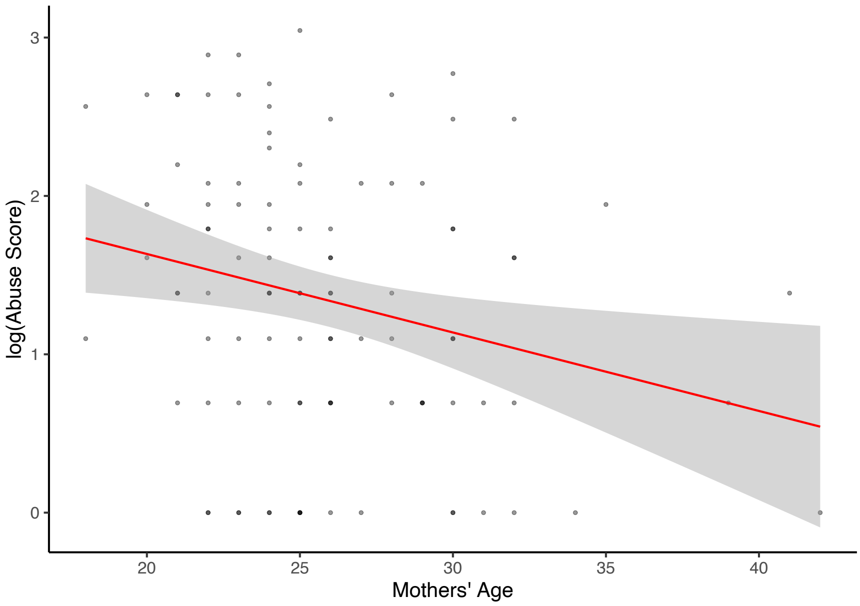


Figure 2.5: Relationship between IPV scores and mothers' age

Sense of Empowerment The final model in Stage 1, on sense of empowerment, revealed the main effects of Mothers' age and of infants' gender (see Table 2.5). As shown in figure 2.6, the older that the mothers were, the greater their sense of empowerment. In addition, mothers of female infants (M = 6.87, SD = 2.50) scored higher for sense of empowerment than mothers of male infants (M = 5.71, SD = 2.60). No significant interactions were found. It must be noted that the overall model was not significant once again (F(7, 130) = 1.87, p = .08, R^2 = .04). Therefore, findings should be interpreted with caution.

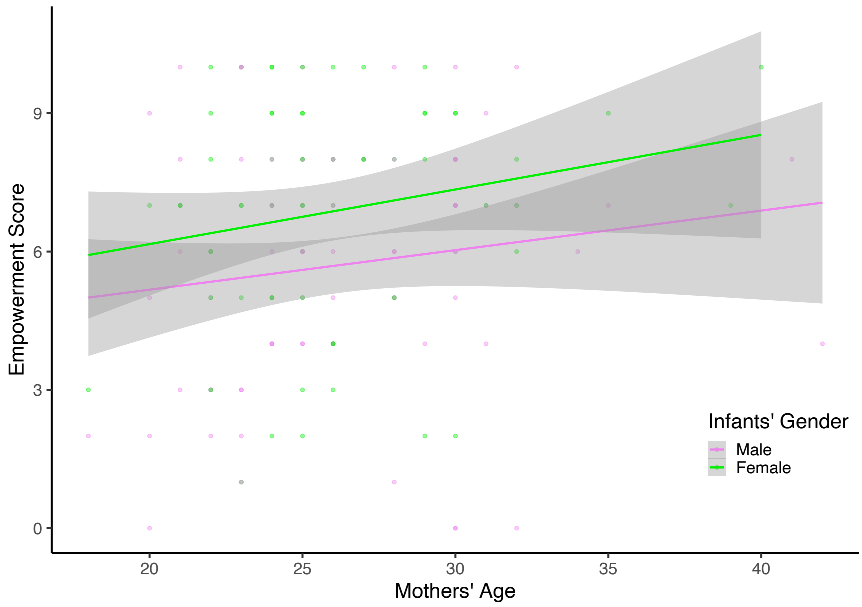


Figure 2.6: Relationship between Sense of empowerment scores, mothers' age and infants' gender.

2.4.2 Development of VWM Base Model

The probability of looking to the target (change side) over time for each participant was fit with a binomial, logistic mixed-effects regression model estimated with Laplace approximation using glmmTMB package version 1.0.2.90 (M. E. Brooks et al., 2017) in the R programming language. The number of orthogonal time turns was chosen based on the shape of the curve. We modelled time using quintic orthogonal polynomials (Mirman, 2014), that is, time, time squared up to time to the power of five but scaled and centred so as not be correlated with one another. Each time term was nested as a random effect within the interaction between participants and load. Our model contained fixed effects of Load (low, medium, and high), scaled SES Scores, first look side (change vs no change side) and the time terms. The first look side was determined by the first data frame available within the time window from 750-4500ms; for the majority of trials, the first look was determined during the second display prior to the onset of the third display (i.e., between 750-

1500ms; see Spencer et al., 2023). Previous studies have computed the VWM scores by dividing the total looking time to change side by the total looking time to both change and no change sides to obtain the measure of change preference score (VWM Score; e.g., Delgado Reyes, Wijeakumar, Magnotta, Forbes, & Spencer, 2020). The model was assessed using half-normal plots. As in Stage 1, fixed effects were tested using Wald chi-squared.

As seen in table 2.6, results revealed the main effects of load and first look, with performance decreasing the greater the load and when participants first looked at the no-change side. All two-way interactions between each time term and load were significant. Additionally, the two-way interactions between time terms and first look were significant with the exception of the quadratic time term. Time terms, except for the linear and cubic time terms, also interacted with SES scores. Results further revealed significant two-way interactions between load and first look, and between SES scores and first look. Last, all three-way interactions between first look, SES scores and the time terms were significant except for the linear time term. Together, the results provide evidence that the time course of looking to the change side varies by load, first look, and SES Scores.

The model fit to the raw data can be seen in Figure 2.7. We see a clear effect of load when the first look is to the no-change side. That is, when there were fewer items to consolidate in working memory and participants looked to the no-change side first, they released the fixation more quickly in order to look to the change side. By contrast, when the first look was to the changing side, participants generally continued looking to that side, regardless of the memory load. The interaction between load and the first look was further qualified by a three-way interaction with SES scores. In particular, when higher SES participants first looked to the no-change side in the high load condition, they released fixation more robustly to look to the changing side relative to lower SES participants. Taken together, results show little variability when participants start by looking at the change side. In contrast, when looking at the no change side first, the relationships between load and SES, and the proportion of looking time become apparent. Based on these findings, we focused subsequent analyses throughout this thesis on the proportion of looking to the change side when the first look was to the no-change side.

A key point to consider here relates to chance. When studying a behaviour or response with two possible outcomes, a proportion of .50 is considered to display chance (proportion of 1 divided by number of options). It logically follows that the further away that scores are from a proportion of.50 (whether above or below), the greater the certainty that responses are not due to chance. In our analysis, however, chance largely happens on the first look (i.e., the chance that the participant will happen to be looking at the change or no change side at first). When separating scores based on first look, this chance is already accounted for, leading to distinct patterns of responses. Inevitably, the first look biases the proportion of looking to the change side, making a score of .50 not representative of chance. Indeed, as shown in Figure 2.7, when participants happen to be looking at the no change side first, scores are lower and tend to remain below .50. Thus, the increase in proportion of looking time (approaching .50) cannot be interpreted as ”approaching chance levels” and is better understood in terms of the time taken for participants to release their fixation rather (as discussed above; see also Spencer et al., 2023).

For the following analysis, we aggregated the proportion of looking time across the 1500-6500ms time window. A correlation matrix with a scatter plot and histogram for the proportion of looking to change side for each load along with its predictor variable (i.e. EPDS, Empowerment Scores, IPV, SES Scores

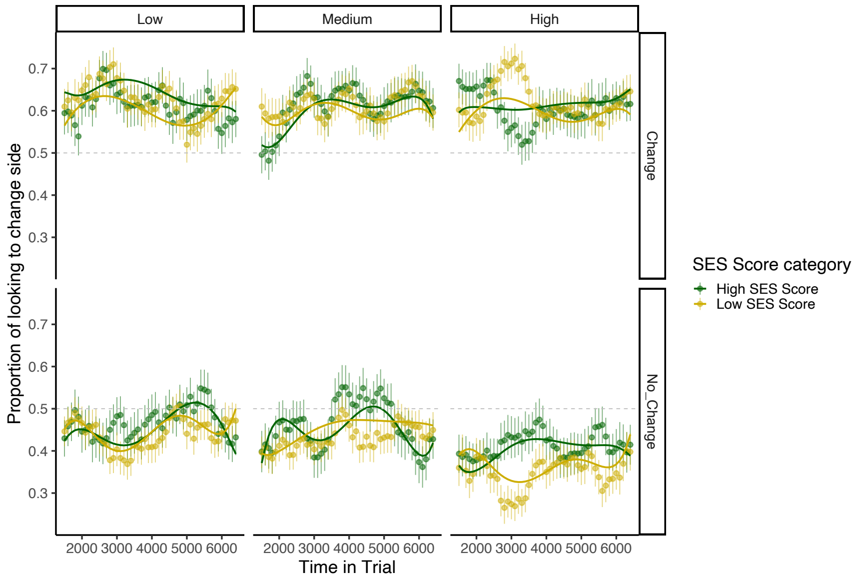


Figure 2.7: Model predicted proportion looking to change side by the load by First Look and SES Score. The grey dotted line depicts chance performance (0.50). SES Scores were median split for the purpose of visualisation.

and Mothers' Age) can be found in Figure G.1 for low load, Figure G.2 for medium load and Figure G.3 high load conditions.

We statistically modelled these data using a linear mixed-effects model using the lmer function from the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R (R. C. Team, 2017) to predict the proportion of looking to change side when the first look was no change side. We included a random effect for participants and fixed effects of Load (low, medium, high), centred SES Scores, and Mother's Age. Preliminary analyses compared a model with main effects only to models with candidate two and three-way interactions. We first ran a three-way interaction model, which included the main effect of Load, SES Scores, and Mothers' Age, as well as two-way interactions between load and SES, Load and mothers' age, and SES Scores and mothers' age, and the three-way interaction (see table in Appendix H.1 for results). We then ran a model including only the main effects of the three variables (see table in Appendix H.2 for results). Although the model with no interactions was the most streamlined (an important consideration given the planned inclusion of socio-emotional variables in the later models), it missed potentially important interactions between task difficulty and maternal age (see table in Appendix H.1). Thus, we ran a final model that included the main effects of Load, SES Score, and mother's age and a two-way interaction between Load and mothers' age. Model fit was assessed using ANOVA and comparing Akaike's Information Criterion (AIC; Wagenmakers & Farrell, 2004). The best model fit was achieved by the final model (see table in H.3).

Results of the final model revealed a main effect of load and interaction between load and mothers' age (see table 2.7 for detail on Wald chi-squared test). The main effect of load is visualised in Figure 2.8. As was evident in the timecourse model, the proportion of looking to the changing side was higher in the low and medium loads relative to the high load condition. That is, participants released fixation from the non-changing side more robustly when there were fewer items to consolidate in working memory.

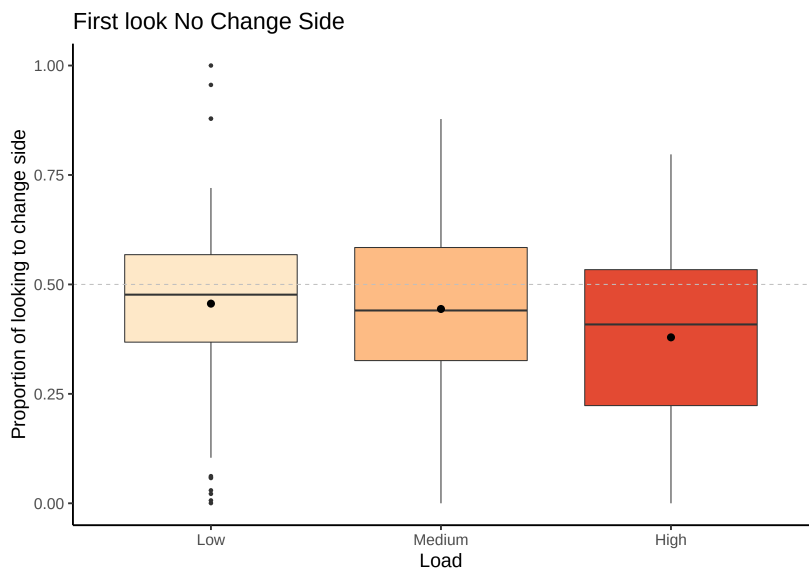


Figure 2.8: Aggregated proportion of looking to change when the first look is no change side. The grey dotted line depicts chance performance (0.50).

Figure 2.9 shows the interaction between load and mother's age. The older that the mothers were, the higher the change preference scores in the low load condition. The association was reversed in the medium load condition, with increasing changes in preference scores being found the younger that the mothers were.

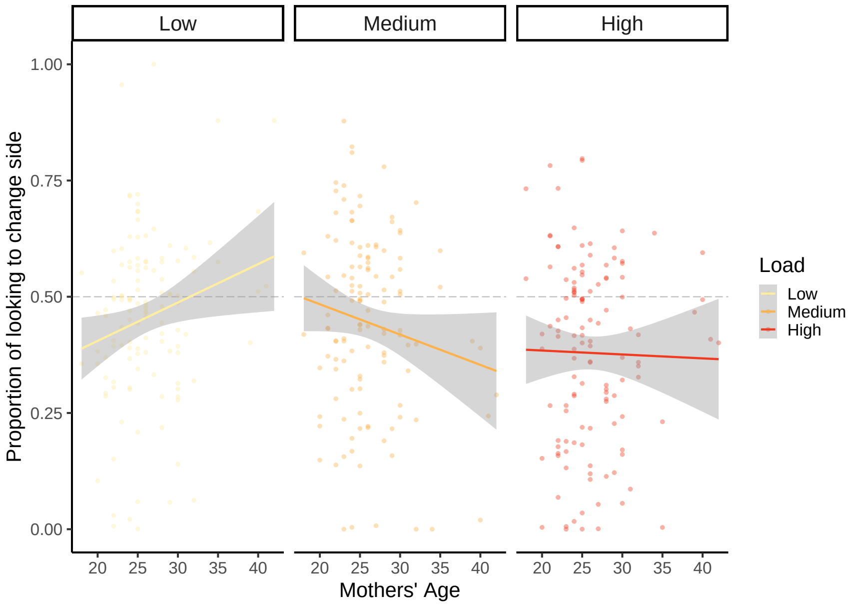


Figure 2.9: Mothers VWM base model predicted proportion looking to change side by load and mothers' age. The grey dotted line depicts chance performance (0.50).

2.4.3 VWM in relation to Socio-emotional factors

Maternal EPDS scores and VWM To understand how maternal depressive thoughts predict VWM performance, we added the main effect of log-transformed EPDS scores to the base VWM model. Two-way interactions between EPDS and the predictors in the base model (load, SES, and mothers' age) were also included. Additionally, we modelled the three-way interaction between load, mothers' age, and depression to see if depression moderated the earlier significant findings. Consistent with our VWM base model, results revealed the main effect of load on participants' VWM performance. However, the interaction between mothers' age and the load was not significant in the current model. Additionally, depressive thoughts did not significantly predict VWM performance on their own or through interactions (see table 2.8 for statistics).

Maternal Abuse Scores (IPV) and VWM In addition to the base model, we included a main effect of log Abuse scores. The two-way interactions between IPV and all predictors in the base VWM model were also included as predictors. The three-way interaction between load, mother's age, and IPV was also modelled. The main effect of load from the base model ceased to be significant in the current model, but the interaction between load and the mother's age was significant. Regarding the IPV scores, there was no main effect of the log abuse score on the proportion of looking to the changing side. There was, however, a significant 2-way interaction between log abuse scores and SES Scores was revealed. As seen in figure 2.10 A, there was a negative relationship between the proportion of looking to the changing side and log abuse scores only among mothers with higher SES Scores. That is, for participants of higher SES, the higher the abuse scores, the poorer the VWM performance.

There was also a significant 3-way interaction between load, mothers' age and abuse scores (see Figure 2.10 B). In low load conditions, both older and younger mothers' showed poorer VWM performance the higher the abuse scores. At the medium load, the same relationship was evident for younger mothers; however, older mothers tended to perform better in the VWM task the higher the Abuse scores. In the high load condition, there was no clear relationship between abuse and VWM performance (see table 2.9 for Wald chi-square statistics).

Maternal Sense of Empowerment and VWM The third linear mixed-effects model building on the VWM base model includes the empowerment score

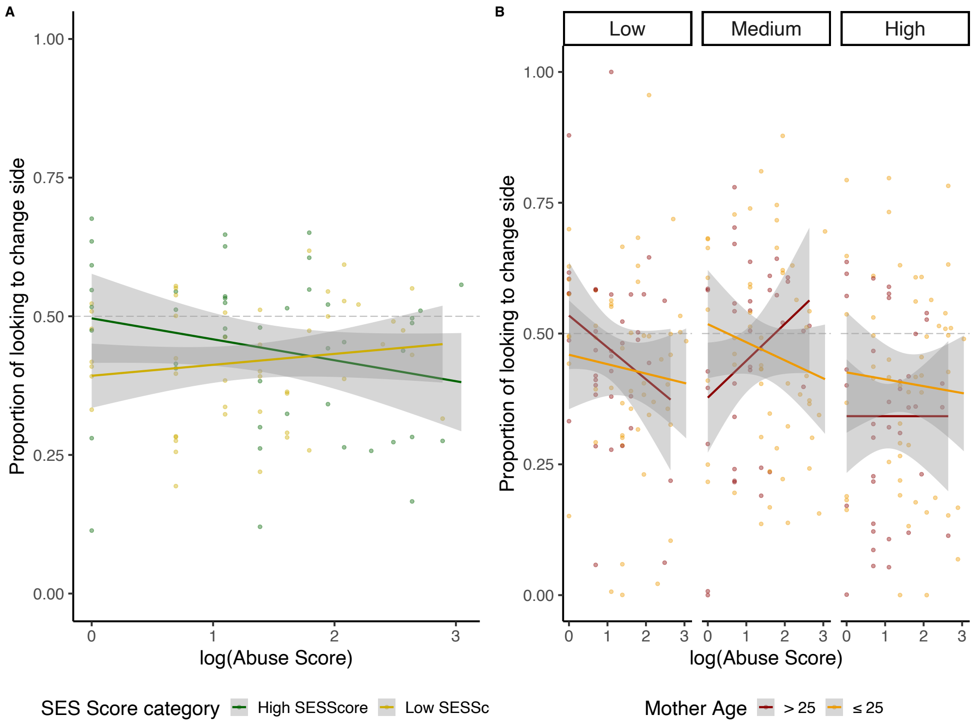


Figure 2.10: A) Model predicting proportion looking to change side with log(Abuse scores) and SES Score as dependent variables. B) Model predicting proportion looking to change side with log(Abuse scores), mothers' age and task load set size 2,4,6) as dependent variables. SES scores and mothers' age were median split for visualisation purposes. The grey dotted line depicts chance performance (0.50)

as well as infants' gender as predictors. We added gender in this analysis to be consistent with the socio-emotional analyses above, where infants' gender predicted maternal sense of empowerment. As in previous models, two-way interactions between empowerment and the three predictors of the base model, as well as the three-way interaction with the mother's age and load were included in the analysis. The same 2-way interactions with the base model were tested in relation to gender. An interaction between gender and empowerment was added as a predictor as well. To elaborate on the potential 2-way interaction between empowerment and gender, 3-way interactions between empowerment, gender and load, SES score, and mother's age were modelled. Last, the four-way interaction between load, mother's age, empowerment, and gender was also included among predictors (see Table 2.10 for Wald chi-square statistics).

Results revealed no main effects on the proportion of looking to the changing side. The interaction between the mother's age and load from the base model was also not significant. The only significant 2-way interaction was found between SES Scores and infants' gender. Among the 3-way interactions, significant results were found for the interaction between SES Scores, infants' gender, and sense of empowerment scores (see figure 2.11). The visual representation of the data suggests that for lower SES mothers of male infants, there was a positive relationship between sense of empowerment and the proportion of looking to the changing side. That is, the lower SES mothers of male infants performed better on the VWM task the higher their sense of empowerment. The relationship was reversed among mothers of female participants, which showed that mothers of lower SES performed worse the higher their sense of empowerment. The opposite was found for higher SES mothers. Among those who had female infants, the higher sense of empowerment, the better the performance on the VWM task. For high SES mothers of male infants, the higher the empowerment, the poorer the performance.

To further evaluate the 3-way interaction, we split the data by gender and ran follow-up analyses (with the same predictors) on VWM for each gender separately. The analysis on mothers with male infants showed no significant main effects of the maternal sense of empowerment on her VWM performance and no interactions. That is, among mothers of boys, the relationship between empowerment and VWM was not different for mothers of higher versus lower SES (c.f., apparent interaction in Figure 2.11). However, analyses on mothers with female infants indicated a significant main effect of SES Score on VWM performance, with VWM improving the higher the SES. A significant 2-way

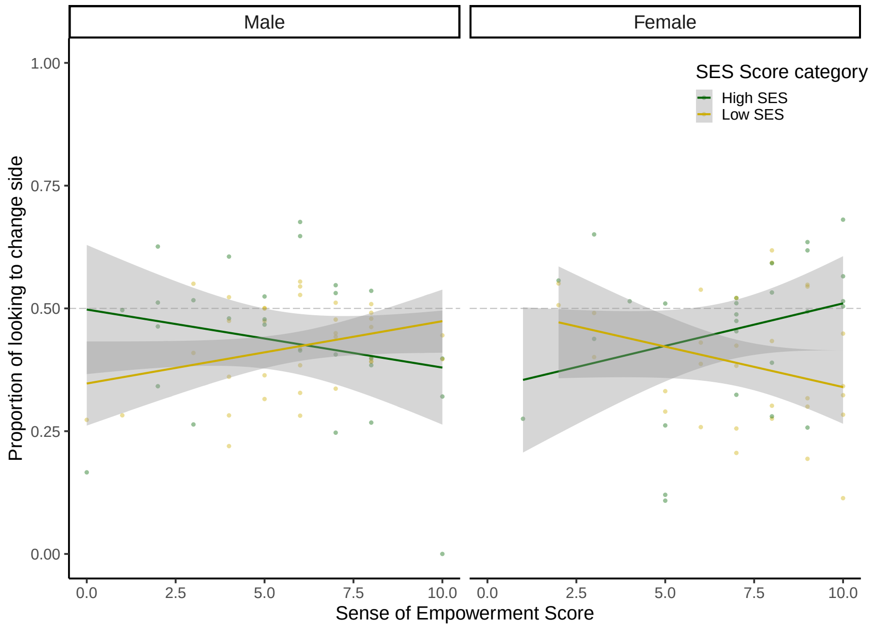


Figure 2.11: Model predicting proportion looking to change side with sense of empowerment, infants' gender and SES Score as dependent variables. SES Scores were median split for visualisation purposes. The grey dotted line depicts chance performance (0.50)

interaction between SES Score and maternal sense of empowerment on VWM performance was also found. Maternal VWM performance increased together SES scores but only for mothers with a higher sense of empowerment.

2.5 Discussion

Results of Study 1 serve to contextualise our sample and their VWM performance. Initial analyses served to corroborate the expected overlap between socio-emotional variables of interest (depressive thoughts, IPV, and sense of empowerment) as well as their distinctiveness. Interestingly, depressive thoughts among our participants were found to be independent of demographic variables of interest, not being predicted by maternal age, SES score, the gender of the infant, or their interactions. However, both IPV and sense of empowerment were predicted by maternal age, with mothers experiencing less IPV and feeling more empowerment the older they were. Mothers of female infants also reported a greater sense of empowerment.

Previous research has indicated that the gender of children in one's family is a relevant predictor of depression and IPV among Indian mothers due to the social desire for a male child (see Nongrum et al., 2014; Patel et al., 2002; Savarimuthu et al., 2010). However, in our sample, not only did the gender of the infant fail to predict these outcomes, but it also predicted empowerment in the opposite direction to what might be expected based on the literature, with mothers of female infants reporting a greater sense of empowerment. It is important to treat this result with caution as mothers in the present study often had multiple children and our analyses only focused on the gender of the infant enrolled in the study. Thus, mothers of female participants in our study could also have male children of other ages. In regards to sense of empowerment, social preference for a male child may result in greater family focus and involvement on child-rearing decisions when the infant is male. Greater paternal pressure and control (as well as enhanced interference from the in-laws) linked to the upbringing of a male child compared to a female child could be related to our findings on sense of empowerment. Family pressure and control may be particularly low if the infant is female and a male has already been born. Researching potential interactions between the gender of the infant and the gender of their older siblings will help clarify some of the complexities and nuances related to maternal feelings of empowerment. Moreover, accounting for the influence of mothers' relationship with their natal kin may further serve to enhance a holistic and contextualised understanding of empowerment (Bloom, Wypij, & Das Gupta, 2001).

Findings from the VWM base model served to replicate previous research on VWM with adults (Simmering, 2016a). The time course analysis revealed that our sample showed decreased VWM performance as the task became more demanding as the load increased. Critically, the load interacted with the side that participants' were looking at first, so that the effects of task difficulty were clearer when participants first focused on the no-changing side. Moreover, performance also changed across SES scores, with higher SES participants releasing fixation more quickly when starting on the no-changing side, suggesting that high SES mothers consolidated the items in VWM more quickly at this load.

Findings on the aggregated VWM scores further revealed an age-related decline in VWM performance for the medium condition with a weaker declining trend in the high load condition. This fits with results from previous studies indicating that VWM performance peaks around the age of 20 (our youngest participants were 18 years old) and then declines over the lifespan (Brockmole & Logie, 2013; Costello & Buss, 2018). The finding showing a reverse trend for the low-load condition, however, creates some tension with the literature. A potential explanation for the results is that, in the easiest condition, mothers with a stronger VWM (the youngest) may have disengaged. The medium load, however, may have provided the "Goldilocks" spot to encourage engagement and display the developmental trajectory across age (with a noticeable decline over time). Trends in the high load condition still suggested a negative relationship with performance decreasing over time. Nevertheless, the greater difficulty of the task may have brought all performances closer together. That is, the task was sufficiently difficult for all mothers to perform relatively poorly. Indeed, the average performance was lowest in the high load condition independently of age and other variables.

With respect to understanding the impact of socio-emotional context on mothers' VWM we found no association between maternal depressive thoughts and her VWM in our sample. Results revealed both negative and positive impacts of IPV, with the latter being unexpected. Overall, experiencing domestic violence was associated with decreased working memory performance in the higher SES group. IPV also interacted with mothers' age and load. In the low-load condition, higher scores in IPV were associated with lower VWM across all ages. In contrast, in the medium-load condition, greater IPV was associated with higher working memory performance for the older mothers in our sample and with lower performance for the younger mothers. A possible explanation for this might be that older mothers who have, perhaps, experienced abuse for a long time may be hyper-vigilant and hypersensitive to detecting changes in the environment. Indeed, hypervigilance has been argued to be linked to the search of threats in the environment through excessively scanning the environment, broadening attention, and focusing on threat-related stimuli (Eysenck, 1997; Richards, Benson, Donnelly, & Hadwin, 2014). Given the overlap between hypervigilance, anxiety, and the deployment of attention in search for threats (Richards et al., 2014), future research on IPV and cognition should consider the influence of anxiety as a measure of relevance.

In regards to sense of empowerment, findings indicated that among mothers with female infants, VWM performance was higher as SES scores improved, but only for mothers with higher sense of empowerment. Results suggest that, on their own, improvements in the socio-economic status of a given population (e.g., increased access to education) may not be sufficient to achieve specific outcomes such as the improvement of VWM. However, further research is necessary on order to enhance the understanding of the influence of empowerment on cognitive outcomes. For instance, research can be aimed at teasing out the unique contribution of different aspects of empowerment (e.g., decision making, accessing resources, freedom of movement) on working memory (also see Bliznashka, Udo, Sudfeld, Fawzi, & Yousafzai, 2021).

Taken together, the results evidence the complex interplay of numerous variables on maternal visual cognition, from socio-economic measures to experiences of IPV and sense of empowerment. It must be noted, however, that our research studied the socio-emotional measures of depressive thoughts, IPV, and sense of empowerment independently, without controlling for the shared variance between them. Our analytical approach is inseparable from considerations of statistical power, given that including all variables in the same model would have resulted in the loss of data from participants who did not complete all measures plus further losses due to the transformation of scores. Future research, with larger samples, may address the unique contribution of each variable. Considering mediational chains, such as increases in maternal age leading to higher empowerment and lower IPV, and the latter variables predicting VWM, will also serve to further polish our understanding of the complex array of influences on VWM.

Despite limitations, Study 1 invites a range of new research, provides insights into analytical strategies when researching VWM with preferential-looking tasks, and succeeds in highlighting the importance of individual differences as well as socio-economic characteristics and conditions for the study of VWM. Moreover, through addressing maternal VWM and its predictors, it successfully sets the stage for Study 2, which focuses on infant VWM. Mothers play a crucial role in infants' early development, and their socio-emotional context and cognitive abilities are likely to impact what mothers bring to the table during parent-infant interaction. Study 2, thus seeks to extend Study 1 to understand *'whether'* and *''how'* maternal socio-emotional context and cognitive abilities influence their infants' visual cognitive skills.

2.6 Study 2: Are mothers' socio-emotional context & visual cognition related to infants' visual cognition?

Study 2 builds on Study 1 by relating the maternal socio-emotional context and VWM performance to their infants' VWM performance. As argued in Chapter 1, infant cognition does not develop in isolation but through countless forms of interaction with others (primarily their caregivers). Thus, the context in which a family lives, the resources available, and the characteristics of infants' social partners are all likely to play a part in the development of infants.

VWM, marked by highly limited capacity, develops early in life, can be assessed as early as 4 months of age with a remarkably noticeable improvement from 6 months of age (Oakes, Baumgartner, Barrett, Messenger, & Luck, 2013; Ross-sheehy et al., 2003), and continues to develop through infancy. Although VWM capacity peaks around the age of 20 (Brockmole & Logie, 2013), it already reaches an adult-like level during childhood (Riggs, McTaggart, Simpson, & Freeman, 2006) displaying a rapid increase throughout infancy. A reliable method of measuring VWM of infants is the preferential-looking task developed by (Ross-sheehy et al., 2003), which has the advantage of not requiring verbal communication. Recent research by Wijeakumar et al. (2019) has successfully used the VWM preferential-looking task in rural India. Behavioural results indicated that 6- and 9-month infants successfully detected change in the VWM preferential looking task at low and medium conditions. The authors suggest medium load condition as the key behavioural marker for VWM performance and note that infants from Higher SES backgrounds performed better at medium load conditions than infants from lower SES backgrounds. A key point of consideration is that low, medium, or high load conditions adhere to considerations of task difficulty and, thus, are relative to age (Ross-sheehy et al., 2003). That is, the same number of items held in the working memory could be classified as difficult (i.e., "high-load") for younger infants but would become of medium difficulty (i.e., "middle-load") for older infants. Taken together, variables of infants' age, load and SES background are central to understanding infants' VWM.

In regards to maternal influences on the development of VWM among infants, studies from high-income countries have reported a negative impact of maternal depression on infants' later cognitive and socio-emotional development (Murray & Cooper, 1996; Feldman et al., 2009). Maternal depressive symptoms can include the loss of interest in daily tasks, fatigue, negative affect, sleep disturbances, concentration problems, agitation, and feelings of worthlessness (Clay & Seehusen, 2004). Such depressive symptoms measured in both clinical (Brookman et al., 2020) and non-clinical populations can have a negative impact on the quality of mother-infant interaction (Skotheim et al., 2013; Vieites & Reeb-Sutherland, 2017). For example, loss of interest or irritability can make mothers pay less attention to infants or be less tolerating of their behaviour (Lovejoy, Graczyk, O'Hare, & Neuman, 2000). Mothers with depressive symptoms tend to respond less contingently to their infants and with lower sensitivity, tending to be withdrawn during social interaction compared to non-depressed mothers (Cohn, Matias, Tronick, Connell, & Lyons-Ruth, 1986).

While early childhood adversities such as child abuse and neglect, and their long-term effects (including social and cognitive developmental outcomes), have been well-documented (Kessler et al., 2010; Reuben et al., 2016), research on the effects of caregivers' experience of Intimate Partner Violence

(IPV) on infants' cognitive outcomes is scarce. To our knowledge, there is only one study, from the U.S., on the link between caregiver's experience of IPV and 24, 30 and 36-month-old toddlers' executive function (EF) at 60 months of age (Gustafsson, Coffman, & Cox, 2015). Results showed that the association between maternal experience of IPV and infants' EF was mediated by maternal sensitivity. That is, higher IPV was related to lower sensitivity and through it, to lower EF. Findings are not surprising given that maternal responsiveness and sensitivity are important factors that influence infant cognitive and social development (Landry et al., 2008).

Just like IPV can be expected to negatively affect victims and those around them, women's sense of empowerment can lead to a better quality of life for both women and their family members. Positive outcomes, of course, can occur through multiple pathways. Women's autonomy, including decision-making in household activities, employment, education, access and control over resources, and freedom of movement (e.g. going to the market) have been shown to be associated with a variety of outcomes for their infants. For instance, Sethuraman et al. (2006) studied the relationship between women's empowerment and their 6- to 24-month infants' nutritional status in rural and tribal communities in South India. They found that women who were empowered to make decisions in the family tended to have infants with better nutritional status. Similarly, women with greater autonomy and freedom of movement (e.g. going to the market, visiting health centre) tended to have children with significantly less stunted growth (Shroff, Griffiths, Adair, Suchindran, & Bentley, 2009) and have been shown to be more likely to seek antenatal care (Bloom et al., 2001).

While there is a large body of literature from low and middle-income countries (LMIC) examining the relationship between maternal empowerment and infants' nutritional status and growth (Cunningham et al., 2015; Sethuraman et al., 2006; Shroff et al., 2009), to our knowledge there is only one study, from a sub-Saharan context, that has looked at the association between maternal empowerment and infants' cognitive development (Bliznashka et al., 2021). Although the study only found a weak association between maternal empowerment and their infant's cognitive development, more empowered women were found to be less likely to have infants with sub-optimal cognitive development.

Caregiving can be physically and emotionally demanding, requiring dynamic coordination of the brain and behaviour to reach multiple goals. For caregivers in a naturalistic environment, interacting effectively with infants requires several cognitive processes such as motivation, attention, and cognitive flexibility (Barrett & Fleming, 2011). For instance, caregivers need to respond contingently in order to attend to their infants' needs. These contingent responses require maintaining and manipulating information in the environment while attending to the infant and thus are connected to VWM. Moreover, contingent responses to infants have been found to result in improved developmental outcomes, including language acquisition (K. E. Smith, Landry, & Swank, 2006; Tamis-LeMonda, Bornstein, & Baumwell, 2001), which are partially underpinned by the infant's developing VWM. Exploring the connection between the VWM of mothers and that of their infants, thus, can serve to connect the dots provided by the literature and develop a clearer picture of the development of infants.

Based on the literature available, we hypothesised that infant age will predict VWM positively and load will predict it negatively. Moreover, SES scores are expected to predict VWM-PL scores positively (Wijeakumar et al., 2019). In regards to maternal socio-emotional characteristics, the literature invites hypothesising that caregiver depressive symptoms and experiences of IPV will negatively predict their infant's VWM-PL scores (see Feldman et al., 2009; Gustafsson et al., 2015)). Links to empowerment, however, are less clear in the literature reviewed, so the variable was included in an exploratory manner. Although logical speculation suggests that empowerment should positively predict infants' VWM due to its close links with IPV, the lack of evidence from previous research precludes stating a priori hypotheses. Last, we hypothesised that caregivers' and their infants' VWM scores will be positively associated. The hypothesis is informed by the demands of caregiving increasingly hindering engagement with one's infant the lower the resources available (including cognitive ones). Considering the inheritability of cognitive skills (Cuevas et al., 2014), the shared genetics and environment between caregivers and infants further support the expected relationship between their characteristics. Once again, we take a step-by-step approach, which starts by evaluating infants' VWM performance on its own. We then explore the maternal socio-emotional factors addressed in Study 1, namely maternal depressive thoughts, IPV, and sense of empowerment, in relation to infants' VWM performance. Last, we consider the links between mothers' and infants' VWM performances and explore the modulating role of maternal socio-emotional variables.

2.7 Method

2.7.1 Participants

Infants aged 6-months ±15 days of age or 9-months ±15 days were eligible for the study. Like in Study 1, the recruitment of participants happened within the sample of a larger project. Due to the nature of the VWM-PL task, parents of infants were screened for colour vision deficits. Infants of parents with any congenital problems, or gestational age <26 weeks at birth, were excluded from the study. Initially, 257 families came to the lab for the VWM-PL assessment in year 1. However, 17 families (10 6-month-old infants; 7 9-month-old infants) did not complete the assessment and were excluded from the study. The remaining 240 families were followed up for the duration of the study which included the following: (1) a laboratory-based VWM-PL assessment in year 1 at 6 or 9 months of age; (2) a laboratory-based VWM-PL assessment in year 2 at 18 or 21 months of age. Enrolment for this study was distributed over 4 phases. Each phase was separated by three months and involved enrolling approximately 60 infants for data collection.

In Year 1, 228 participants contributed to the visual working memory data. This includes infants aged 6 months old (N = 119; N females = 58) and infants aged 9 months old (N = 109; N females = 56). Out of the 228 participants in Year 1, 188 participants contributed visual working memory data in Year 2. This includes infants at 18 months of age (N = 94; N females = 45) and infants at 21 months of age (N = 94; N females = 46). Participants in Study 2 were the infants of the mothers who participated in Study 1. It must be noted that more data was lost for mothers, who completed multiple measures (i.e., EPDS, IPV, empowerment), than for infants, who only completed the VWM task. That means that our sample of infants was larger (N = 228) for the evaluation of VWM on its own, but was later reduced when linking infant VWM to the maternal socio-emotional measures (EPDS, N = 131 dyads; IPV, N = 104 dyads; sense of empowerment, N = 138 dyads) after the log transformation of EPDS and IPV scores. The sample with all measures needed to connect maternal VWM to infant VWM was 120 dyads. As mentioned in Study 1, all participants were from the Shivgrah block of Uttar Pradesh, India. The experiment was conducted with the understanding and written consent of each participant's parents. Where parents were illiterate, a witness gave signed consent accompanied by a thumb impression of the caregiver in place of a signature. At the end of each laboratory session, families received a small token of appreciation.

2.7.2 Materials

2.7.2.1 VWM-PL Task infant

We administered the same VWM-PL task, using the same stimuli as Study 1 (see figure 2.3)with the exception of variations in the size of the sets of squares. In Year 1, when participants were 6- and 9- months of age, the stimuli consisted of 1, 2 and 3 coloured squares (set size) corresponding to low, medium and high load conditions respectively. Previous studies have shown that infants around age 13 months of age show above chance performance for change preference scores at set sizes 1,2,3 whereas older infants and toddlers show above chance change preference scores at loads 2, 4 and 6 (Ross-sheehy et al., 2003; Simmering, 2016b). Thus, we varied the VWM load between 1, 2 and 3 squares on each side for infants at 6- and 9 -months of age. Similarly, we varied the load between 2, 4 and 6 items for children in Year 2, when they were of 18 and 21 months of age. Participants were presented with a total of 36 total trials in six blocks of 6 trials. Like in Study 1, each trial lasted for 10s. In terms of the set-up, a 42-inch LCD monitor, connected to a PC running Experiment Builder, was used to display the stimuli. Looking data were collected using an Eyelink 1000 Plus eye-tracker (SR Research) operating in binocular remote mode, at 500Hz. In the cases in which eye-tracking data were not available (due to reflection, poor lighting, or unwillingness of the infant to wear the calibration sticker), looking data were collected with a webcam. This was hand-coded on a frame-by-frame basis offline for left or right looking by a blind observer using Datavyu (D. Team, 2014). Infants sat on their mother's lap on a chair approximately 100cm from the screen. A target sticker was placed on the infant's forehead for the eye-tracking system to track the infant's head movement. Each display contained coloured squares that subtended a visual angle of 0.2°.

2.7.3 Procedure

Participants with their families were welcomed and seated in a common playroom, where testing procedures were explained and consent was obtained. Parents and their infants were, then, accompanied to a quiet room where testing was carried out. Parents were seated on a chair in front of the camera and infants sat on their laps. The task began with a short animated video of local cartoons while the participant settled in. A 5-point standard calibration sequence was administered to ensure correct eye-tracking at the top, bottom, left, right and centre of the screen. Following this, the VWM-PL task began.

Due to a range of constraints (e.g. power cuts, long travelling hours), we could not always rely on reviewing the data in real-time or while in India. Therefore, where infants and parents were willing to continue, additional blocks of data were sometimes collected. Each additional block contained 3 trials (one for each load) for each change side. Where necessary, participants could take a break between blocks. Participants completed on average 20.99 trials in Year 1 (*SD* = 9.72) and 26.25 trials in Year 2 (*SD* = 9.66). Each trial was preceded by an attention-getter presented at the centre of the screen until the infant looked toward it.

2.7.4 Measure Construction and Methods of analysis

The eye-tracking data was exported frame by frame using Data Viewer (SR-Research, Ontario, Canada). Since we also gathered video data to reduce data loss, the area of interest around the two objects on the screen was increased such that the eye-tracking data would match video-coded data. The primary distinctions were looking to the left, right, or away. In case of no recorded eye-tracking data or where the tracker was unable to register any looking on the screen during the 10-second trial, the hand-coded video data from that trial was included in place of the eye-tracking data. Hand coding was done frame-by-frame (30 frames per second) to determine looking to the left, centre, and right of the screen.

Cohen's Kappa was used to examine the reliability of the coded data. 17% of the data was re-coded to check reliability. The mean Kappa for the 6-month cohort of 0.73 and a mean Kappa for the 9-month cohort of 0.83 suggests good reliability. Note that Kappa values from 0.6 - 0.8 indicate substantial reliability; scores greater than 0.8 indicate almost perfect agreement. For the change preference analysis, we focused on the analysis period from 1500ms to 6500ms. This is the same window as maternal VWM. We trimmed the last few seconds of data from each trial as the number of eye-tracking samples diminished as infants' attention waned.

Maternal demographic and socio-emotional measures were computed in the same way as in Study 1. Logarithmic transformations were carried out once again for scores of EPDS and IPV.

2.7.4.1 Analytic strategy

As in Study 1, the analytic strategy was divided across different stages, focusing on 1) the infants' VWM unrelated to maternal measures, 2) the exploration of maternal socio-emotional measures as predictors of infants' VWM, and 3) the relationship between the VWM of mothers and their infants. Based on findings from Study 1, all analyses focused on the preferential looking performance when the first look was directed at the no change side.

In Stage 1, descriptive statistics and correlations between key variables are addressed. Next, infants' VWM is assessed on its own, with predictors directly related to them (i.e., infant age, Year of the study, and SES scores) and to the task (i.e., Load) being included in order to develop a minimal baseline model. No maternal measures were included in this Stage.

Stage 2 focuses on the influence of maternal socio-emotional measures on infants' VWM. As in Study 1, the loss of data resulting from including multiple variables in the same model (i.e., not all mothers completed all measures) and from transformations of skewed data prevents us from computing a single model which controls for all variables. Thus, three separate models are evaluated, one for each socio-emotional measure (depressive thoughts, experience of IPV, and sense of empowerment) in addition to the predictors in the base model from Stage 1. Within each of the three models, interactions between the socio-emotional measure and the predictors in the baseline model are explored.

Last, in Stage 3 links the VWM performances of infants and their mothers. The model includes the VWM scores of mothers and the infants' baseline model as predictors. Given that analyses from Study 1 indicated that maternal VWM performance was more noticeable at medium and high load conditions, we decided to exclude low load conditions and aggregate the scores from medium and high load conditions. When compared to the alternative linear mixed model, with the main and interactive effects of all maternal VWM scores (centred), the simplified model improved the AIC score, and therefore, the model's relative quality (assessed using the Anova function in R).

2.8 Results

2.8.1 Infant VWM baseline

The baseline model for the change preference measure of VWM was a linear mixed-effect model including the year (1 or 2), load (low, medium or high), SES Score (centred), and age cohort (cohort 1 = 6 and 18 months, cohort 2 = 9 and 21 months) as predictors. Year and age cohorts were difference-coded, and the load was input as a factor. To allow for individual differences across year and load, the model included a random intercept for each participant. To arrive at a minimal baseline model, we began with a model that only included the main effects. We then introduced two-way, three-way, and four-way interactions, only including interactive effects that showed evidence of improving the model fit (for detailed step-wise strategy and results on interaction effects, see Spencer et al., 2023). The final baseline model only included the main effects. All models were assessed for fit based on a Q-Q plot of the residuals and using the R package DHARMa (Hartig, 2021). Analyses are reported as Wald Chi-squared tests (see table 2.12). Results indicated the main effect of age and the main effect of load on infants' visual working memory. As the load increased, VWM performance decreased. Moreover, the cohort of 9-month-old infants (21-month-old in Year 2) performed significantly better than the other cohort. Figure 2.12 shows a visual representation of the results of the base model.

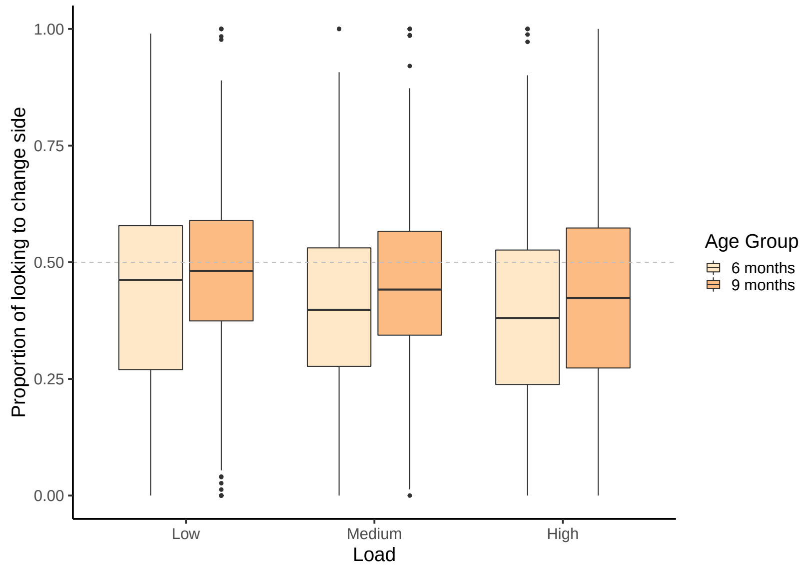


Figure 2.12: Infants had lower visual working memory scores with increased load. 9-month cohort infants show higher visual working memory scores across the loads (dark orange) compared to 6-month cohort infants (light orange).

2.8.2 Infant VWM and maternal socio-emotional measures

We ran three linear mixed-effect models with the socio-emotional measures as well as the infants' VWM base model (i.e., year, load, SES Scores, and age cohort) as predictors. Each of the socio-emotional measures (EDPS, IPV, and sense of empowerment) was included in a separate model (see Figure I.1, Figure I.2 and Figure I.3 in Appendix for Scatterplots between infants VWM-PL scores and maternal socio-emotional measures). We, also, introduced two-way interactions between the maternal socio-emotional measures (EPDS, IPV, or sense of empowerment scores) and the load, SES Score, and Year. No other interactions were modelled.

EPDS Results indicated no significant main effect of log EPDS scores on infants' VWM scores. No modelled interactions were found to be significant for maternal depressive thoughts either. Only infants' age predicted VWM performance (see table 2.13 for Wald Chi-squared test).

IPV As in the EPDS model, we found no significant main or interaction effects of maternal experiences of IPV on their infants' VWM scores (see table 2.14 for Wald Chi-squared test). Significant effects from the infants' VWM base model (age and load) remained significant after including IPV and its interactions.

Sense of Empowerment Consistent with the analysis in Study 1, we added the infants' gender (male, female) as a predictor along with the variables included in the EPDS and IPV models. Given the addition of gender, two-way interactions were added between it and empowerment, load, SES Score, and year. Additionally, three-way interactions were modelled. All three-way interactions included Empowerment and Gender, with each one also including the Load, Year, or SES. Results indicated no significant main or interaction effects of the sense of empowerment on infants' VWM scores (see table 2.15 for Wald Chi-squared tests). As in the EPDS model, only age remained significant from the base model.

2.8.3 Infants' & Mothers' VWM

The model on infants' VWM included maternal VWM scores (aggregated across the medium and high load) and maternal age as well as the SES score as predictors (based on results from Study 1). Additionally, we included fixed effects of infant age (6 or 9 months), Year (1 or 2), infants', and VWM Load condition (low, medium, high; based on infant VWM base model). All two-way interactions between SES Score (centered), maternal age (centered), and maternal VWM, and the three-way interaction between them, were also modelled. A random effect for participants was included to account for variability within the participants. The model was assessed for fit based on a Q-Q plot of the residuals and using the R package DHARMa (Hartig, 2021).

For scatterplot, histogram and correlation between infants' VWM-PL scores (divided by Year and Load) and their predictors (i.e., Maternal VWM-PL scores, Maternal Age, SES Scores, and Infant age, see Figure J.1, Figure J.2, Figure J.3, Figure J.4, Figure J.5, and Figure J.6 in Appendix.

As in the infants' base model, age and load were found to be significant predictors of the infants' VWM scores. Additionally, a main effect of maternal VWM and the three-way interaction effect of SES Scores, maternal VWM scores, and maternal age were also significant (see Table 2.16).

Figure 2.13 graphically depicts the data suggesting that maternal VWM performance for younger mothers across the SES background, as well as older mothers from higher SES background, tends to have an inverse relationship with infants' VWM performance. However, older mothers' from lower SES backgrounds tend to have a positive relationship with infants' VWM scores - though not strong as the other group.

2.8.4 Discussion

Study 2 examined the VWM of infants over time and explored its relationships with the maternal socio-emotional context and VWM. Results from the infants' VWM base model indicated that, overall, the older cohort (9 months old in Year 1 and 21 months old in Year 2) performed better at VWM tasks than the younger cohort (at 6 and 18 months of age). Additionally, both age groups showed a decline in performance as the task got more difficult. That is, as the load increased, infants' proportion of looking to the change side decreased. These findings converge with previous research indicating developmental changes in infants VWM (Ross-sheehy et al., 2003; Oakes et al.,

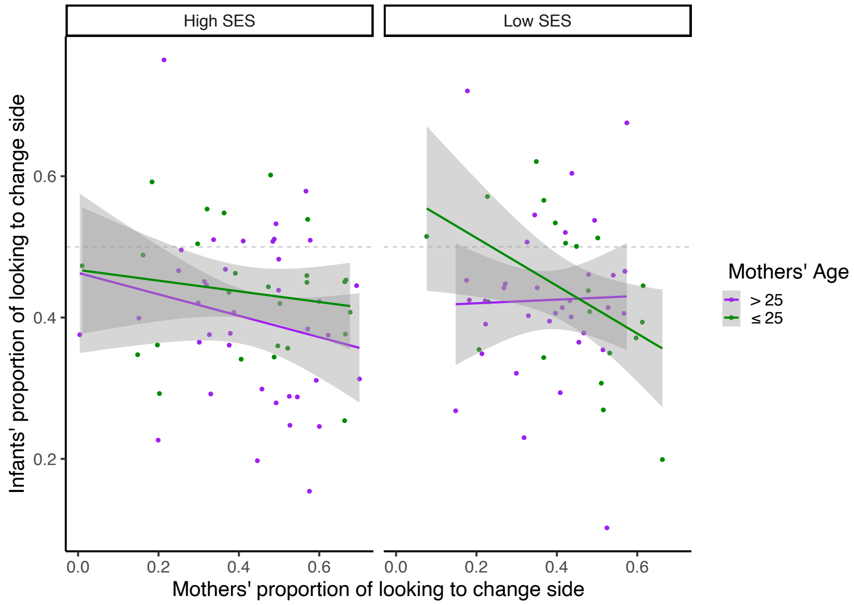


Figure 2.13: Effect of maternal VWM scores, age and socio-economic score on infants' VWM scores. Maternal age and SES Scores were median split for the purpose of visualisation.

2013). It must also be noted that the Year of data collection did not predict the VWM performance of infants. Results are not surprising given that the task was more difficult in Year 2 (for 18 and 21 months-old infants). In Year 1, the stimuli consisted of 1, 2, and 3 coloured squares for each load condition, but in Year 2, the number of squares per load condition was doubled. The fact that performance was equivalent across years (it was neither better nor worse) indicates that the change in square set sizes served to keep the difficulty of the task stable relative to the developmental growth of participants over time.

With regards to maternal socio-emotional context and infants' VWM scores, results indicated no main or interaction effects of maternal depressive thoughts, the experience of IPV, or sense of empowerment on infants' VWM performance. It is worth considering, however, that scores of depressive thoughts and IPV were very low in our sample, precluding a comprehensive evaluation of their potential influences. The fact that increases among the lowest scores on the scales do not predict infant VWM does not necessarily entail that findings would replicate with the inclusion of greater levels (even if only mid-range) of depressive thoughts and experiences of IPV. It is also important to note that in the population that we studied, families are larger than in western settings, and infants frequently have multiple caregivers, including older siblings (often sisters) and grandparents. It may be that a more collectivistic approach to caregiving serves as a protective factor in the development of infants against the potential negative influences of maternal depressive thoughts, experiences of IPV, and lack of empowerment. Future research that takes into account not only the maternal socio-emotional characteristics but also the degree to which caregiving time is shared with other family members and the interaction between both will help enhance our understanding of the development of infants' visual cognition in relation to maternal characteristics in collectivistic cultures.

In terms of mothers' performance on the VWM task and its relation to infants' VWM, results indicated that the higher the VWM capabilities of mothers, the lower the VWM capabilities of their infants. There was also an interaction between mothers' VWM performance, her age, and SES Scores which indicated that the negative trend of results did not apply to the infants of older mothers from the lower SES. The overall negative association between the VWM of mothers and their infants clashes with what we would have logically expected based on the literature. That is, because caregiving interactions are underpinned by maternal visual cognition, which allows handling information and coping with caregiving demands, higher VWM should enable interactions of better quality, which would in turn foster the development of infant VWM. Moreover, neither bio-genetic views on the inheritability of traits and capabilities nor socio-contextual explanations serve to intuitively explain the trends in results given that mothers and infants in our study would share both genes and environmental influences to some degree.

Findings may be explained by the potential effects of a mismatch between the abilities of the mother and those of the infant. Taking the earlier consideration (in Study 1) regarding people's potential disengagement from simple tasks, it may be that mothers with lower VWM capabilities have a greater tendency to remain focused on their infants and the objects to which they are attending. In contrast, mothers with higher VWM capabilities may disengage, or at least divert their attention and that of their infants towards a broader range of elements. This could result in them overloading the infants' underdeveloped cognitive capabilities and inadvertently hindering their development. Moreover, mothers with greater VWM capabilities may multitask to a larger degree, which may, in turn, reduce the quality of the mother-infant interaction. A key limitation of the study, thus, is that the use of laboratory tasks does not allow evaluating the mother-infant interaction in a naturalistic setting. Explanations that consider said interactions remain speculative unless the laboratory data is related to the real-life behaviour of mothers and their infants. Note that we directly look at this issue in the final empirical chapter.

Last, looking at the overall picture of our findings, it is important to bear in mind that considerations of sample size and missing data guided our analytic approach toward focusing on direct effects. However, it is possible that although there was no direct relationship between maternal socio-emotional characteristics and infant VWM in our sample, there may be indirect associations through relevant variables including maternal VWM. Indeed, findings of Study 1 revealed connections between socio-emotional variables and maternal VWM and Study 2 linked the latter to the cognition of their infants. Other indirect relationships may be revealed through maternal responsiveness and sensitivity during caregiver-infant interaction. Future research could consider the role of more proximal antecedents of infant VWM in connecting maternal socio-emotional characteristics to their infants' developmental outcomes.

2.8.5 General Discussion

Studies in the current chapter have provided an initial understanding of our sample, of how the socio-contextual characteristics of mothers relate to their VWM capabilities, and of how these relate to the VWM of their infants. Study 1 revealed that, in our sample, specific elements of the maternal socio-emotional context, including the experience of domestic violence and the sense of empowerment, were associated with their visual cognition. Moreover, Study 2 served to link maternal VWM to that of their infants. Interestingly, findings in Study 2 opposed an intuitive application of knowledge from the literature, with higher VWM among mothers relating to poorer VWM performance among their infants. Potential explanations for the findings remain speculative and require connecting the laboratory findings to real-life caregiver-infant interactions.

Taken together, Studies 1 and 2 also invite future research. Larger and more comprehensive models will enable us to account for indirect effects from the socio-emotional characteristics of mothers to the visual cognition of their infants via maternal cognitive capabilities. Similarly, further research connecting laboratory measures to real-life interactions will serve to clarify our findings and enhance the understanding of infants' cognitive development. It is the latter point what this thesis will aim to cover in Chapter5. However, before proceeding to address the real-life data of mother-infant dyadic interactions, it is imperative to elaborate on the methodological approach. Chapter 3, therefore, addresses the selection and extraction of variables of visual cognition from video recordings which can be linked to the laboratory data. It also elaborates on the development of the machine learning pipeline that enables the analysis of our data.

Table 2.1: Summary of key demographic features of our participants from Shivgarh, UP, India.

Participant Demographics				
Completed Measures		EPDS n = 187	IPV & Emp n = 138	VWM n = 125
Age in Years	Mean	26.4	26.1	26.2
	(SD)	(4.8)	(4.4)	(4.5)
	Median	25	25	25
	[min, max]	[18, 47]	[18, 42]	[18, 42]
Maternal Education				
Primary or lower	n (%)	102 (54.5)	72 (52.2)	65 (54.4)
Secondary or high		85 (45.5)	66 (47.8)	60 (48.0)
Income in INR				
<= 50,000	n (%)	85 (45.5)	51 (37.0)	49 (39.2)
50,001 - 1,19,999		67 (35.8)	57 (41.3)	51 (40.8)
>= 1,20,000		35 (18.7)	30 (21.7)	25(20.0)
Caste				
Scheduled Caste (SC) and Scheduled Tribe (ST)	n (%)	122 (65.2)	91 (66.0)	83 (66.4)
Other Backward Caste (OBC)		56 (29.9)	38 (27.5)	34 (27.2)
General		9 (4.8)	9 (6.5)	8 (6.4)
No. of Family Members				
	Mean	6.8	6.8	6.8
	(SD)	(3.1)	(2.9)	(3.0)
	Median	6	6	6
	[min, max]	[3, 25]	[3, 16]	[3, 16]
No. of Children Under 5				
	Mean	1.6	1.7	1.7
	(SD)	(0.7)	(0.7)	(0.7)
	Median	2	2	2
	[min, max]	[1, 4]	[1, 4]	[1, 4]
Access to Electricity				
Yes	n (%)	114 (61.0)	87 (63.0)	77 (61.6)
No		73 (39.0)	51 (37.0)	48 (38.4)
Cooking Fuel Type				
Wood	n (%)	136 (72.7)	113 (81.9)	102 (81.6)
Cow dung		22 (11.8)	8 (5.8)	6 (4.8)
LPG		29 (15.5)	17 (12.3)	17 (13.6)
Toilet Type				
Toilet at home (flush or pit)	n (%)	43 (23.0)	31 (22.5)	32 (25.6)
Open defecation		144 (77.0)	107 (77.5)	93 (74.4)

Note. EPDS, Edinburgh Post-natal Depression; IPV, Intimate Partner Violence; Emp, Empowerment; VWM, Visual Working Memory.

Table 2.2: Means, standard deviations and correlations between key maternal measures

Maternal Measures	Mean	SD	Range	1	2	3	4	5	6
1. EPDS	5.11	5.23	0-21	1	.29 ***	-.28 **	.13	-.13	.06
2. IPV	4.15	4.66	0-21		1	-.53 ***	-.24***	-.14	-.02
3. Empowerment	6.23	2.64	0-10			1	.16	.01	-.22 ***
4. Mother Age	26.16	4.35	18-42				1	-.14	.04
5. SES Score	9.73	3.82	4-26					1	-.04
6. Infant Gender									1

*Note. EPDS stands for Edinburgh postnatal depression score. IPV stands for Intimate Partner violence. SD stands for standard deviation. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table 2.3: Regression results for Maternal Measures with EPDS Scores as criterion

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	sum sq	df	F statistics	p-value
(Intercept)	281.39	1	396.49	<.001 ***
SESScore_s	0.27	1	0.384	.53
MotherAge_s	1.84	1	2.59	.12
Gender_s	1.55	1	2.19	.14
SESScore_s:MotherAge_s	1.59	1	2.24	.14
SESScore_s:Gender_s	0.07	1	0.099	.75
MotherAge_s:Gender_s	0.58	1	0.83	.36
SESScore_s:MotherAge_s:Gender_s	2.11	1	2.97	.09 .
Residuals	88.00	124		

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table 2.4: Regression results for Maternal Measures with IPV as criterion

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	sum sq	df	F statistics	p-value
(Intercept)	168.18	1	225.88	.000
MotherAge	4.34	1	5.83	<.05 *
SES Score	0.00	1	0.00	.96
Gender	0.02	1	0.03	.87
SES Score:MotherAge	0.03	1	0.05	.83
SES Score:Gender	1.16	1	1.56	.21
MotherAge:Gender	0.53	1	0.72	.40
SES Score:MotherAge:Gender	1.78	1	0.27	.61
Residuals	72.22	97		

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table 2.5: Regression results for Maternal Measures with Empowerment Scores as criterion

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	sum sq	df	F statistics	p-value
(Intercept)	5317.02	1	794.13	.00
SES Score	0.03	1	0.00	.94
MotherAge	31.76	1	4.74	<.05 *
Gender	43.86	1	6.55	<.05 *
SES Score:MotherAge	3.73	1	0.56	.46
SES Score:Gender	10.12	1	1.51	.22
MotherAge:Gender	0.11	1	0.02	.90
SES Score: MotherAge:Gender	1.78	1	0.27	.61
Residuals	870.40	130		

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table 2.6: Mothers' VWM-PL regression results using proportion looking to change side over time as the criterion

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	1.24	1	.27
ot1	1.69	1	.19
ot2	0.05	1	.82
ot3	0.89	1	.35
ot4	0.02	1	.88
ot5	0.76	1	.38
Load	6092.21	2	<.001 ***
FirstLook_s	75669.66	1	<.001 ***
SESScore_s	1.76	1	.19
ot1:Load	464.71	2	<.001 ***
ot2:Load	1110.29	2	<.001 ***
ot3:Load	64.02	2	<.001 ***
ot4:Load	14.91	2	.001 ***
ot5:Load	87.18	2	<.001 ***
ot1:FirstLook_s	73.37	1	<.001 ***
ot2:FirstLook_s	0.25	1	0.615
ot3:FirstLook_s	224.77	1	<.001 ***
ot4:FirstLook_s	62.43	1	.001 ***
ot5:FirstLook_s	45.93	1	<.001 ***
Load:FirstLook_s	4981.08	2	<.001 ***
ot1:SESScore_s	1.26	1	.26
ot2:SESScore_s	8.83	1	<.01 **
ot3:SESScore_s	0.88	1	.35
ot4:SESScore_s	4.77	1	<.05 *
ot5:SESScore_s	9.13	1	<.01 **
Load:SESScore_s	742.36	2	<.001 ***
FirstLook_s:SESScore_s	163.39	1	<.001 ***
ot1:Load:FirstLook_s	512.78	2	<.001 ***
ot2:Load:FirstLook_s	71.16	2	<.001 ***
ot3:Load:FirstLook_s	584.00	2	<.001 ***
ot4:Load:FirstLook_s	146.95	2	<.001 ***
ot5:Load:FirstLook_s	191.89	2	<.001 ***
ot1:Load:SESScore_s	352.99	2	<.001 ***
ot2:Load:SESScore_s	87.58	2	<.001 ***
ot3:Load:SESScore_s	170.64	2	<.001 ***
ot4:Load:SESScore_s	241.16	2	<.001 ***
ot5:Load:SESScore_s	748.83	2	<.001 ***
ot1:FirstLook_s:SESScore_s	0.83	1	.36
ot2:FirstLook_s:SESScore_s	1251.70	1	<.001 ***
ot3:FirstLook_s:SESScore_s	371.96	1	<.001 ***
ot4:FirstLook_s:SESScore_s	31.26	1	<.001 ***
ot5:FirstLook_s:SESScore_s	655.76	1	<.001 ***
Load:FirstLook_s:SESScore_s	822.67	2	<.001 ***
ot1:Load:FirstLook_s:SESScore_s	1431.26	2	<.001 ***
ot2:Load:FirstLook_s:SESScore_s	1268.27	2	<.001 ***
ot3:Load:FirstLook_s:SESScore_s	482.87	2	<.001 ***
ot4:Load:FirstLook_s:SESScore_s	148.99	2	<.001 ***
ot5:Load:FirstLook_s:SESScore_s	1182.75	2	<.001 ***

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table 2.7: Regression results for maternal VWM-PL performance with the proportion of looking to change side Scores as criterion

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	493.95	1	<.001 ***
Load	14.41	2	<.001 ***
SESScore_s	0.04	1	.85
MotherAge_s	0.03	1	.85
Load:MotherAge_s	9.51	2	<.01 **

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table 2.8: Results for maternal VWM-PL scores with EDPS scores as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	77.44	1	<.001 ***
Load	8.39	2	<.05 *
SESScore_s	0.08	1	.78
MotherAge_s	0.46	1	.50
log(DepTotal)	0.18	1	.68
Load:MotherAge_s	1.73	2	.42
Load:log(DepTotal)	2.61	2	.27
SESScore_s:log(DepTotal)	0.0005	1	.98
MotherAge_s:log(DepTotal)	0.62	1	.43
Load:MotherAge_s:log(DepTotal)	0.44	2	.80

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table 2.9: Results for maternal VWM-PL scores with IPV scores as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	105.79	1	<.001 ***
Load	5.06	2	.08 .
SESScore_s	3.43	1	.06 .
MotherAge_s	0.25	1	.62
log(AbuseScore)	0.11	1	.74
Load:MotherAge_s	13.82	2	<.001 ***
Load:log(AbuseScore)	1.09	2	.58
SESScore_s:log(AbuseScore)	4.99	1	<.05 *
MotherAge_s:log(AbuseScore)	1.22	1	.27
Load:MotherAge_s:log(AbuseScore)	10.98	2	<.01 **

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table 2.10: Results for maternal VWM-PL scores with Sense of Empowerment scores as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	51.32	1	<.001 ***
Load	2.47	2	.29
SESScore_s	1.17	1	.28
MotherAge_s	0.05	1	.82
EmpowermentScore	0.10	1	.75
Gender_s	0.29	1	.59
Load:MotherAge_s	0.35	2	.84
Load:EmpowermentScore	0.18	2	.91
SESScore_s:EmpowermentScore	0.99	1	.32
MotherAge_s:EmpowermentScore	0.01	1	.93
Load:Gender_s	2.18	2	.34
SESScore_s:Gender_s	7.66	1	<.01 **
MotherAge_s:Gender_s	0.02	1	.88
EmpowermentScore:Gender_s	1.0	1	.30
Load:MotherAge_s:EmpowermentScore	1.93	2	.38
Load:MotherAge_s:Gender_s	0.98	2	.61
Load:EmpowermentScore:Gender_s	1.84	2	.40
SESScore_s:EmpowermentScore:Gender_s	8.32	1	<.01 **
MotherAge_s:EmpowermentScore:Gender_s	0.02	1	.90
Load:MotherAge_s:EmpowermentScore:Gender_s	1.16	2	.56

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table 2.11: Results for maternal VWM-PL scores with Sense of Empowerment scores as a predictor for mothers of female infants only.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	19.45	1	<.001 ***
Load	1.44	2	.49
SESScore_s	5.46	1	<.05 *
MotherAge_s	0.001	1	.97
EmpowermentScore	0.20	1	.66
Load:MotherAge_s	0.59	2	.75
Load:EmpowermentScore	0.61	2	.74
SESScore_s:EmpowermentScore	5.91	1	<.05 *
MotherAge_s:EmpowermentScore	0.0004	1	.98
Load:MotherAge_s:EmpowermentScore	1.97	2	.37

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table 2.12: Infant VWM-PL regression results with Load, Year, SES Score and Age as criterion (Infant VWM Base Model)

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	1257.21	1	<.001 ***
Load	11.11	2	<.01 **
Year_s	2.22	1	.14
SESScore_c	0.59	1	.44
Age_s	11.67	1	<.001 ***

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table 2.13: Infant VWM-PL regression results with maternal log EPDS scores as criterion

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	206.18	1	<.001 ***
Age_s	9.86	1	<.01 **
Load	1.15	2	.56
Year_s	0.17	1	.68
SESScore_c	0.03	1	.86
log(DepTotal)	0.84	1	.36
Load:log(DepTotal)	2.92	2	.23
Year_s: log(DepTotal)	0.10	1	.75
SESScore_c:log(DepTotal)	0.06	1	.80

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table 2.14: Infant VWM-PL regression results with maternal log IPV scores as criterion

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	159.16	1	<.001 ***
Age_s	5.40	1	<.05 *
Load	7.50	2	<.05 *
Year_s	0.21	1	.65
SESScore_c	1.44	1	.23
log(AbuseScore)	3.53	1	.06 .
Load:log(AbuseScore)	2.52	2	.28
Year_s:log(AbuseScore)	0.03	1	.85
SESScore_c:log(AbuseScore)	1.96	1	.16

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table 2.15: Infant VWM-PL regression results with maternal sense of empowerment scores as criterion

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	124.13	1	<.001 ***
Age_s	6.97	1	<.01 **
Load	2.88	2	.24
Year_s	0.10	1	.75
SESScore_c	0.23	1	.64
EmpowermentScore	0.34	1	.56
Gender_s	0.09	1	.76
Load:EmpowermentScore	0.63	2	.73
Year_s:EmpowermentScore	0.15	1	.70
SESScore_c:EmpowermentScore	0.77	1	.38
Load:Gender_s	1.56	2	.46
Year_s:Gender_s	2.10	1	.15
SESScore_c:Gender_s	0.11	1	.74
EmpowermentScore:Gender_s	0.58	1	.45
Load:EmpowermentScore:Gender_s	1.63	2	.44
Year_s:EmpowermentScore:Gender_s	2.79	1	.09 .
SESScore_c:EmpowermentScore:Gender_s	0.11	1	.74

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table 2.16: Regression results for Infants VWM scores with Maternal age and VWM scores as predictors

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	224.85	1	<.001 ***
Age_s	6.33	1	<.05 *
Load	6.12	2	<.05 *
Year_s	1.94	1	.16
SESScore_c	0.31	1	.58
Prop_mom_NC	4.11	1	<.05 *
MotherAge_s	0.05	1	.83
SESScore_c:Prop_mom_NC	0.67	1	.41
SESScore_c:MotherAge_s	3.32	1	.07 .
Prop_mom_NC:MotherAge_s	0.11	1	.75
SESScore_c:Prop_mom_NC:MotherAge_s	4.98	1	<.05 *

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Chapter 3

Leveraging technological advances to overcome challenges in assessing visual cognition in naturalistic low-resource settings

Chapter 2 established the association between the visual cognitive abilities of the caregivers and their infants within our sample, i.e., mother and infant VWM are inversely related to each other in a lab-based task. However, in real-world settings, infants' visual experience is more complicated than simply attending to the stimuli on a screen. In this cluttered world, we want to be able to understand how infants deploy their visual attention to interact with the objects and people around them. For example, we might assess the duration of looks to an object, how often they shift between objects or objects and faces, and so on. Given that in the real-world setting, infants engage with their social partners who have their own visual experience, and attention abilities, we also need to understand how both deploy visual cognition when interacting in a naturalistic setting.

In this context, it is important to capture the dynamic interaction between parents and their infants in a meaningful way by extracting key measures of visual cognition (e.g., look duration, and shift rate) within the ongoing stream of interaction. Previous anthropological and psychological research on parent-infant interaction has used a third-person view to capture parent-infant interaction in low-resource settings and analysed data by annotating the videos frame by frame. This has created two main issues: 1) a limited amount of participants' data can be collected because of the time and effort that goes into frame-by-frame coding, and 2) the third-person perspective offers only limited insight into parents' and infants' own visual experiences. The current chapter focuses on using technological advances to create a pipeline to collect, process, and analyse parent-infant interaction data from low and middle-income countries using open-source and freely accessible algorithms.

Recently researchers in western, high-resource contexts have quantified parent-infant interactions using innovative technology (Aslin, 2009; L. B. Smith et al., 2011, 2015; Yoshida & Smith, 2008; C. Yu & Smith, 2012). One such technology is the head-mounted eye-tracker which has been established to be reliable equipment to quantify parent-infant interactions in naturalistic settings. Unlike other technologies (e.g., hand-held cameras), head-mounted cameras and eye trackers provide the closest possible approximation to what individuals see and how they deploy their attention (e.g., being sensitive to head movements). Using these technologies, researchers have found that, in an active context, the visual dynamics of parent and infant are very different compared to a third-person's view. For example, L. B. Smith et al. (2011) recorded a ten-minute toy play session between parents and their 17- to 19-

month-old infants using head-mounted cameras and eye-trackers with the aim to understand infants' first-person visual experiences. Results suggested that the infants' first-person view was highly selective with one dominating object (toy) in view thus blocking the view of other objects around the infant. The study defined a dominating object as an object whose relative size in view was at least double the size of all other objects combined. In contrast, the parents' first-person view was broader and more stable. Parents tended to shift their gaze between visual targets (toys, hands and infant's face) rapidly with all objects equally in view. Interestingly, infants' momentary visual experience included more hands (own or partners') and hands manipulating objects than parents' momentary visual experience (also see Yoshida & Smith, 2008). Therefore, head-mounted eye trackers and cameras enable us to capture first-person visual experiences which can systematically differ from a third-person perspective (Aslin, 2009) and are not intuitive relative to the adult third-person view (Yurovsky, Smith, & Yu, 2013).

A key explanation for the difference in visual experience comes from the fact that infants' bodily movements such as turning heads or reaching for a toy, together, have a major influence on their visual dynamics (Schneiberg, Sveistrup, McFadyen, McKinley, & Levin, 2002; Yoshida & Smith, 2008; L. B. Smith et al., 2011). In addition to infant's own bodily dynamics playing a role in selecting their visual experience, parents also played a complementary role in selecting targets for the infant's momentary visual experience (Xu, Chen, & Smith, 2011) allowing infant and caregiver to spontaneously choose their visual stimuli within the constraints of their own bodily actions. Franchak, Kretch, Soska, and Adolph (2011) examined the visual experience of 14 months old, mobile infants, using head-mounted eye-trackers during a parent-infant play session. They noted that infants frequently fixated on caregivers' hands and bodies instead of caregivers' faces. Moreover, infants were more likely to look at the mother's face if the mother was sitting down at the infant's eye level versus standing upright. Given the infant's own bodily constraints, the caregiver's contribution to the infant's visual experience, and the fact that it is common for caregivers to sit on the floor while interacting with their infants in rural India, we wanted to examine these key factors that might impact infants' visual dynamics.

Extending the use of these tools to rural India is also important given the western sampling bias in psychological research. Much of the research in psychology comes from middle-class white families. Lack of technological infrastructures such as consistent access to electricity or power cuts as well as lack of portable equipment have made it difficult for researchers to collect data from rural low- and middle-income countries (LMIC). Given the aforementioned technological advances such as head-mounted eye-trackers, it is now possible to collect data from countries with a lack of technological infrastructure. To our knowledge, no study has incorporated head-mounted eye-trackers and head cameras in low-resource settings. In the present study, we use head-mounted eye trackers and head cameras to extract measures of visual cognition during parent-infant interaction in Norwich, UK (urban UK) and Shivgarh, India (rural India). By having both lines of video recording together, we can look at the dynamic interaction between parents' and infants' visual worlds, thus extracting measures of visual cognition across social contexts.

Technology enabling data collection goes hand in hand with the development of data processing tools such as machine learning algorithms. Machine learning algorithms can enable us to process data quickly and efficiently. Because of this, we are now able to collect data such as videos from head-mounted cameras from a large data set, extracting variables of interest such as joint attention and sustained attention.

In this context, the core of Chapter 3 focuses on the process of creating a common methodological pipeline to quantify parent-infant interaction that can be used across high and low resource settings, in a way that ensures that we are extracting meaningful data connected to what parents and their infants view. Our goal is to quantify how long each partner in the dyad is looking at an object or how much they are shifting between objects and faces, whether they are looking at the same object at the same time, how long they sustain attention on objects and faces, and who is leading and who is following each attentional episode across both settings. Moreover, the current chapter focuses on using open-source machine learning algorithms with the goal of avoiding laborious frame-by-frame hand coding. This can enable the processing of large datasets as well as objectively code data in a way that is fully transferable across cultures.

The present chapter proceeds as follows. First, we introduce the data collection methods used in both the UK and rural India. Next, we discuss the machine learning tools used to process the data set, including a face-detection network and an object recognition network. For each machine learning approach, we discuss how the networks were trained and validated, including the key metrics we focused on and methods used to optimise performance. We then present an overview of the full processing pipeline that applied these machine learning tools to the processing of the eye-tracking and head camera data, including a discussion of a toolkit for analysing the resultant time series data. Next, we apply this pipeline to a subset of 12 infant-caregiver dyads: 4 from a 6-month-old UK cohort, 4 from a 6-month-old India cohort, and 4 from a 9-month-old India cohort. Here, we validate the pipeline performance by quantifying the accuracy of the resultant data set relative to hand-coded data. We also quantify – for the first time – differences in dyadic interactions across high- and low-resource settings. This work sets the stage for a larger cross-cultural comparison of dyadic interactions which is presented in Chapter 4.

3.1 Data Collection Methods

3.1.1 Participants

We present data from 4 parent-infant dyads (3 females) that were recruited by the Developmental Dynamics Lab at the University of East Anglia, UK, for an ongoing longitudinal project on early brain development. Inclusion criteria for dyads included (1) normal or corrected-to-normal vision; (2) uncomplicated single birth between 37 and 42 weeks; (3) no reports of alcohol or drug illicit use during pregnancy; (4) no pre-existing neurological conditions or major head injury; (5) no familial history of major depressive or psychiatric illness confirmed during the parental interview during enrolment. Parents were informed of the experiment's aim and procedure, and written consent was obtained. Remuneration comprised of 20 pounds, travel expenses, a t-shirt and a toy for each participant.

The Indian sample was recruited with the help of the Community Empowerment Lab (CEL) which works in the rural area of Shivgarh in the state of Uttar Pradesh. Uttar Pradesh constitutes one of the highest infant mortality rates in India with 60 deaths per 1000 live births for children under five years of age (see National Family Health Survey report 2019-2021). CEL, in partnership with the community, created an intervention that helped reduce the infant mortality rate in the intervention villages by 54% (Kumar et al., 2008).

The current study reports a subset of data from a longitudinal project which recruited 240 families. In the current study, we processed and analysed data from eight dyads. Of the eight dyads, four infants were aged 6 months ±15 days (3 females, M = 6.06 months, SD = 0.23 months) and four infants were aged 9-month ±15 days (3 females, M = 9.02 months, SD = 0.42 months). All the infants were full-term and typically developing.

This project was supported by the Bill & Melinda Gates Foundation Grant No. OPP1164153 and the NIH Grant No. R01HD083287, both were awarded to Prof. John P. Spencer. The data reported here is part of a larger study examining infant brain health in India and probing the neural basis of visual working memory in early development in the UK.

3.1.2 Materials

Mobile eye-tracking. Parent eye movements and caregivers' and infants' visual fields were recorded using light-weighted (36gms) mobile eye-trackers developed by Pupil Labs (Kassner, Patera, & Bulling, 2014). The eye tracker was used with the software Pupil Capture (versions 0.09 to 0.9.15). The eye-tracker has an infrared eye camera, placed close to the eye, that recorded monocular pupil and corneal reflections from the images of one of the eyes at a resolution of 640x480 pixels and a sampling rate of 120 Hz. The sampling rate from the world camera was captured at 30 Hz at a resolution of 1280x720 pixels. The set-up includes connecting both the eye-tacker and head-mounted camera to Pupil Labs recommended mobile phones. For data collection in Shivgarh, we used Nexus 5XN4F2T mobile phones. For the UK, we used Google Pixel 2 mobile phones. Each mobile phone in both contexts consisted of Pupil Mobile apps that captured the data on the mobile phones. These mobile phones were in turn connected to a laptop (HP laptop in India and Mac laptop in the UK) through a WiFi network for the simultaneous streaming of the video. Here, the experimenter can leave the room and see a live gaze position in the form of a red dot relative to the headset. These eye-trackers interface with the world camera and record the subject's field of view while the eye-tracker records the eye movement. This open-source software combines the world camera information with pupil position information. The parent wears the head-mounted eye-trackers like glasses with the nose-piece placed on the nose. The infant's head-mounted camera is embedded into a headband for comfort as well as to avoid slippage. In line with previous research using head-mounted cameras, we placed the camera low on the infant's forehead (L. B. Smith et al., 2011).

3.1.3 Stimuli

Ten toys were organized into two sets with each set containing five toys in the UK. The toys include utensils, animals, and/or different shaped blocks of single main colour (see fig 3.1). If and when the toys broke during data collection, they were replaced by another toy. For example, a toy apple in the UK was replaced with a toy pear.

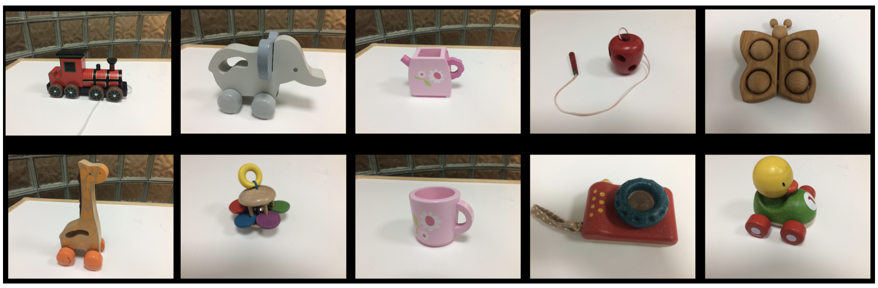


Figure 3.1: Set of toys used in the UK. Top (left to right) Train, Elephant, Kettle, Apple, Butterfly. Bottom (left to right) Giraffe, Rattle, Cup, Camera, Duck.

Similarly, ten toys were used in Shivagrh including objects used as toys in the local community (e.g. plate, spoon), familiar toys (e.g rattle, ball) and novel toys (see figure 3.2). If the toys broke during the one-year phase of data collection, we either replaced them with another toy or taped them with the exception of one toy (GLOW) as parents and infants continued to play with the dismantled toy.

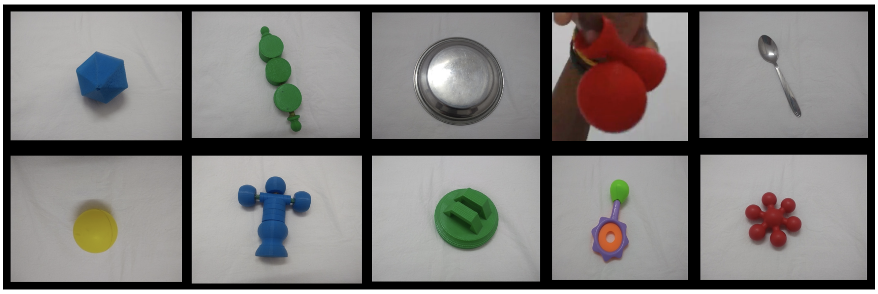


Figure 3.2: Set of Toys used in India. For some objects, we made up names to create labels to train them in the YOLO algorithm. Top (left to right) Blue ball, Candy, Plate, Glow, Spoon. Bottom (left to right) Yellow ball, Man, Green, Ratte, Puzzle.

3.1.4 Setup UK

Trained researchers visited the participant's home at the time when the parents confirmed that the infants were usually awake and fed. After obtaining the consent form from the parents, the eye trackers and tripod cameras were set up in an area where the parents typically played with the infant. Infants wore a vest to which the mobiles were attached at the back for freedom of movement. If the infant was on a boppy pillow, then the cable and phone were left on the side. Parents were briefed that they were free to pick up their infants and move the phone/cable as desired. The experimenter adjusted the scene and eye cameras when necessary. Parents wore a lab coat with a pocket or velcro at the back of the lab coat to attach the phone for freedom of movement. Two tripod-mounted cameras captured the play session from a third-person view (see figure 3.3 for an example of the setup). A LENA device was placed near parents and their infants.

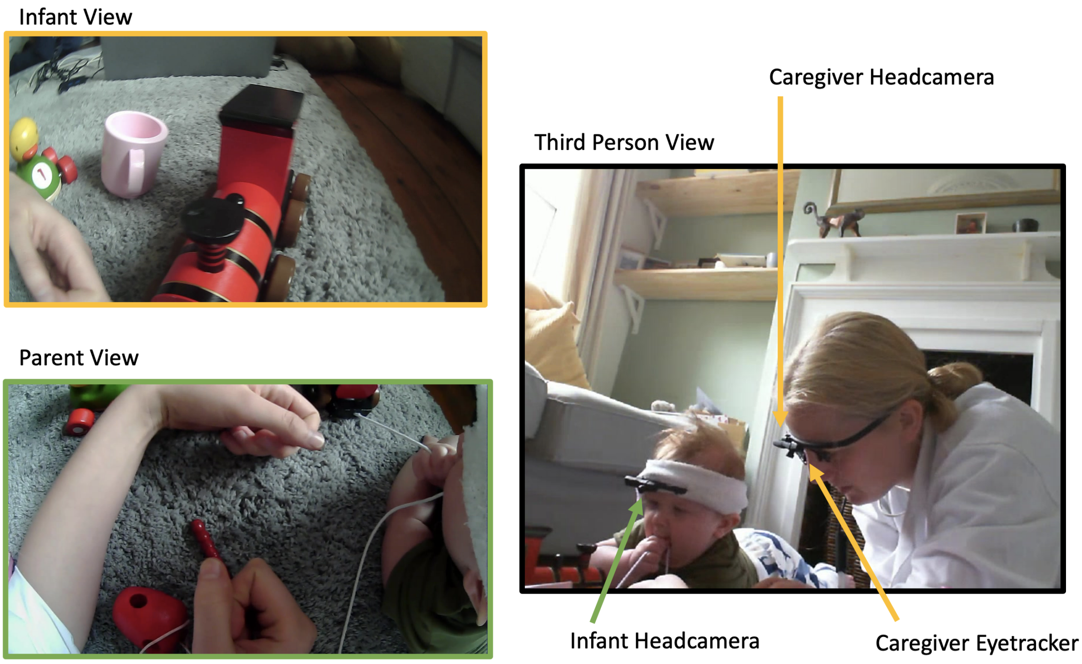


Figure 3.3: Caregiver-Infant interaction set up in the UK. Caregiver-infant dyads played together with a set of toys in a naturalistic setting. Each wore head-mounted cameras to collect egocentric video (left) and the caregivers also wore eye-tracker to track gaze positions. A stationary camera also recorded from a third-person perspective (right).

3.1.5 Setup India

Due to infrastructural and technological constraints (e.g. lack of electricity), the parent-infant interaction study took place in several rooms set up as an open laboratory space in a palace where we could bring in a generator for power cuts. Before the sessions began, families toured the laboratory while all procedures were explained to them. Families were shown the equipment, explained its function and were given the opportunity to ask any questions. They were, then, seated in a common playroom where consent was given.

The parent-infant interaction room consisted of a mattress on the floor as a play area. Two cameras were placed on opposite walls such that they record parent-infant interaction from a third-person view. Figure 3.4 shows an example of the setup for the head-mounted eye-trackers and cameras in India. The set-up for eye-trackers was the same as in the UK with the exception that caregivers in India did not wear lab coats and hence mobile phones were placed next to them. Caregivers felt uncomfortable wearing white lab coats for two reasons: 1) In the Hindu community in India, white is the colour of mourning, and 2) mothers' would feel shy wearing a lab coat that was very different from their usual attire. Caregivers were briefed that they were free to move around holding their mobile phones if needed. Culturally, sitting on the floor, cross-legged or in a squatting position is an everyday practice. The caregivers hardly ever moved around the room or even displayed the need to move around the room with their infants.

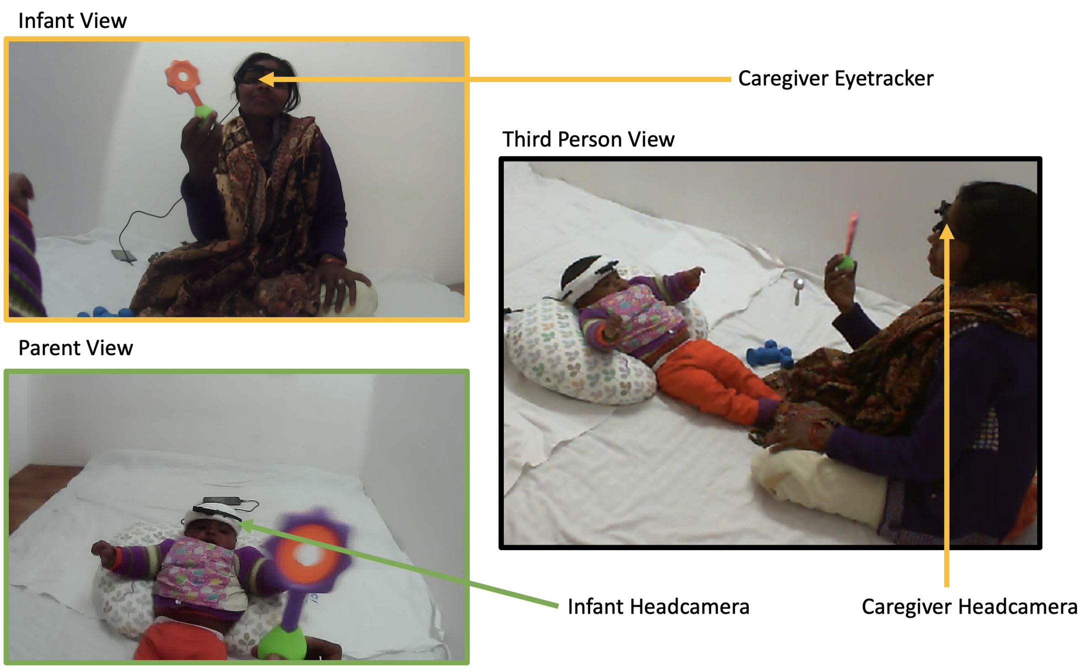


Figure 3.4: Caregiver-Infant interaction setup in Shivgarh, India. Caregiver-infant dyads played together with a set of toys in a naturalistic setting. Each wore head-mounted cameras to collect egocentric video and the caregivers also wore eye-tracker to track gaze positions (left). A stationary camera is recorded from a third-person perspective (right).

3.1.6 Procedure

There were two experimenters in the room. In India, one of the experimenters was a staff member from the Community Empowerment Lab (CEL) organisation and one was from the local community. Experimenter 1 helped the caregiver wear the head-mounted eye-tracker. Prior to calibrating the eye-tracker, Experimenter 1 would ask the parent to follow their finger (left, right, top, bottom) to make sure that the pupil was captured properly and to ensure that it was visible in the world camera. We start recording the parent eye-tracker to calibrate with a minimum of nine calibration points performed using Pupil Capture software (for more information see https://docs.pupil-labs.com/core/). During calibration, the caregiver fixated on the calibration marker (Appendix L.1) while keeping the head stationary, and the experimenter moved the marker around while staying within the participant's visual field (about 1.5-2m away). The experimenter moved the calibration marker in such a way that it covered the 2D screen that was monitored by experimenter 2. Following the calibration, one of the two experimenters would distract the infant with a toy while the other experimenter would place the headband on the child's head. Experimenter 1 moved the toy in different directions (top, down, left, right) while Experimenter 2 adjusted the angle of the camera to ensure that the toy was in the infant's field of view when they moved their heads in different directions. Once the cameras were set up, we started the infants' head camera and the LENA device recording. The experimenter placed the toys near the dyads within reach. Caregivers were instructed to play as they usually would with their infants. The experimenter placed a clapperboard between the parent and their infants' head camera such that it was visible on both the recordings as well as close to the LENA device and clap it three times to synchronise the onset of the play session. Both the experimenters would leave the room (India) or move to a corner of the room (UK) and check the play session as it streamed live on the laptop for any issues (such as removing the head camera, technical issues, software errors etc). The play sessions were recorded for approximately 10 minutes.

3.2 Setting up Machine Learning Tools

In recent years, the use of CNNs (convolutional neural networks) has gained popularity due to its vision-related applications including face detection (K. Zhang, Zhang, Li, & Qiao, 2016), object recognition (Redmon, Divvala, Girshick, & Farhadi, 2016) and image-based diagnostic applications such as detecting anomalies in X-ray and MRI images (K.-H. Yu, Beam, & Kohane, 2018). Here, we used several specific CNN tools to objectively detect the presence of faces and toys in the video data collected from the mothers' eye-tracker and the infants' head cameras.

3.2.1 Multi-Cascade Convolutional Neural Network for Face Detection

We used a publicly available multi-cascade convolutional neural network (MTCNN) available at https://github.com/ipazc/mtcnn focusing on the MTCNN face detection network built by K. Zhang et al. (2016). MTCNN is a fast, efficient and robust face detection algorithm (N. Zhang, Luo, & Gao, 2020) built to account for various illuminations and occlusion in real-world environments. It uses three steps to identify faces.

1. It proposes candidate facial windows in the image.

2. Then, it refines the facial windows by rejecting a large number of non-

face windows.

3. And finally, it refines the result to produce five landmark positions (see figure 3.5 for an example).

In Developmental science, we are aware of only one study using machine learning algorithms for face detection (Long, Sanchez, Kraus, Agrawal, & Frank, 2022). The study compares open-source neural networks such as Open Pose and MTCNN to understand changes in infants' visual fields across 8- to 16-month old infants. Long et al. (2022) note that while both MTCNN and Open Pose outperform ViolaJones classifier (see Viola & Jones, 2004), overall, MTCNN shows slightly better performance in face detection than Open Pose thus suggesting good accuracy for face detection. Moreover, unlike OpenPose, MTCNN is a specialised face detector that can be used both on static images and videos.

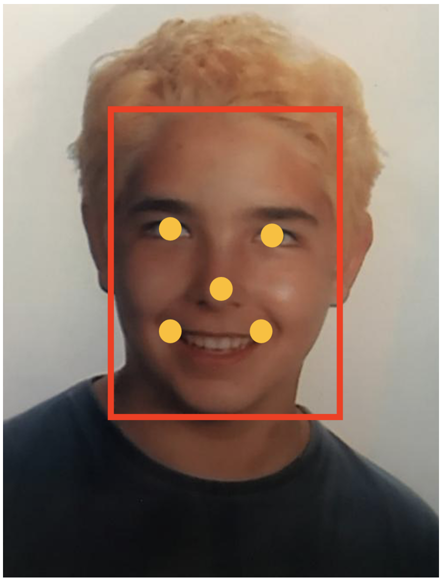


Figure 3.5: An example of five facial landmark positions for MTCNN detection on the left and right eye, nose, left and right edges of the mouth (circled in yellow). The red bounding box indicates the facial window.

We set up the MTCNN environment using Anaconda Navigator, a graphical user interface that helps manage the virtual environment and consists of several data science packages required for machine learning. We used the Tensorflow implementation of the MTCNN algorithm available at https://github.com/ipazc/mtcnn

To test the accuracy of the MTCNN algorithm, we extracted 10 images from 10 dyads (5 from the parent head camera and 5 from the infant head camera), including dyads that are not included in the current analyses. The final test data set for UK and India dyads consisted of 100 images each. Out of the 100 images, each data set consisted of 80 face images and 20 non-face images. Half of the images were of infant faces and another half of caregivers' faces. Images with various orientations, lighting and distance were selected (see figure 3.6 for an example of raw data). All the images were hand-labelled for faces using ImageJ, an open-source image processing and analysing software (Schneider, Rasband, & Eliceiri, 2012) yielding an a.ROI file for each image containing the coordinates of the labelled box. These hand-labelled data sets were called the "Ground Truth" (GT) data set. Faces were classified as present if at least half of the face was visible.

Figure 3.6: Example of a subset of raw data with and without faces. We used MTCNN algorithm for face detection for both India (left) and the UK (right) cohort.

To account for the impact of variation in lighting in the dyadic videos (e.g. sometimes poor lighting in the home context or face against the light), we use Contrast Limited Adaptive Histogram Equalization (CLAHE) which takes care of over-amplification of contrast (see Zuiderveld, 1994). We used CLAHE to check if it made any difference in the accuracy of face detection. We, then, ran these images through the MTCNN algorithm with and without CLAHE. This gives us a prediction file with the following information: 1) image name, 2) prediction confidence, and 3) box coordinates of the prediction (see figure 3.7 and figure 3.8 for an example output of predictions with and without CLAHE).

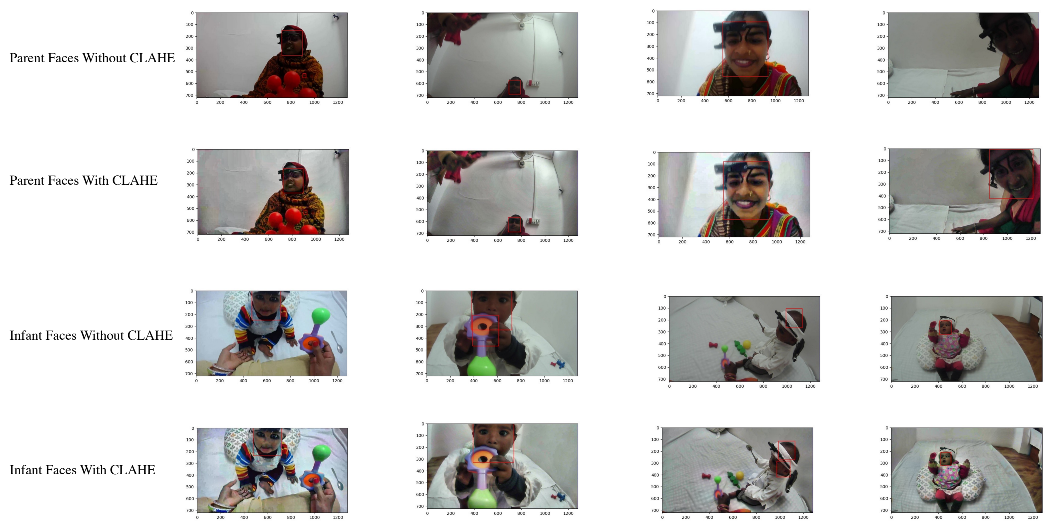


Figure 3.7: Example of MTCNN evaluation output with and without CLHAE for India cohort.

To evaluate the performance of the MTCNN face detector on our test data sets, we used an open-source evaluation metric developed by Padilla, Passos, Dias, Netto, and da Silva (2021). The source code was downloaded from GitHub onto a mini mac computer and a Python environment was created to run the software. The metric used Intersection Over Union (IoU) to measure

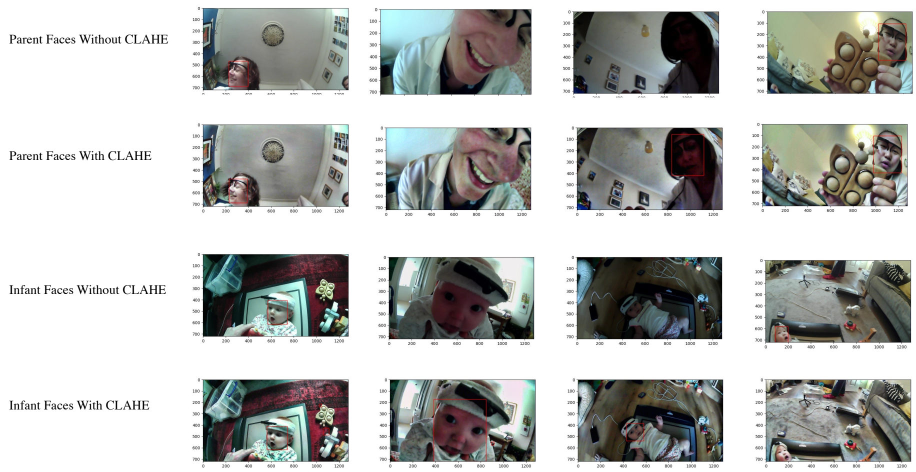


Figure 3.8: Example of MTCNN evaluation output with and without CLHAE for UK cohort.

the accuracy of detection (see figure 3.9). IoU provides us with information on the extent of overlap between the GT and the prediction (the greater the overlap, the greater the IoU). Figure 3.10 displays the GT in green and the prediction in red, with increasing overlap across images. Figure 3.10 A has the least overlap and IoU, and Figure 3.10 C has the largest.

To evaluate the precision and recall of the detections, it is necessary to establish an IoU threshold. The larger the threshold, the larger the IoU, and therefore overlap required. Figure 3.11 serves to further visualise the evaluation process. Each of these images, except figure 3.11G, has at least one target object of the class face. The GT labels are the bounding boxes in green. MTCNN predicted eight faces, represented by the red bounding boxes. In this case, the IoU would be higher in figure 3.11B than in figure 3.11E as the extent of overlap between the GT and the prediction is larger. The evaluation metric

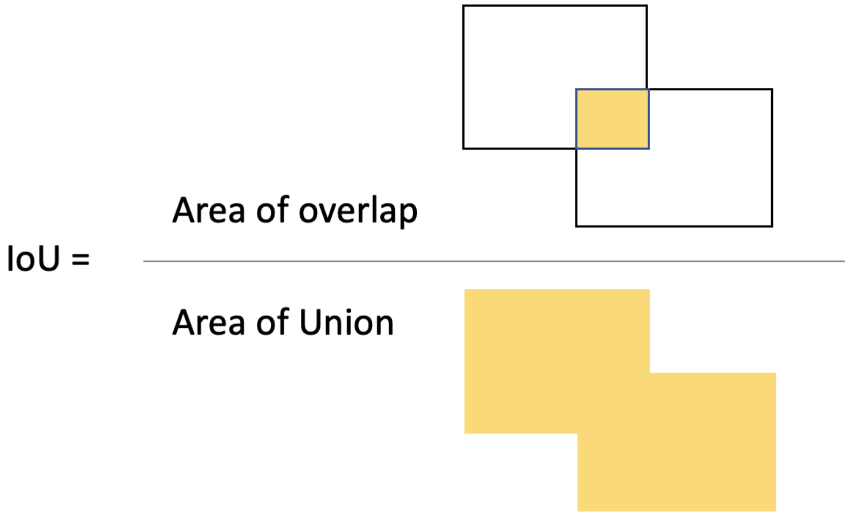


Figure 3.9: Illustration of Intersection over Union (IoU). The IoU is computed by dividing the area of overlap between the predicted and ground truth bounding boxes by the area of union between the predicted and ground truth bounding boxes.

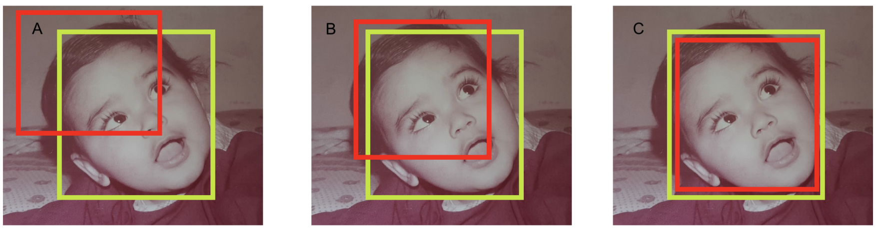


Figure 3.10: An example of detecting a face in the image and calculating IoU. Ground Truth is depicted in green and the predicted bounding box is depicted in red. Image A (extreme left) has the least overlap between the ground truth and predicted bounding boxes, and Image C (extreme right) has the most overlap.

will classify each detection as:

1. True Positive, if IoU between the GT and prediction is greater than the IoU threshold (e.g., figure 3.11 A);

2. False positive, if the IoU between GT and prediction is less than the IoU threshold or if there is a prediction for no associated GT (e.g. figure 3.11 image C and G respectively);

3. False Negative, if the GT has no associated prediction (e.g. figure 3.11 image D);

4. True Negative, if the frame has no GT and prediction.

Therefore, if IoU threshold is, say 0.5, and the IoU value for the prediction is 0.7, then we classify the prediction as True Positive. On the other hand, if IoU is 0.3, we classify it as a False Positive.

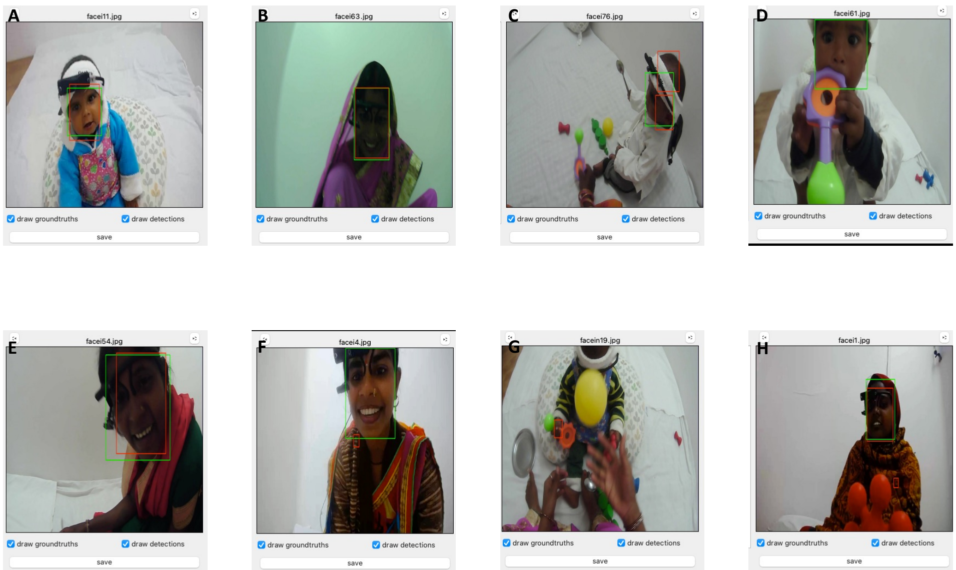
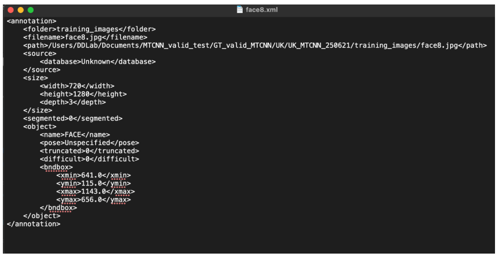
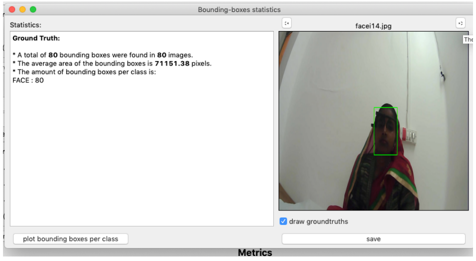


Figure 3.11: An example of detecting faces in the validation data set and calculating IoU in the India sample. Ground Truth is depicted in green and the predicted bounding box is depicted in red. The closer the predicting bounding box to the ground truth bounding box, the larger the IoU.

To use the evaluation metric, we convert the ROI images to a .csv file with image path, image name, size/coordinates of the bounding box and label name (in this case, face) for the GT data. We, then, created .xml files from the .csv files for the GT data and loaded the .xml files and raw images into the evaluation metrics software. We then checked the GT statistics to examine if any labels were missing or incorrectly labelled. Figure 3.12 shows visual example of this step.



(a) Example of .xml file



(b) Sanity check for manually annotated ground truth labels.

Figure 3.12: Example of steps that lead into the evaluation metrics. Figure (a) shows a .xml file output consisting of image path, name, coordinates of bounding boxes and label name. Figure (b) shows a visualisation process of the metrics that enables us to check the ground truth for the missing label.

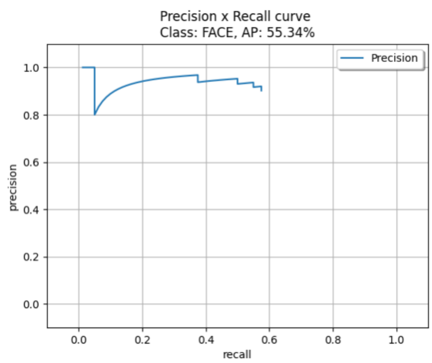
We ran two evaluation metrics for each cohort (India, UK). One evaluation included MTCNN predictions with CLAHE and another without CLAHE. Within each output, we extracted results using an IoU threshold of 0.50. Figure 3.13 shows the evaluation metric output for the UK cohort. Precision, here, is the total number of true positives divided by the total number of true positives and false positives, in other words, it is the correctly identified faces out of all the identifications. Similarly, the recall is the ratio of true positives and total GT positives, i.e., true positives divided by the sum of true positives and false negatives. The trade-off between precision and recall performance can be manipulated by adjusting the IoU threshold such that with a less restrictive IoU threshold, higher recall values can be obtained with the highest precision. For instance in figure 3.13 (a), the detector can retrieve about 55.34% of the total ground truths without any miss detection. Similarly, in figure 3.13 (b), the detector can retrieve 62% of total ground truths without any miss detection. AP in the figures denotes average precision. AP is a single value obtained by summarising the precision-recall curve, i.e., precision averaged across all unique recalls. Precision indicates how precise our model is for each class using the equation -

$$Precision = \frac{TP}{TP + FP}$$

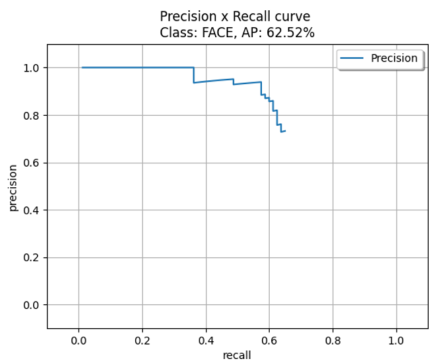
where TP denotes the True positive and FP is the False positive. Recall, here, indicates how good the model is at recalling the class (that is, Face) from all the inputted face images. In other words, out of all the face images in the dataset, how many faces was the model able to detect? See below equation for calculation of Recall -

$$Recall = \frac{TP}{TP + FN}$$

where TP denotes the True positive and FN is the False negative. A precision-recall curve is used to visualise the trade-off between the precision and recall for different thresholds.



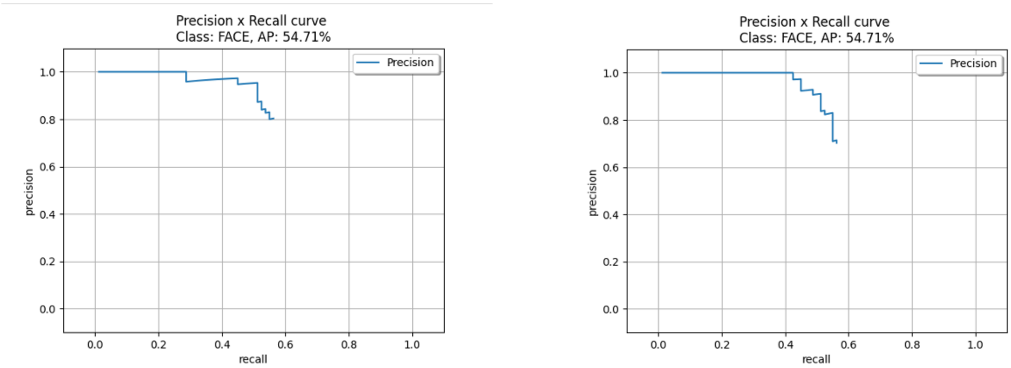
(a) Evaluation metric without CLAHE



(b) Evaluation metric with CLAHE.

Figure 3.13: Evaluation metric output for the UK at IoU 0.50. AP, average precision, i.e., precision averaged across all unique recalls. The blue line denotes the Precision-Recall curve.

We ran the same evaluation metric for the India cohort. Figure 3.14 shows that for both outputs with (b) and without (a) CLAHE, the detector can retrieve 54.71% of total ground truths without any miss detection. Based on these evaluation metrics, we decided to use CLAHE for both India and UK cohorts.



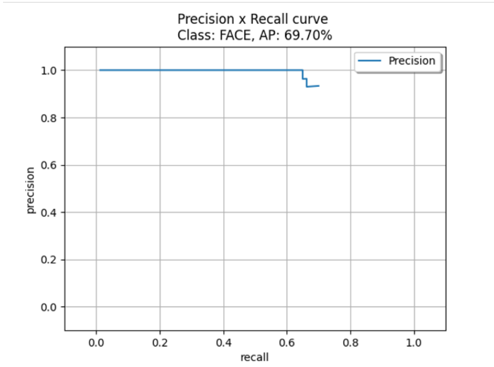
(a) Evaluation metric without CLAHE

(b) Evaluation metric with CLAHE.

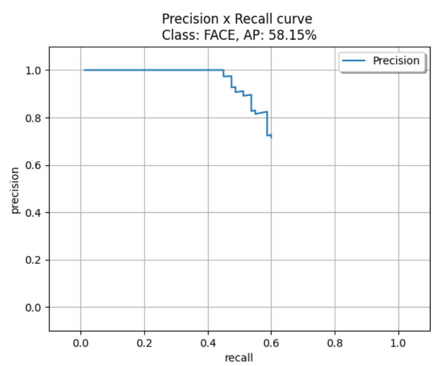
Figure 3.14: Evaluation metric output for India at IoU 0.50. AP, average precision, i.e., precision averaged across all unique recalls. The blue line denotes the Precision-Recall curve.

To increase the precision, we need to decrease the false positives. To assess potential filters needed to decrease the number of false positives, we ran a dyadic video from each cohort through the MTCNN algorithm and visualised the predictions. Visualisations were then converted into a video format, at 90fps, for a qualitative check. To do this, we manually coded the videos using an open-source, video coding software called BORIS – Behavioral Observation Research Interactive Software (Friard & Gamba, 2016). BORIS allows us to create a list of observations, add comments for observation, choose between state (with duration) and point (no duration) events, extract plots and time budgets of the events and has a coding pad that allows coders to start/stop the event in real-time (see http://boris.readthedocs.io for full documentation). After making a qualitative pass on the chosen dyadic video, we noted potential filtering steps needed to improve accuracy.

The first filtering step required manipulating the size and confidence of bounding boxes. Therefore, we filtered the data through different sizes of bounding boxes and confidence thresholds (ranging from 60-80%) of bounding boxes until we got the best precision. The best bounding box size and confidence threshold that suited both cohorts were 5 and 70% respectively at an IOU threshold of 0.30. Figure 3.15 shows that the average precision for the UK cohort increased by 7.18% and for India increased by 3.45%.



(a) Evaluation metric for UK



(b) Evaluation metric for India

Figure 3.15: Evaluation metric with CLAHE at IoU threshold 0.30, with the minimum size of the bounding box 5 and confidence threshold set at 70%. AP, average precision, i.e., precision averaged across all unique recalls. The blue line denotes the Precision-Recall curve.

As a point of reference, prior research by Long et al. (2022) has shown a precision of .94 and recall of .62 when using MTCNN with a random sample (c.f., figure 3.15). Although precision and recall were comparable, if not superior, for our UK dataset, the evaluation metrics for the India sample were lower. To increase the accuracy further, we incorporated four additional filtering steps. The final filtering steps consisted of:

1. Filter by confidence. Any detection with the confidence of <70% was discarded;

2. Filter by minimum size. Any bounding box with size <5% of the image was discarded;

3. Filter by asymmetry. If an edge of the bounding box was five times longer than the adjacent side, then it was discarded;
4. Removal of duplicates. If there were two bounding boxes for the same item (e.g. parent face) on the same frame, then this filtering step removed the duplicate. Removal was prioritised primarily based on the size of the bounding box (i.e. the largest bounding box was kept) and secondarily based on the confidence (i.e. if two bounding boxes were the same size, we kept the bounding box with the highest confidence value).

3.2.2 You Only Look Once for Object Recognition

You Only Look Once (YOLO) is a state-of-the-art object detection algorithm developed by Redmon et al. (2016). It uses Convolutional Neural Networks (CNNs) to detect objects with better than real-time performance, typically processing in excess of 40 frames per second. The advantage of using YOLO over other algorithms, such as a Fast Regional Convolutional Neural Network (FRCNN), is the highly efficient method it employs to detect objects. While other approaches to object detection go through each image hundreds of times to detect objects, YOLO runs through each image only once to return a complete list of detected objects; hence its name, You Only Look Once. YOLO's speed comes with a slight trade-off in terms of the accuracy of detection that it provides (69% mean average precision for YOLO versus 70% for FRCNN). However, given the speed of detection, it is the preferred approach for a number of real-life applications where real-time processing is required. For example, YOLO has been used to identify cars, people and traffic signals on busy roads for autonomous driving applications (Boukerche & Hou, 2022; Masmoudi, Friji, Ghazzai, & Massoud, 2021; Pouyanfar et al., 2019). Hence, YOLO can be a useful method for detecting objects.

Here, we use YOLOv5, freely available at https://github.com/ultralytics/yolov5.

The network uses the following steps (see figures 3.16, 3.17, 3.18, 3.19 and 3.20). First, the YOLO algorithm overlays a grid on the chosen image (see figure 3.16) and makes a prediction for each cell in the grid (e.g., figure 3.17 A). Based on the predictions across cells, it sets bounding boxes and confidence values (see figure 3.17 B). Predictions, bounding boxes, and confidence values are set for all cells, regardless of whether there are any salient objects or specified targets within them (see figure 3.18). Logically, these will tend to have lower confidence values. Then, YOLO expands the bounding boxes by an amount proportional to the confidence of each prediction (figure 3.17 B versus figure 3.18 B). The process, therefore, allows a threshold parameter to be specified which defines the minimum probability that a bounding box must reach in order to be said to contain an object. As a result, YOLO creates a map with multiple bounding boxes ranked by their confidence value (see figure 3.19 A), which serves to identify where the objects are located in the image.

In order to detect what the objects are, the algorithm predicts the conditional class probability for each grid cell. That is, it predicts the probability that the image contains an object of a given class (e.g., plant) at each specific location. Predictions are conditional at this stage because they do not specify the presence of an object. Instead, they set the condition that ”if” there is an object within the cell, that object will be of the given class. Figure 3.19 B, depicts a coarse segmentation map, with the conditional class predictions for each cell. Within this map, the condition would be, for example, that if there is an object located within the red section, then it will be a plant.

Once YOLO has set the location of bounding boxes and the probable classes,

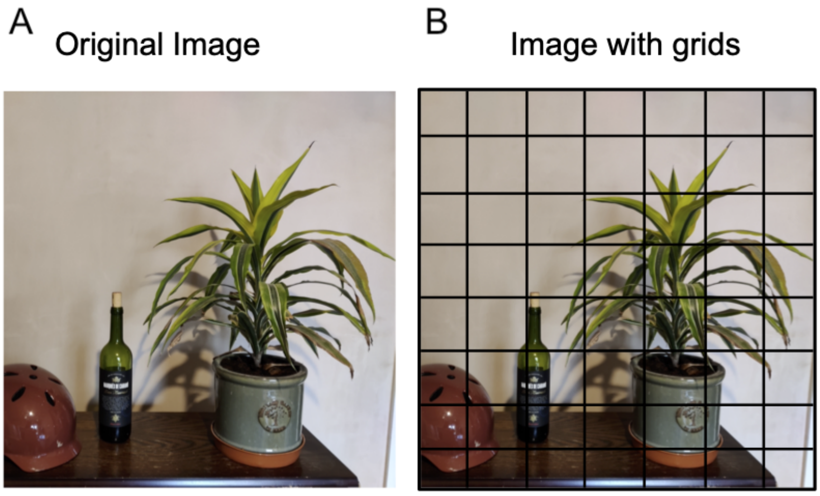


Figure 3.16: Image A depicts the raw image containing three classes of objects, (left to right) helmet, bottle and plant. YOLO divides the raw image into grids as seen in Image B.

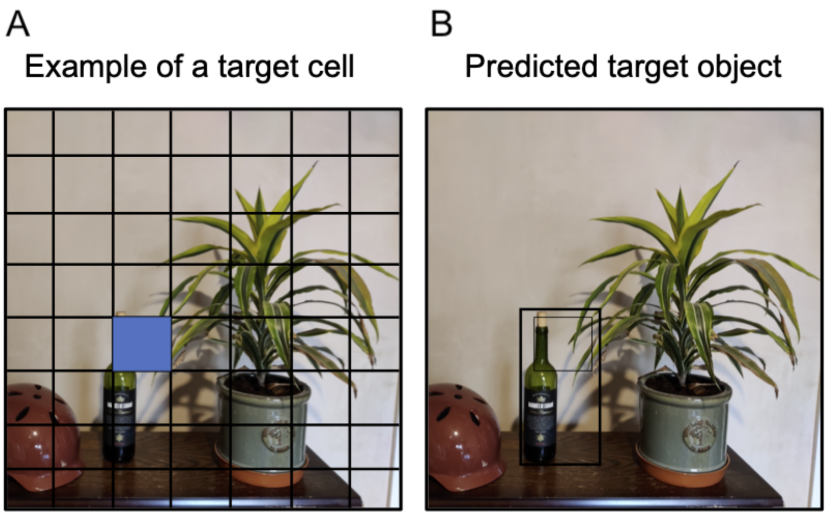


Figure 3.17: Image A shows an example of each grid predicting a bounding box and image B shows an example of each grid predicting the confidence value of the bounding box. In Image B, the thick bounding box depicts a higher confidence value than the thinner bounding box.

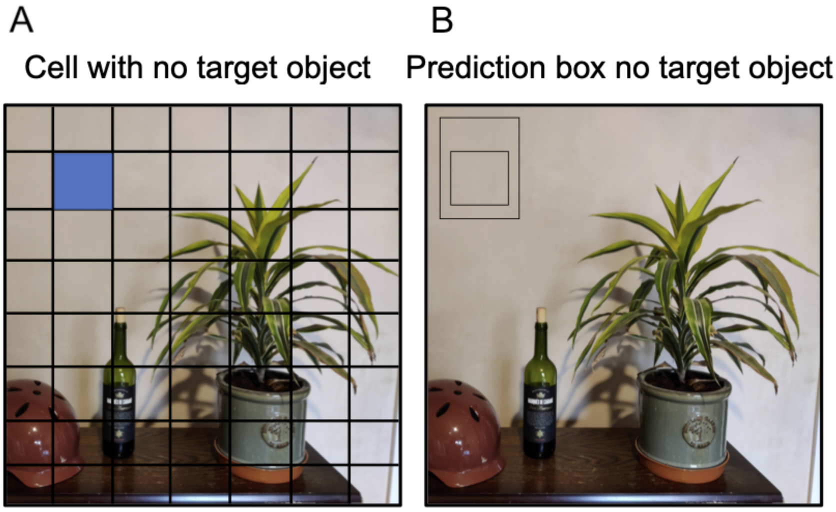


Figure 3.18: Image A shows a prediction cell with no target image. Image B shows the predicted bounding box and confidence for a no-target bounding box. Note, that the thin bounding boxes depict a lower confidence value of prediction.

the information is merged. First, the algorithm multiplies the conditional probability with the objects' confidence value. This results in bounding boxes weighted by their actual probabilities of containing the desired object(s) (fig 3.6 a). Then, it discards predictions with lower confidence values, thus reducing redundancy and narrowing down the location of the target object with the greatest probability of containing it (figure 3.20 B).

We trained separate YOLO models for the UK and India dyads. To train the models for the 11 UK toys and 10 India toys (and an additional class for a mobile phone), we manually annotated a small subset of data to be used as training and validation data. For the UK dyads, we labelled 1295 frames, sampled from 28 UK dyads, and for the India dyads, we labelled 1238 frames sampled from 113 dyads. The frames were extracted from both the parent and infants' head-mounted cameras. The initial selection of frames was random, however, we then inspected the selection and added additional frames

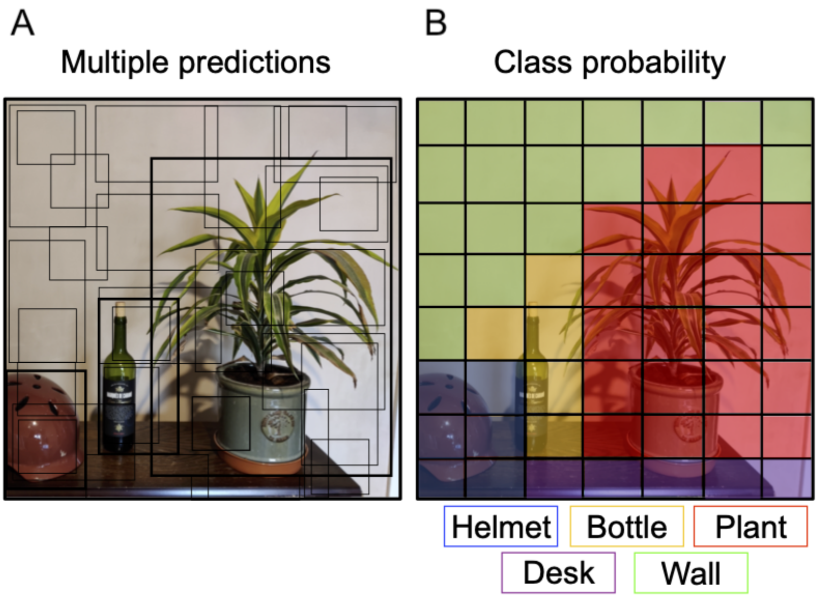


Figure 3.19: Image A visualises multiple predictions in the raw image. Predictions with lower confidence values are depicted in thin bounding boxes and predictions with higher confidence values are depicted in thick bounding boxes. Image B depicts an example of a coarse segmentation map of class probability. Blue cells depict the areas with a conditional probability for the class helmet. Yellow cells identify the class bottle. Red cells depict the class plant.

to make sure that the data contained a good variety of different scales and orientations of each toy (see figure 3.21 for an example).

All frames were labelled by trained research assistants using a software package called LabelImg (Tzutalin, 2015). LabelImg is an open-source software package written in Python. It is used to annotate images for the purpose of providing data for training and validating object detection software. We used it to create training and validation data sets for training YOLO (see figure 3.22 for an example). The annotations were saved in a text format, i.e., one text file per image, with each line containing the numeric representation of the class label for each object in the image. Subsequent columns identified the location and size of each object's bounding box. From each cohort's training data set, we extracted 128 (UK) and 117 (India) images to create a val-

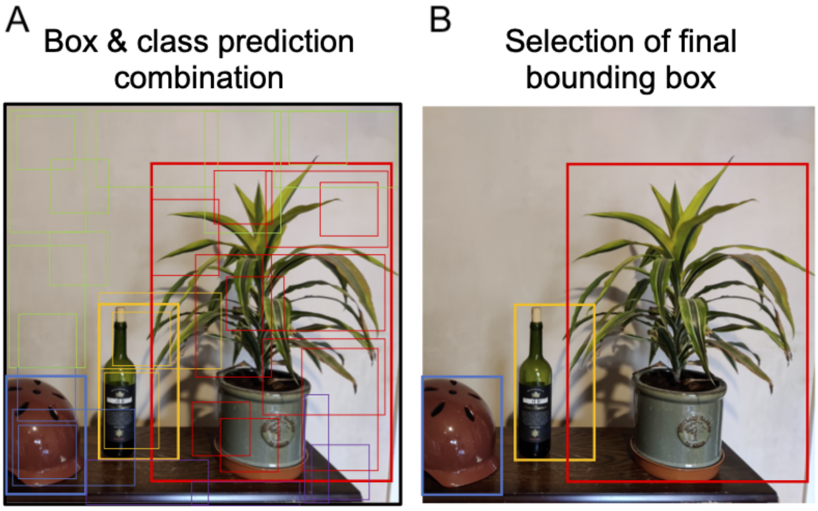


Figure 3.20: Image A depicts a combination of the bounding box, confidence and class predictions. Image B depicts the final output after discarding predictions with low confidence.

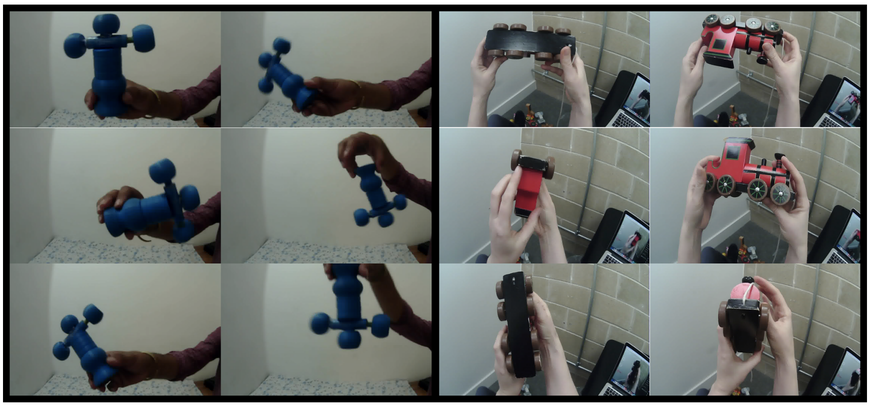


Figure 3.21: Example of different orientations of a toy from India (left) and UK (right).

idation dataset to assess the YOLO model's performance, leaving us with 1167 images to train the UK model and 1120 to train the India model. The accuracy of the models was assessed using the precision (of the toys detected, how accurate were the detections), recall (how many toys were successfully detected out of the total number of toys) and mAP (mean average precision; the mean of average precision for all classes). All models were run at a threshold of 0.50

for Intersection Over Union (IoU), which means that the detected bounding boxes had to overlap with the labelled boxes by at least 50% in order to count a successful detection.

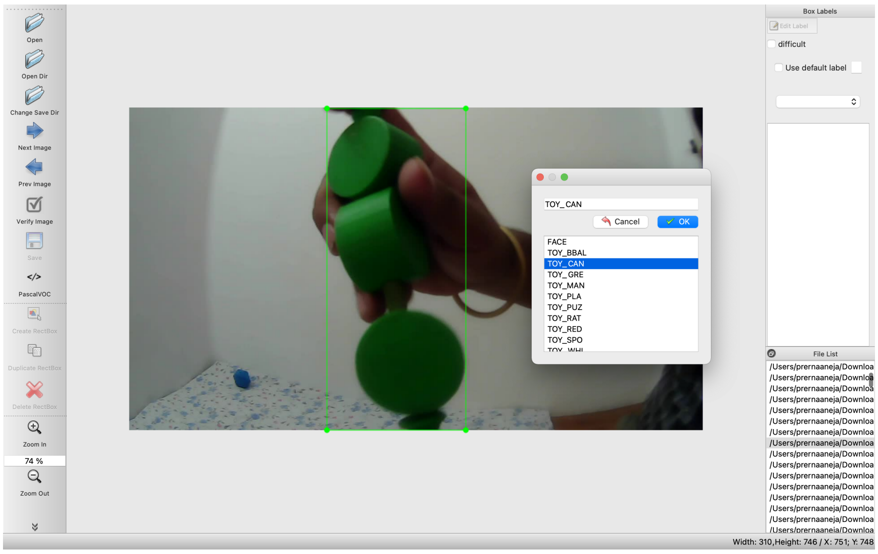


Figure 3.22: Example of labelling toys using LabelImg annotation tool with a minimal enclosing bounding box.

The training, validation and testing of the YOLO algorithm were undertaken at the University of East Anglia's High-Performance Computing Cluster (HPC). The HPC allows for the processing of multiple dyads in parallel, rather than sequentially, as is more often the case on a conventional computer. The image data for training, validation and testing were uploaded to the HPC and organised into directories of the same name. The training and validation directories contained a sub-directory named "labels", in which the annotations from the LabelImg software were stored. A configuration file (a yaml file) was populated with the directory locations of the training and validation directories.

The version of YOLO used in our work was YOLOv5 (Jocher et al., 2022). The source code for this toolbox was downloaded from GitHub onto the HPC and a Python environment was created to enable the software to run. The

YOLO toolbox contains separate Python scripts for training and testing. The model training used by YOLO is a form of transfer learning, meaning that a pre-trained model is first loaded, and then the weights in that model are fine-tuned for the specific task at hand. The pre-trained model will already have been trained using many hundreds of thousands of images, drastically reducing the amount of new data required to adapt the model. The metric used to monitor the training performance is called loss. Loss is the difference between the model input and the input reconstructed by first passing the data through and then backward through the network. If a network is training properly, loss is expected to fall as the training progresses and the network becomes better able to represent the data passed through it.

Training is conducted in a series of steps called epochs. In each epoch, small chunks of data (batches) are passed through the network and the weights at each node of the network are tuned. We chose to train the model to 250 epochs and used the default batch size of 12 images. YOLO uses an early stopping criteria causing the training to stop automatically if the loss stops reducing significantly between epochs. The training process produces two models: The model produced by the final epoch and the best model, as identified by the model with the lowest loss. We use the best model for processing our validation and test data.

An HPC job submission script was created to call the YOLO training script with the parameters required for training. These parameters included: the location of the yaml configuration file, the location of initial model weights for YOLO, the number of epochs to train the model for, and the size of the batch. YOLO automatically processes the validation data through the network at each epoch, providing the mAP and other performance metrics. Note that the validation data is used only for the purpose of inference (testing) and does not influence the model training itself. Finally, the validation data is also passed through the best model generated during training.

Once the best YOLO model has been generated, it is ready to apply to the dyads that were set aside for testing. A bash script was written which generates and submits separate job submissions scripts for each parent and child, in each dyad. Each submission script calls YOLO's detection script, which is the script used for testing. The testing script requires several parameters: The path of the best model generated during training, the directory containing a dyad's images, and the selection of the type of data output. In this case, we save the detected toys in a YOLO label file for each dyad, which is the same format as those annotated for the training data.

YOLO model evaluation UK Table 3.1 shows the metric output from model 1 consisting of 1167 images in the training data set and 128 images with 352 labels in the validation data set. An mAP of 0.832 was achieved with a precision of 89% and a recall of 77.9% across all toys. All toys with the exception of the camera (TOY_CAM), elephant (TOY_ELE) and giraffe (TOY_GIR) had an mAP of over 80%.

Next, we incrementally increased the number of frames for the poorly performing toys to see if this improved the accuracy. We started by adding 20 toy elephant labels to the training dataset. Table 3.2 shows an overall improved performance across toys with an mAP of 0.858 with a precision of 91.7% and a recall of 77.7%. Model 2 showed a slightly increased accuracy for toy elephant as well as for camera and giraffe.

On looking through the validation data set, we noted some missing as well as mislabelled data. Fixes included relabelling as well as creating new labels for missing toys. Therefore, Model 3 consisted of the same number of training

Table 3.1: Performance of object detector at IoU Threshold 0.50 with 1167 images in the training dataset, 128 images in validation data set and 352 labels in the validation dataset.

Model 1				
Class	Label	Precision	Recall	mAP@.5
All	352	0.89	0.779	0.832
TOY_APP	24	0.875	0.792	0.873
TOY_BUT	30	0.993	0.867	0.893
TOY_CAM	26	0.914	0.692	0.75
TOY_CUP	35	0.921	0.8	0.865
TOY_ELE	32	0.872	0.639	0.733
TOY_GIR	37	0.673	0.568	0.609
TOY_DUC	34	0.97	0.941	0.974
TOY_GRE	27	1	0.865	0.893
TOY_KET	41	0.967	0.722	0.831
TOY_RAT	29	0.753	0.897	0.861
TOY_TRA	37	0.853	0.784	0.875

Note. APP, Apple; BUT, Butterfly; CAM, Camera; CUP, Cup, ELE, Elephant, GIR, Giraffe, DUC, Duck; GRE, Green; Ket, Kettle; RAT, Rattle; TRA, Train.

Table 3.2: Performance of object detector at IoU Threshold 0.50 with 1187 images in the training dataset, 128 images in validation data set and 352 labels in the validation dataset.

Model 2				
Class	label	Precision	Recall	mAP@.5
all	352	0.917	0.777	0.858
TOY_APP	24	0.928	0.75	0.902
TOY_BUT	30	1	0.763	0.845
TOY_CAM	26	0.952	0.769	0.852
TOY_CUP	35	0.958	0.829	0.908
TOY_ELE	32	0.898	0.688	0.745
TOY_GIR	37	0.793	0.519	0.628
TOY_DUC	34	0.91	0.971	0.976
TOY_GRE	27	1	0.81	0.891
TOY_KET	41	0.891	0.805	0.881
TOY_RAT	29	0.821	0.897	0.933
TOY_TRA	37	0.932	0.746	0.877

Note. APP, Apple; BUT, Butterfly; CAM, Camera; CUP, Cup, ELE, Elephant, GIR, Giraffe, DUC, Duck; GRE, Green; Ket, Kettle; RAT, Rattle; TRA, Train.

Table 3.3: Performance of object detector at IoU Threshold 0.50 with 1167 images in the training dataset, 128 images in validation data set and 364 labels in the validation dataset.

Model 3				
Class	label	Precision	Recall	mAP@.5
All	364	0.922	0.797	0.868
TOY_APP	24	0.912	0.866	0.94
TOY_BUT	35	1	0.682	0.748
TOY_CAM	27	1	0.784	0.862
TOY_CUP	35	0.909	0.857	0.915
TOY_ELE	32	0.954	0.719	0.778
TOY_GIR	37	0.874	0.676	0.728
TOY_DUC	35	0.969	0.893	0.939
TOY_GRE	27	0.891	0.889	0.922
TOY_KET	45	0.929	0.756	0.908
TOY_RAT	31	0.871	0.871	0.944
TOY_TRA	36	0.836	0.778	0.866

Note. APP, Apple; BUT, Butterfly; CAM, Camera; CUP, Cup, ELE, Elephant, GIR, Giraffe, DUC, Duck; GRE, Green; Ket, Kettle; RAT, Rattle; TRA, Train.

images as model 1 (i.e. 1167 images) and increased labels in the validation data set (from 352 to 364). Table 3.3 shows an improved accuracy for object detection with an overall mAP of 0.868 across all toys with 91.7% precision and 77.7% recall.

Our final model included updated training data set with 60 new labels (20 elephants, 20 giraffes and 20 cameras) as well as the updated validation set with 364 labels. As shown in table 3.4, the overall mAP across toys was 0.856 with 96.2% precision and 75.2% recall.

YOLO model evaluation India India models were generated after the UK models, hence we looked through the data prior to running models to ensure that there were no missing or mislabelled labels. One of the toys in India broke, and parent and their infants started using them as novel toys so we decided not to change them. Therefore, during labelling for the India training dataset, we labelled the TOY_GLO as a whole and its dismantled parts as TOY_RED and TOY_MIX.

Table 3.4: Performance of object detector at IoU Threshold 0.50 with 1227 images in the training dataset, 128 images in the validation data set and 364 labels in the validation dataset.

Model 4				
Class	label	Precision	Recall	mAP@.5
all	364	0.962	0.752	0.856
TOY_APP	24	1	0.762	0.841
TOY_BUT	35	1	0.667	0.748
TOY_CAM	27	1	0.739	0.841
TOY_CUP	35	1	0.785	0.933
TOY_ELE	32	0.999	0.688	0.789
TOY_GIR	37	0.797	0.568	0.684
TOY_DUC	35	1	0.841	0.925
TOY_GRE	27	0.993	0.815	0.92
TOY_KET	45	1	0.721	0.913
TOY_RAT	31	0.89	0.871	0.927
TOY_TRA	36	0.908	0.82	0.9

Note. APP, Apple; BUT, Butterfly; CAM, Camera; CUP, Cup, ELE, Elephant, GIR, Giraffe, DUC, Duck; GRE, Green; Ket, Kettle; RAT, Rattle; TRA, Train.

Table 3.5 shows evaluation metric output from the first India YOLO model. The overall mAP across toys was 0.839, with 84.1% precision and 81.4% recall. However, labels such as TOY_RED, TOY_MIX and TOY_GLO showed a relatively poor performance. Therefore, to enhance the results for these labels, we added CLHAE image processing to our training data set (previously used in MTCNN).

As can be seen from table 3.6 adding CLHAE did not increase the overall performance of the detector. However, it did increase the accuracy of performance on TOY_MIX and TOY_GLO but not for TOY_RED. For our next model, we decided to combine labels of TOY_GLO, TOY_MIX and TOY_RED into one label, namely, TOY_GLO given that they were all dismantled parts of the same toy.

Model 3 showed an improved accuracy for object detection with an overall mAP of 0.898 across all toys with a 93.5% precision and 83.3% recall. As can be seen in table 3.7, the performance for TOY_GLO increased by merging the toys under one label.

Table 3.5: Performance of object detector at IoU Threshold 0.50 with 1120 images in the training dataset, 117 images in the validation data set and 364 labels in the validation dataset.

Model 1				
Class	Label	Precision	Recall	mAP@.5
All	324	0.841	0.814	0.839
TOY_BBAL	21	0.93	1	0.978
TOY_CAN	38	0.913	0.895	0.933
TOY_GRE	5	0.826	0.8	0.803
TOY_MAN	46	0.872	0.913	0.938
TOY_PLA	49	0.95	0.816	0.876
TOY_PUZ	4	0.848	1	0.995
TOY_RAT	6	1	1	0.995
TOY_RED	12	0.623	0.667	0.684
TOY_SPO	49	0.904	0.673	0.767
TOY_YEL	8	0.9	1	0.995
TOY_MIX	17	0.345	0.882	0.585
TOY_GLO	42	0.965	0.31	0.599
MOBILE	27	0.852	0.63	0.763

Note. BBAL, Blue Ball; CAN, Candy; GRE, Green; MAN, Man; PLA, Plate; PUZ, Puzzle; RAT, Rattle; RED, Red; SPO, Spoon; YEL, Yellow Ball; MIX, Mix; GLO, Glow.

Table 3.6: Performance of object detector at IoU Threshold 0.50 with 1120 images in the training dataset, 117 images in the validation data set, 364 labels in the validation dataset and CLAHE histogram equalizer.

Model 2				
Class	Label	Precision	Recall	mAP@.5
All	324	0.853	0.972	0.816
TOY_BBAL	21	0.939	1	0.963
TOY_CAN	38	0.961	0.895	0.943
TOY_GRE	5	0.927	0.8	0.8
TOY_MAN	46	0.909	0.87	0.9
TOY_PLA	49	0.073	0.738	0.802
TOY_PUZ	4	0.677	1	0.995
TOY_RAT	6	0.854	1	0.972
TOY_RED	12	0.694	0.567	0.582
TOY_SPO	49	0.933	0.633	0.707
TOY_YEL	8	0.959	1	0.995
TOY_MIX	17	0.417	0.882	0.591
TOY_GLO	42	1	0.248	0.631
MOBILE	27	0.846	0.667	0.722

Note. BBAL, Blue Ball; CAN, Candy; GRE, Green; MAN, Man; PLA, Plate; PUZ, Puzzle; RAT, Rattle; RED, Red; SPO, Spoon; YEL, Yellow Ball; MIX, Mix; GLO, Glow.

Table 3.7: Performance of object detector at IoU Threshold 0.50. The training dataset included 1120 images. The validation dataset included 117 images and3 64 labels. CLAHE histogram equalizer was added to the model. TOY_-RED, TOY_MIX and TOY_GLO were combined into TOY_GLO.

Model 3				
Class	Label	Precision	Recall	mAP@.5
All	324	0.935	0.833	0.898
TOY_BBAL	21	0.942	1	0.95
TOY_CAN	38	0.969	0.835	0.943
TOY_GRE	5	0.927	0.8	0.8
TOY_MAN	46	0.918	0.87	0.91
TOY_PLA	49	0.949	0.767	0.81
TOY_PUZ	4	0.839	1	0.995
TOY_RAT	6	1	0.729	0.995
TOY_SPO	49	0.975	0.633	0.747
TOY_YEL	8	0.96	1	0.995
TOY_GLO	71	1	0.859	0.925
MOBILE	27	0.804	0.667	0.803

Note. BBAL, Blue Ball; CAN, Candy; GRE, Green; MAN, Man; PLA, Plate; PUZ, Puzzle; RAT, Rattle; RED, Red; SPO, Spoon; YEL, Yellow Ball; MIX, Mix; GLO, Glow.

3.3 Overview of the Pipeline

Now that the CNNs for faces and objects were trained and validated, we moved on to create a full processing pipeline for the data set. Figure 3.23 shows the steps to be followed using this pipeline. The first step includes processing gaze data from the eye-trackers and annotating dyadic videos to identify the synchronised onset and any infant crying for more than 1 minute. Next, we discuss two processes that we ran in parallel 1) extracting synchronised frames from dyadic videos, and 2) optimising the use of object recognition (YOLO) and face detection (MTCNN) by filtering the resultant data. Following this, we discuss event detection for faces and objects, leading to analyses of joint attention using an existing analysis toolkit.

Figure 3.23: A flow diagram demonstrating the steps of the pipeline.

3.3.1 Eye-tracker

We first pre-processed the eye-tracking recordings by mapping gaze points. The pupil player software comes with several visualisation plugins that can be added to the video processing. We used the Vis Circle plugin to visualise gaze position for caregivers' eye-tracker (see Appendix K.1). While the pupil player automatically loads pupil positions that were detected and stored during pupil capture recording, it also allows for post-hoc pupil detection. We used post-hoc pupil detection for the following reasons: 1) to ensure that the calibrated gaze was mapped onto the target circle on the calibration marker and hence validate the calibration, and 2) to enhance accuracy by trimming any dark eyelashes. This can be done using a built-in algorithm where the Region of Interest (ROI) could be adjusted closer to the edges of the eye (Appendix K.2). We noted that adjusting the ROI could be particularly useful for the India cohort that had dark, long eyelashes or for caregivers wearing mascara. If the number of calibrations dismissed by the software was over 40%, the data were excluded. After post-hoc calibration, the software generates the gaze position (visualised using Vis Circle) that maps onto the caregivers' world camera view. This was verified by visually checking the data to ensure that the gaze dot mapped onto the calibration marker. Figure 3.24 shows frames from a processed video from a dyad in India, with the parent gaze mapped onto the world camera. We, then, export the following: 1) the play session video with gaze position overlayed on the video, 2) a .csv file consisting of the pupil and gaze coordinates and confidence, and 3) timestamps in NumPy format (for complete documentation, see Pupil Player docs).

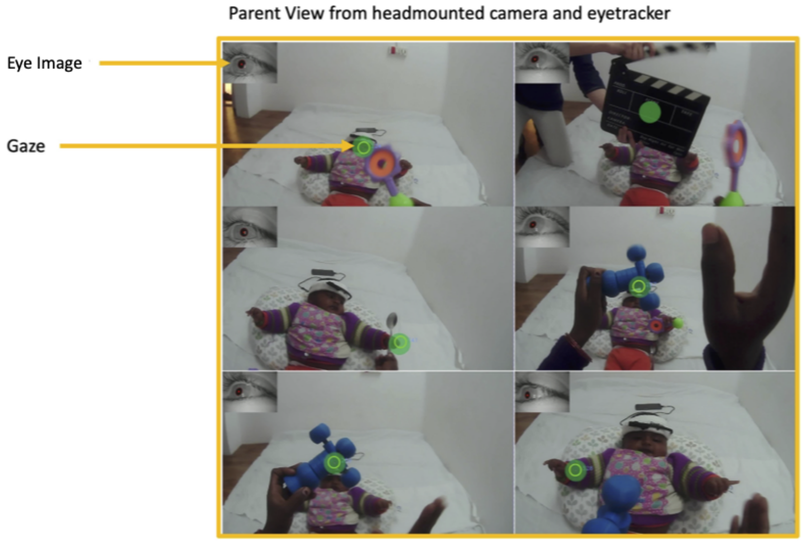


Figure 3.24: Processed dyadic video from India. Here, we extracted a few frames to depict the output after processing the parent eye-tracker and head-camera through Pupil Player. The image of the eye in the top left corner of each frame is captured via the eye-tracker. The green dot depicts the parents' gaze. The yellow circle within the green dot depicts fixation.

3.3.2 Annotation

We used the annotation plugin in pupil player software to synchronise the parent and infant head camera videos. The Clapper board is used as a reference to annotate the synchronised onset of the play session for both parent and infants' head cameras. The offset was synchronised either at the end of the ten-minute play session or when the session ended (in case of infant fussiness). Any crying event for more than a minute was annotated to remove from the processing during the frame extraction phase. The crying event was mainly from the parents' head camera view as they are more likely to see the infants' faces. On exporting the annotation, we get a .csv with world or scene camera index, timestamp (NumPy file), the label of annotation and duration of the event.

3.3.3 Frame Extraction

We organised folders for each participant that consisted of parent and infant raw .mp4 video, its corresponding timestamp (i.e. NumPy file), annotation.csv, and gaze position file for parents only. The synchronised frames are extracted in MATLAB using FFmpeg (frame extraction code available on GitHub). The code compensates for any fluctuations in time sampling. It then uses the parent or infants' annotations as a referent signal, whichever had fewer frames. In the case of annotations related to infant crying, it filters any crying epoch for more than a minute.

The matlab code, then, downsamples the other stream using the timestamp information to pick the frame that best matches the timing of each frame in the reference signal. Once the frames are synchronised, it renames the files such that each synchronised frame for the parent and the infant has the same frame number (e.g. ID_Parent_00001 and ID_Child_00001). Following this, the code finds the gaze timestamps (from the eye-tracker) that match caregivers' synchronised frames (from the head camera). This gives us a new gaze data file with normalised positions X and Y for each gaze point with its corresponding frame for the parents. This step yields an equal number of synchronised frames for parents and the infants' head cameras, along with parents' gaze data that matched the extracted frames (new gaze position file) for each dyad.

3.3.4 Machine learning

We submitted both MTCNN and YOLO script jobs on the HPC. MTCNN included two additional filtering steps to further improve the data quality:

1. Fill length. When an item was detected for at least 3 frames (e.g. bound-

ing box on parent's face), then detection ceased for a frame, and the item was detected again for 3 frames, the filtering step filled the missing frame. This filter served to reduce the number of false rejections (misses);

2. Minimum length of sequential frames with detection. When the identification of items lasted for a single frame, the filtering step discarded the detection. This served to minimise false positives.

For YOLO, we used the weights from Model 4 for the UK cohort (see table 3.4) and Model 3 for the India cohort (see table 3.7) as our final models for object recognition.

3.3.5 Visualisation and BORIS

To verify the MTCNN face detection and filtering approach and to check the accuracy of YOLO object recognition, we processed data for four dyads from the UK and eight dyads from India for manual inspection using BORIS. For this purpose, we visualised the MTCNN and YOLO predictions by converting the synchronised frames for each dyad into a video format at 90fps. Each video consisted of the predictions visualised using green bounding boxes for toys and faces with the predicted accuracy (in %) and a blue gaze position with a blue extended bounding box around it. The gaze position for parents' eye-tracker is based on the gaze output from the pupil player. We created the gaze position for the infants' headcamera at the centre of the frame. Both gaze points included an extended bounding box. The gaze box was determined by the approximate central vision field of view (FOV), the FOV of the camera, and the camera sensor resolution. The camera FOV for pupil labs is 60 degrees and we assumed that the central human FOV is approximately 15 degrees.

We, then, converted the raw gaze data obtained from pupil player processing and converted it to pixel coordinates for the central blue dot. The height and width of the extended gaze box were calculated from the aforementioned FOV values. The top left corner of the extended gaze box was generated by halving the gaze width and height and subtracting that from the gaze x, and y coordinates.

Figure 3.25 shows example frames from a dyad from the UK (images A and B) and India (images C and D). Trained experimenters coded hit for correct detection, miss for incorrect rejection, false-positive for incorrect detection, and true negative for correct rejection for each video. For instance, in figure 3.25 A, detection of a face on the infants' onesie (label 0 with prediction confidence 79.3%) would be coded as a false positive, missed detection of toys puzzle and glow in image C would be labelled as a miss, and all other detection in all four images would be labelled as a hit. For, YOLO, each toy was coded twice, once from a parent's headcamera and once from the infant's headcamera.

3.3.6 Event Detection

After validating both MTCNN and YOLO detection through manual coding in BORIS, we created event detection files for each prediction (MTCNN, YOLO) for each member of the dyad (parent, infant). The event detection script reads in the gaze report and predictions and creates a new csv file that informs us of what object each partner is looking at. The event csv file outputs three columns: 1) the onset frame of each label, 2) the offset frame of that label, and 3) the corresponding label for that duration of the frame (i.e., the object that the parent/infant is looking at). An event was defined as a continuous series of 3 or more frames (or 99msec) looking at the same label (toy, face). In case there is more than one object in the dyad's view (e.g. figure 3.25 A

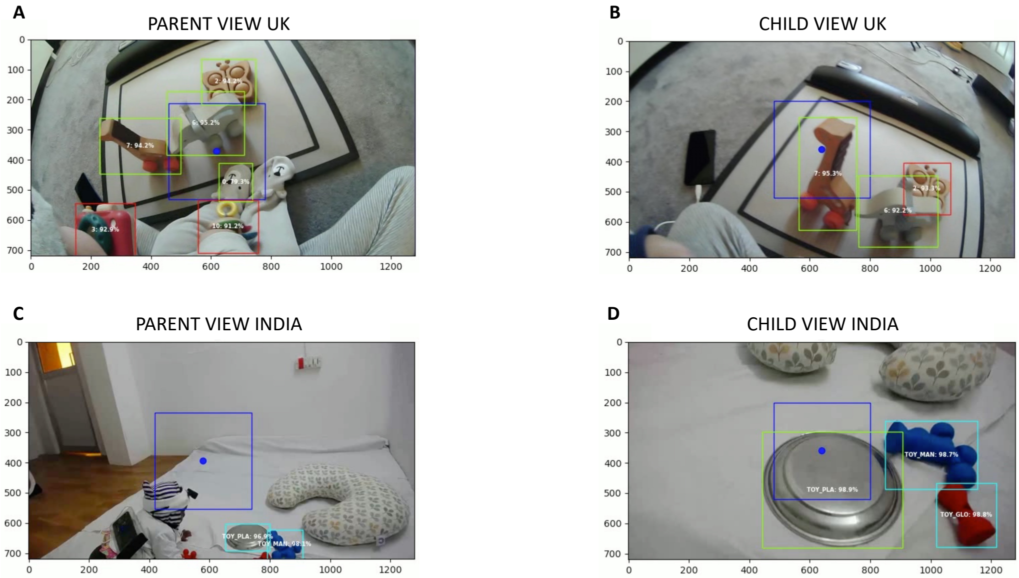


Figure 3.25: Example frames from machine learning visualisation output for a dyad from the UK (image A and B) and India (image C and D). Images A and B depict a frame from the parent's and infant's view, respectively. The blue bounding box with a circle depicts the gaze. The red bounding boxes depict toys predicted by the algorithm that does not overlap with the parent/infant's gaze. Green bounding boxes depict toys and faces predicted by the algorithms that overlap the parent/infant's gaze. Images C and D depict a frame from the parent's and infant's view, respectively from an Indian dyad. A blue bounding box with a circle depicts the gaze. The cyan bounding boxes depict toys predicted by the algorithm that does not overlap with the parent/infant's gaze. Green bounding boxes depict toys and faces predicted by the algorithms that overlap the parent's gaze. Each predicted bounding box consisted of the label for the object or face (in numbers for the UK and in letters for the India dyad) and well its corresponding prediction confidence in percentage.

and B), then the most central bounding box with the largest confidence would be considered as the main object in view (e.g. toy elephant in figure 3.25 A and toy giraffe in image B). The onset and offset for each label consist of a minimum of three frames for it to be considered an event. The YOLO and MTCNN event file is then combined into a single csv file for each dyad.

3.3.7 Time is Very imPortant toolkit (TimeVP)

Unlike screen-based stimuli, parents and their infant's interaction within the 3D world goes beyond simply looking left and right on the screen. Following the extraction of meaningful measures using gaze reports from pupil labs and machine-learning tools (i.e., what infant or parent are looking at), the natural next step is to analyse the behavioural data to understand the dynamic interaction between parent and their infants. Here, we use Time is Very important (TimeVP) toolkit developed by the Developmental Intelligence Lab (also see C. Yu & Smith, 2016), available on GitHub https://github.com/devintel-lab/timevp. TimeVP uses the events files generated in the previous step and computes the variables of interest. First, the toolbox provides us with visualisation tools that can be crucial to checking data quality, validating preprocessing as well as examining patterns in our data (C. Yu, Yurovsky, & Xu, 2012). Next, it provides us with the following key measures of visual cognition to compute descriptive statistics: 1) Mean Look Duration (MLD) at the target (toys, face) in seconds, and 2) Switch Rate (SR) between targets (toys, face) per minute (normalised in 60 seconds). MLD was defined as the duration of looking towards the target (face, toy) without any looks away from the target (that is, offset - onset = MLD). SR was defined as looks between targets (face, toys). Lastly, it enables us to extract coupled behaviours to understand the temporal relations between two events such as episodes of joint attention led by parents vs. infants. Infant-led JA was defined as looking to a target initiated by the infant and could be terminated either by the infants themselves or the parent. Here, the infants' onset time of looking to the target (e.g. blue toy) was less than that of their caregivers' onset to the target (looking to the blue toy). Caregivers' onset look to the object (blue toy) was less that their infants offset to the same object (toy). The overlap between the looks to target was defined as JA. Same process was followed to calculate parent-led JA episodes except that the caregivers initiated the look (onset) to the target object.

To match the data format required for the TimeVP toolbox, each event detection file containing onset and offset frames was converted into onset and offset seconds. Parent and infant event files were saved in each individual dyad's folder (see figure 3.26 for an example). We then ran four MATLAB scripts for each cohort (6 months UK, 6 months India, and 9 months India). The first scripts visualised the sequential temporal events. Next, we ran two individual scripts for parents and infants to compute individual and overall statistics such as proportion, duration and frequency of looks on a target. Last, we ran the paired event script that informs us of the temporal relation between two events such as joint attention and its characteristics.

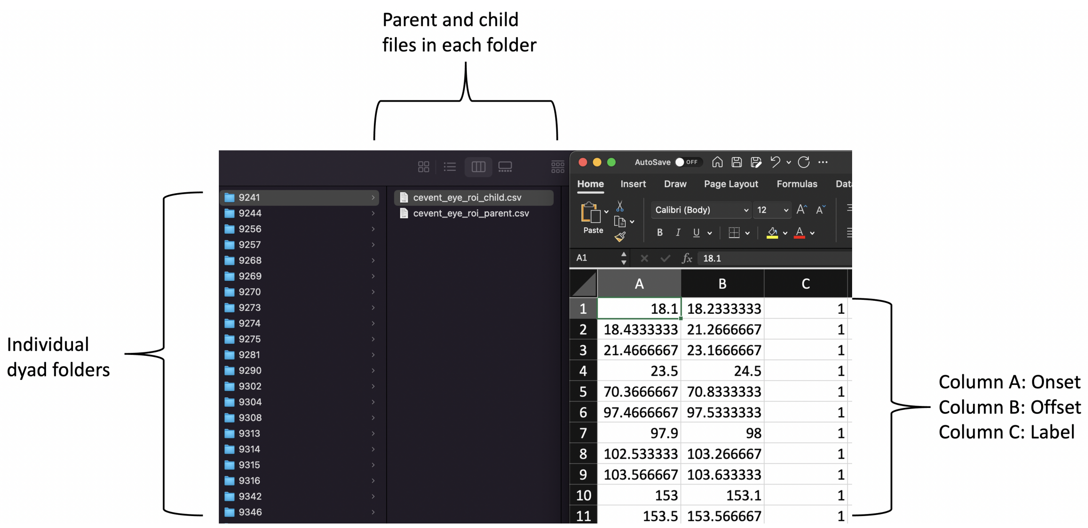


Figure 3.26: Data organisation and format for TimeVP toolbox. Columns A and B (right) show a sequence of events in seconds.

3.4 Results

3.4.1 Evaluation of MTCNN Accuracy

Results from manual coding for MTCNN revealed an overall percentage correct for faces was over 90% across all four dyads in the UK cohort (see table 3.8) and over 88% across all eight days in the India cohort (see table 3.9). The proportion of the hit, miss, false positive, and true negative is evaluated over the video length for each dyad. We added hits and true negatives to calculate all the correct identifications and rejections for each dyad ("Overall Percent Correct"). A visualised example of the parameters coded for a UK and India dyad can be seen in figure 3.27 and figure 3.28 respectively. The mean accuracy across all dyads in the UK was 95.41% (SD = 0.04) with 97.06% precision and 86.43%. The mean accuracy across dyads in India was 96.08% (SD = 0.04) with 85.98% precision and 83.63% recall. Within the UK dyads, the mean accuracy for the data captured via infants' head-camera was 96.93% (SD = 0.03) with a precision of 96.08% and recall of 88.89%, and the mean accuracy for the data captured via parents' head-camera was 93.90% (SD = 0.04) with precision of 97.67% and recall of 85%. Similarly, the mean accuracy for infants' head-camera in the Indian cohort was 97.36% (SD = 0.03) with a precision of 76.81% and recall of 76.24%, and the mean accuracy for parents' head-camera was 94.79% (SD = 0.04) with a precision of 88.31% and recall of 85.48%. Thus, compared to Long et al. (2022), MTCNN performed very well on our data sets with consistently high precision and recall for the UK data set. While the precision for the India data set was slightly lower than that of Long et al. (2022), the recall was consistently higher.

Table 3.8: The proportion of false positives, hits, misses and true negatives for MTCNN (UK).

Participant	False positive (%)	Miss (%)	Hit (%)	True Negative (%)	Overall Percent Correct (%)
Child1	1.6	1.3	6.4	90.7	97.1
Parent1	0.5	3.1	23.3	73.0	96.3
Child2	0.2	0.8	59.7	39.3	99.0
Parent2	0.4	8.6	54.8	36.1	90.9
Child3	0.7	6.6	1.9	90.8	92.7
Parent3	0.7	9.1	17.9	72.3	90.2
Child4	0.5	0.5	5.6	93.3	98.9
Parent4	1.3	0.6	25.3	72.9	98.2
Mean	0.7	3.8	24.4	71.1	95.4

Table 3.9: The proportion of false positives, hits, misses and true negatives for MTCNN (India).

Participant	False positive (%)	Miss (%)	Hit (%)	True Negative (%)	Overall Percent Correct (%)
India Child1	1.1	0.9	1.8	96.1	97.9
India Parent1	1.9	9.8	15.3	73.0	88.3
India Child2	1.0	5.7	1.8	89.4	91.2
India Parent2	0.8	3.3	19.4	76.5	95.9
India Child3	1.1	2.0	0.8	96.1	96.9
India Parent3	0.7	5.4	20.3	73.7	94.0
India Child4	1.4	0	4.2	94.4	98.6
India Parent4	0.9	0.5	14.7	83.8	98.5
India Child5	0.4	0.1	3.8	95.7	99.5
India Parent5	0.8	1.3	23.6	74.4	98.0
India Child6	1.4	0	0.2	98.4	98.6
India Parent6	9.1	0.6	3.5	86.8	90.3
India Child7	0.6	0	13.3	86.1	99.4
India Parent7	1.7	1.0	20.6	76.7	97.3
India Child8	2.3	0.9	4.9	91.9	96.8
India Parent8	2.4	1.6	20.9	75.1	96.0
Mean	1.7	2.1	10.6	85.5	96.1

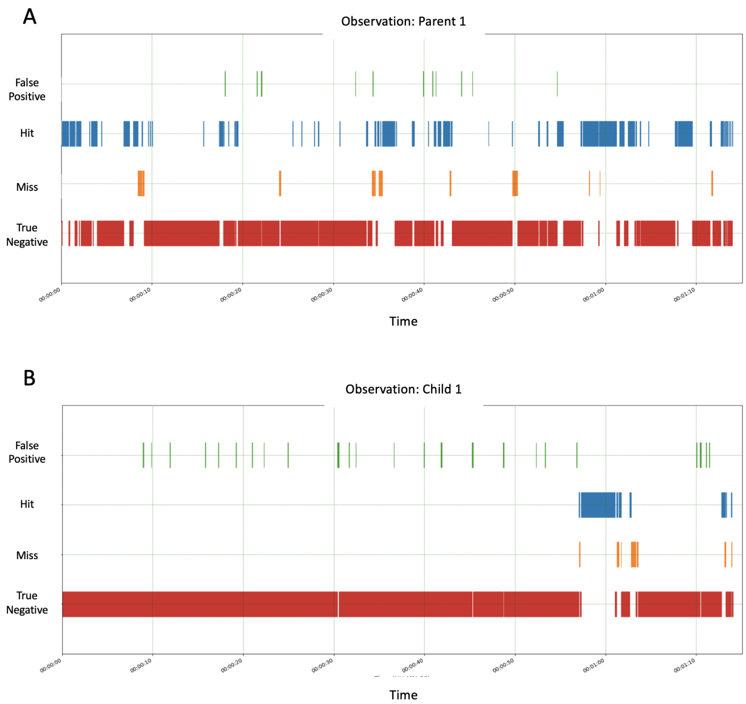


Figure 3.27: MTCNN results from manually coded parameters in BORIS. Plot A is the observation from a parent's headcamera and Plot B is the observation from an infant's headcamera from the UK cohort.

3.4.2 Evaluation of YOLO Accuracy

Table 3.10 and table 3.11 show results from manual coding for YOLO object recognition. The overall percentage correct for object recognition was over 70% for the UK cohort and over 84% for the India cohort (for visualisation see Appendix N.1 for a plot using BORIS for a toy from the UK and appendix N.2 for a toy from India). The mean accuracy across all dyads in the UK was 92.61% (SD = 0.07), and the mean accuracy across dyads in India was 96.52% (SD = 0.06). Within the UK cohort, infants' head-camera yielded a mean accuracy of 94.64% (SD = 0.06) and parents' head-camera of 90.58% (SD = 0.08). Similarly, within the Indian cohort, data captured from infants' head-camera resulted in a mean accuracy of 99.54% (SD = 0.05) and the parents' head-camera resulted in 93.44% (SD = 0.05). Thus, YOLO performed very well on our data sets.

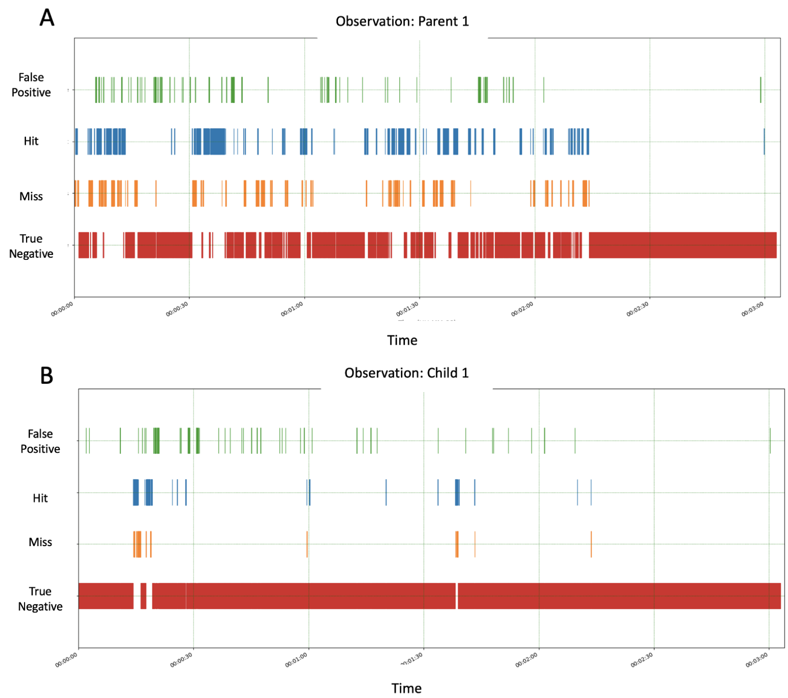


Figure 3.28: MTCNN results from manually coded parameters in BORIS. Plot A is the observation from a parent's headcamera and Plot B is the observation from an infant's headcamera from the India cohort.

Table 3.10: Proportion of false positives, hits, misses and true negatives for toys from dyads in UK.

UK Toys	False positive (%)	Miss (%)	Hit (%)	True Negative (%)	Overall Percent Correct (%)
TOY_BUT Child	0	3.1	61.5	0.35	96.9
TOY_BUT Parent	0	5.0	73.6	21.4	95.0
TOY_RAT Child	0.2	0.8	33.0	66.0	99.0
TOY_RAT Parent	0	4.7	29.1	66.2	95.3
TOY_CAM Child	0	11.3	24.3	64.4	88.7
TOY_CAM Parent	0	12.5	31.7	55.9	87.6
TOY_ELE Child	0	4.2	47.3	48.4	95.7
TOY_ELE Parent	0	8.1	17.5	74.4	91.9
TOY_GIR Child	0	10.0	43.7	46.3	90.0
TOY_GIR Parent	0	8.0	48.7	43.3	92.0
TOY_KET Child	0	2.7	79.9	17.4	97.3
TOY_KET Parent	0	18.3	75.3	6.4	81.7
TOY_APP Child	0	1.0	42.4	56.6	99.0
TOY_APP Parent	0	6.8	68.1	25.1	93.2
TOY_CUP Child	0	17.1	28.4	54.5	82.9
TOY_CUP Parent	0.1	27.4	32.1	40.4	72.5
TOY_TRA Child	0	2.5	65.8	31.7	97.5
TOY_TRA Parent	0.1	3.3	75.1	21.5	96.6
TOY_DUC Child	0	0.6	52.8	46.6	99.4
TOY_DUC Parent	0	0	69.6	30.4	1.0

Table 3.11: Proportion of false positives, hits, misses and true negatives for toys from dyads in India.

India Toys	False positive (%)	Miss (%)	Hit (%)	True Negative (%)	Overall Percent Correct (%)
TOY_RAT Child	0	1.7	32.4	65.9	98.3
TOY_RAT Parent	0	9.1	69.7	21.3	91.0
TOY_PUZ Child	0	1.6	10.3	88.1	98.4
TOY_PUZ Parent	0	0.7	47.7	51.6	99.3
TOY_MAN Child	0	2.2	40.9	56.9	97.8
TOY_MAN Parent	0	0	43.8	56.2	1.0
TOY_PLA Child	0.5	1.1	36.3	62.1	98.4
TOY_PLA Parent	0	5.5	30.2	64.3	94.5
TOY_CAN Child	0.1	0	17.6	82.3	99.9
TOY_CAN Parent	0.1	4.6	38.8	56.5	95.3
TOY_SPO Child	0	3.9	5.6	90.5	96.1
TOY_SPO Parent	0	15.6	6.7	77.8	84.5
TOY_YEL Child	0	0.1	9.7	90.2	99.9
TOY_YEL Parent	0	6.9	49.4	43.8	93.2
TOY_GLO Child	0.3	3.5	8.6	87.6	96.2
TOY_GLO Parent	6.8	3.5	37.9	51.8	89.7

3.4.3 Quantifying the dyadic data using the TimeVP toolkit

First, we looked at the visualised temporal data stream for parents and their infants during toy play. As can be seen in Figure 3.29, the parent-infant interaction varies qualitatively across the three cohorts. In our sample, dyads with 6-month-old infants in the UK tend to look more toward their social partner's faces (indicated in dark blue). There also seems to be a longer looking duration at the regions of interest (longer periods with a single colour). For the 6-month-old infant dyads in the India cohort there seem to be fewer looks to social partners' faces in 6-month-old Indian cohort. Nine-month-old infant dyads in India seem to be looking more at their partner's faces. White areas across the dyads indicate that the member of the dyad is looking at something other than the toys and faces that were labelled as the target of interest (see Appendix M.1 for the colour map for faces and individual toys). There is a good deal of variability in white space across dyads, with perhaps more white space for the 6-month India cohort.

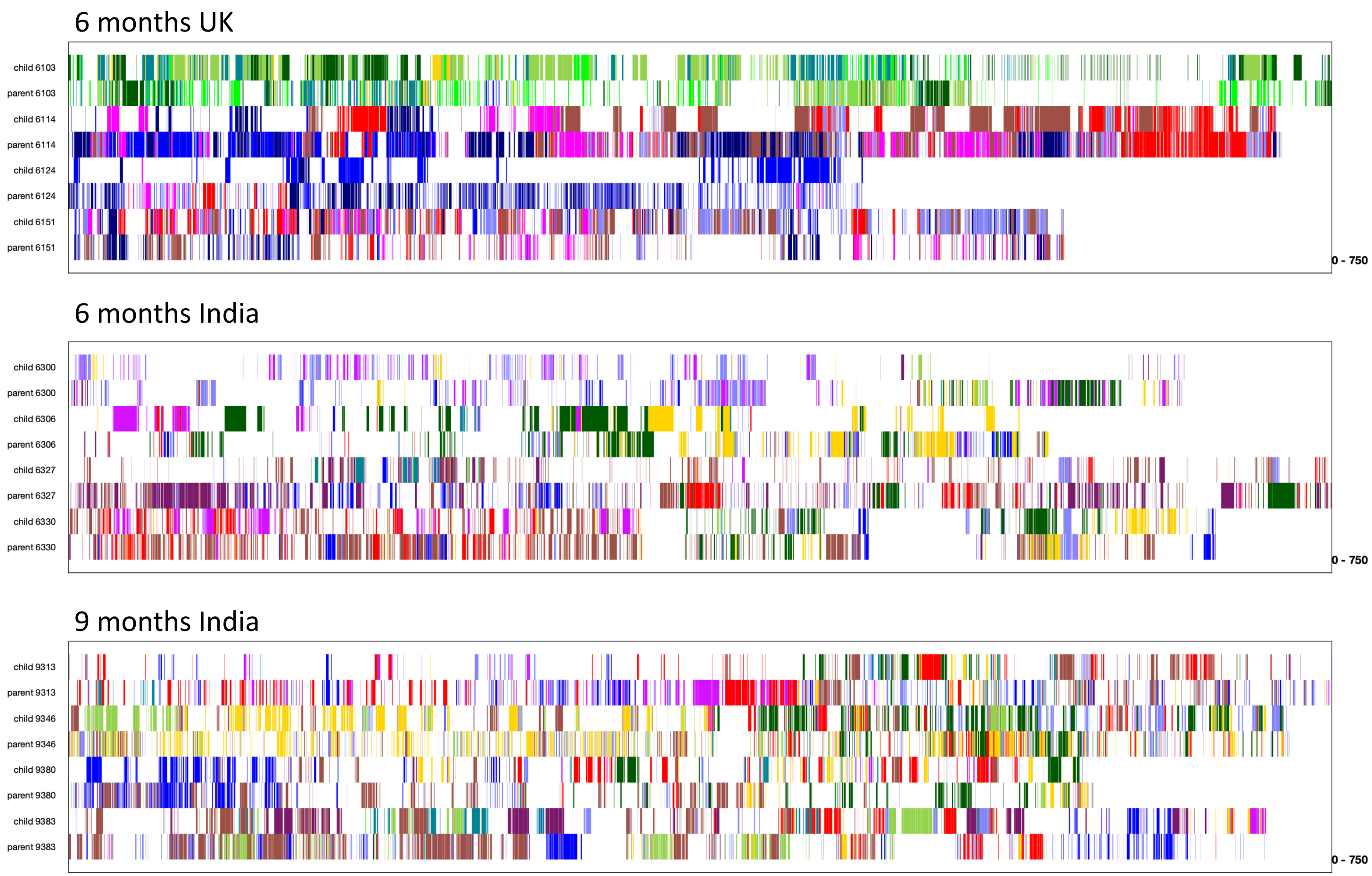


Figure 3.29: Visualisation of the coupled data stream and joint attention for each dyad across three cohorts (6 months UK, 6 months India and 9 months India) using the TimeVP toolkit. For each cohort, the first row indicates data for infants followed by their parents (second row). Each colour indicates a different object or social partner's face. Dark blue indicates the social partner's face.

Next, we look at the overall Mean Look Duration for caregivers and infants across cohorts (6 months UK, 6 months India and 9 months India). Figure 3.30 indicates that 6-month-old infants from the UK had the longest MLD (M = 0.70, SD = 0.25). Within the Indian cohort, the 9-month-old infants (M = 0.50, SD = 0.07) had slightly longer MLD than the 6-month-old infants(M = 0.49, SD = 0.35) during the interaction. Caregivers' in the UK also had the longest MLD during the dyadic interaction M = 0.38, SD = 0.16 compared to the other groups. This was followed by the caregivers of 6 month old infants (M = 0.36 SD = 0.01) and 9-month-old infants (M = 0.34, SD = 0.05) in India.

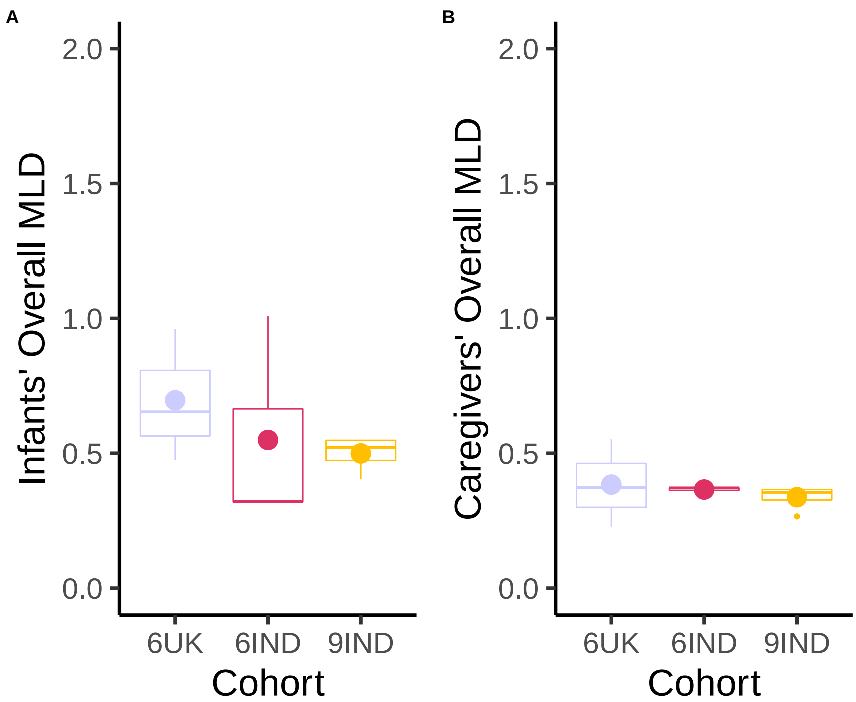


Figure 3.30: Mean Look Duration in seconds. A) Infants' Mean Look Duration for all targets. B) Caregivers' Mean Look Duration for all targets. Cohort 6UK represents 6-month-old infants from the UK; 6IND represents 6-month-old infants from India; 9IND represents 9-month-old infants in India.

In terms of switch rate, figure 3.31 depicts the highest switch rate per minute for the 9-months infant (M= 8.75; SD = 0.50) and their caregiver's' (M = 8.25; SR = 0.50) from India. 6-months-old infants (M = 7.0 ; SD = 1.83) and their caregiver's (M = 8.0; SD = 0.82) had lower switch rates with 6-month old UK dyads showing the least number of switches per minute (M = 4.67; SD = 0.58 , for infants and M = 5.0; SD = 1.0, for caregivers).

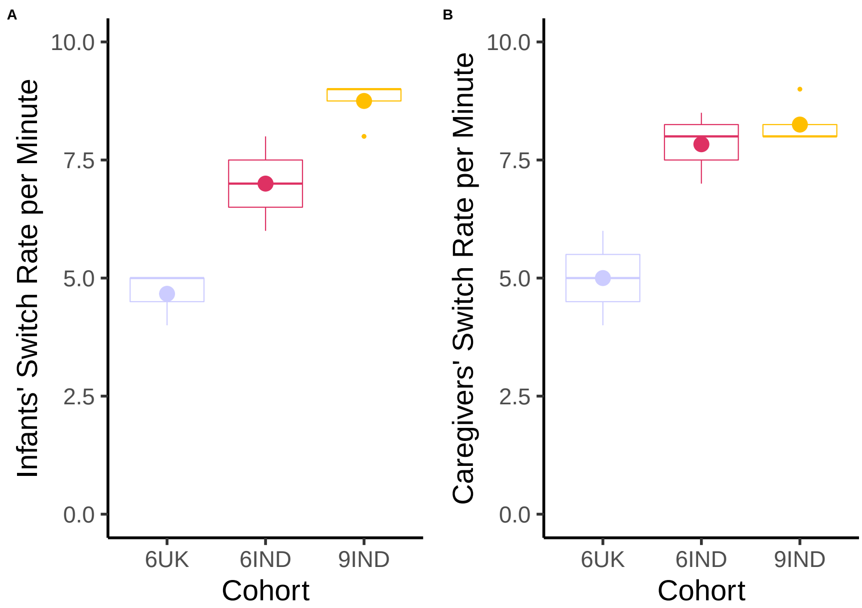


Figure 3.31: Switch rate between the target (toys, faces) per minute. A) Infants switch rates per minute. B) Caregivers switch rate per minute. Cohort 6UK represents infants and caregivers of 6-month-old infants from the UK; 6IND represents infants and caregivers of 6-month-old infants from India; 9IND represents infants and caregivers of 9-month-old infants in India.

Next, we divided MLD into two categories, that is, MLD when looking at a target alone and MLD when looking at a target in a joint attention bout. As seen in figure 3.32, 9- months-old infants from our Indian sample had the longest MLD when looking at a target (toys, face) in a non-joint attention episode (M = 0.46; SD =0.63) followed by 6-months old infants in the UK (M= 0.45; SD = 0.72) followed by 6-month-olds in India who had the shortest MLD (M = 0.43; SD = 0.79). Infants from all three cohorts showed an increased MLD during episodes of joint attention (note the increased range of the y-axis in the figure) with 6-months old infants from the UK showing the longest MLD (M = 1.74; SD = 1.97). This was followed by 6-months-old infants from

India (M = 1.25; SD = 1.50) which did not vary much from the 9-month-old infants' MLD (M = 1.06 ; SD = 1.07).

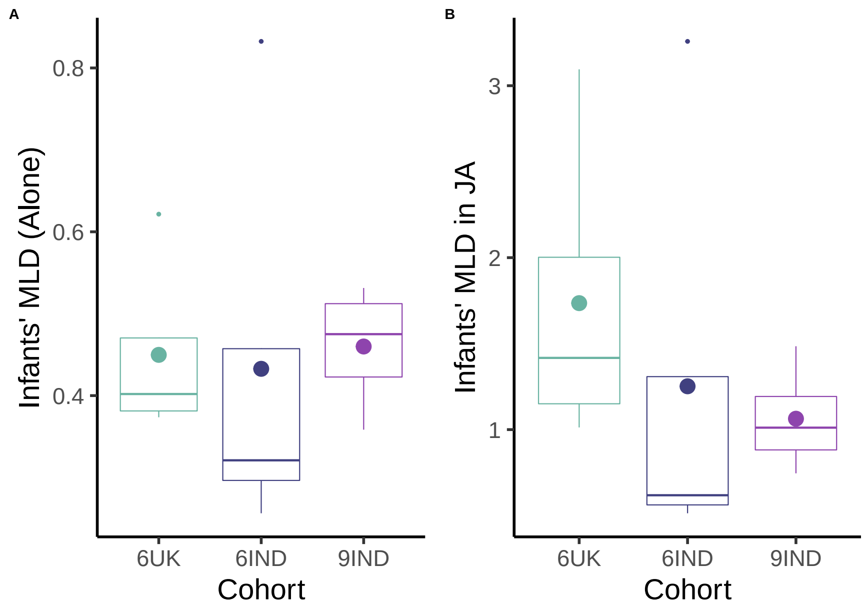


Figure 3.32: Mean Look Duration in seconds. A) Infants' Mean Look Duration alone. B) Infants' Mean Look Duration during Joint Attention. Cohort 6UK represents 6-month-old infants from the UK; 6IND represents 6-month-old infants from India; 9IND represents 9-month-old infants in India.

Similar to infants looking behaviour, caregivers tended to have longer MLD in joint attention episodes versus when looking at the target alone (see figure 3.33). Caregivers of 6 months old infants in the UK showed the longest MLD (M = 0.35; SD = 0.44) when looking at the target alone while caregivers of 6 months old infants (M = 0.35; SD = 0.48) and 9-month-old infants (M= 0.31; SD = 0.38) in India showed shorter MLD while looking at a target alone. Caregivers of 6-month-old infants from India tended to have the highest MLD in joint attention episodes (M = 0.74 , SD= 0.91), followed by caregivers of 6 months old infants in the UK (M = 0.53, SD = 0.69) and caregivers of 9-months-old infants in India (M = 0.64, SD = 0.71).

As seen in Figure 3.34A, over 50% of joint episodes are initiated by infants, suggesting a common trend across cohorts where caregivers tend to follow in

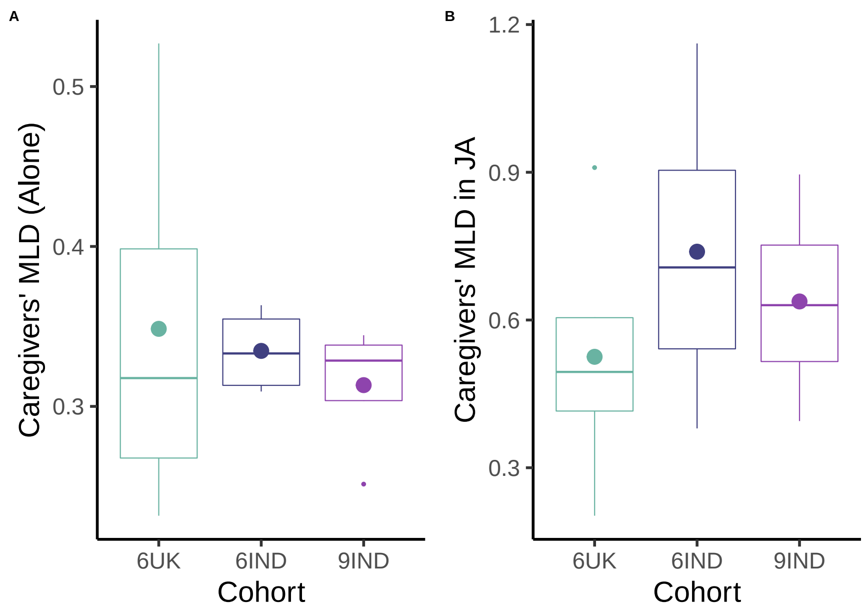


Figure 3.33: Mean Look Duration in seconds. A) Caregivers' Mean Look Duration alone. B) Caregivers' Mean Look Duration during Joint Attention. Cohort 6UK represents caregivers of 6-month-old infants from the UK; 6IND represents caregivers of 6-month-old infants from India; 9IND represents caregivers of 9-month-old infants in India.

on their infants' attention. A large proportion of joint attention was initiated by infants in the UK compared to India. Results did not indicate a large difference between the proportion of joint attention episodes led by 6- and 9- month old infants in India. Contrarily, less than 50% of joint attention tends to be terminated by infants. Figure 3.34B shows that infants from the UK tend to terminate joint attention the least whereas 6-months infants from India tend to terminate joint attention episodes the most in our sample.

Next, we looked at whether there were any differences in how caregivers and infants' MLD varied when looking at each others' faces in joint attention episodes (JA) versus when looking at their partners' faces when not in joint attention (i.e., alone). As seen in figure 3.35, infants looked longer at their caregiver's faces when in joint attention across cohorts compared to when looking at the partners' faces alone. Infants in the UK looked the longest to the care-

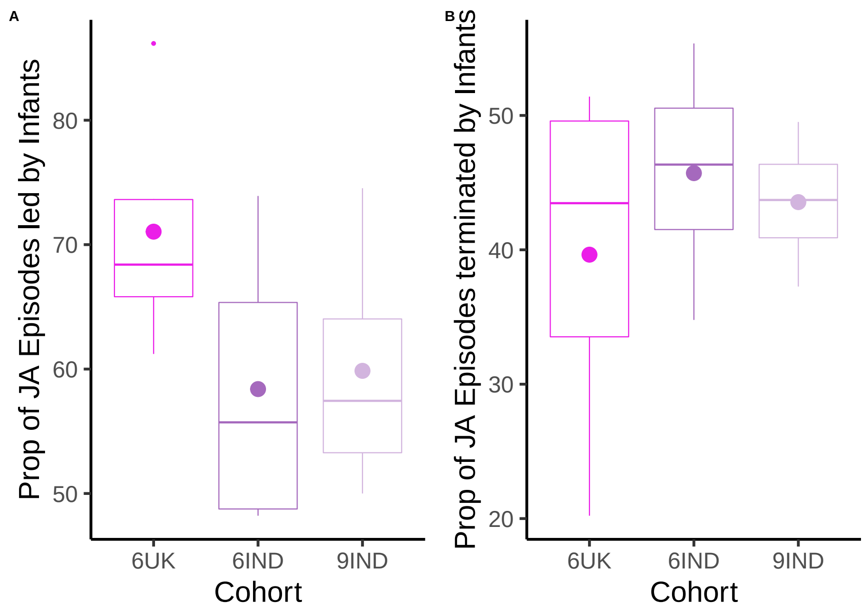


Figure 3.34: A) The proportion of infants led to joint attention in percentage. B) Proportion of joint attention terminated by infants in percentage. Cohort 6UK represents 6-month-old infants from the UK; 6IND represents 6-month-old infants from India; 9IND represents 9-month-old infants in India.

giver's face, both alone (M = 0.34; SD = 0.45) and during the joint attention episode (M = 1.47; SD = 1.18). Six-months old infants from India has shorter MLD to caregiver's faces when alone (M = 0.27; SD = 0.40) with a slight increase in MLD during joint attention (M = 0.29; SD = 0.68). Nine month old infants showed shorter MLD to caregiver's faces when alone (M = 0.26; SD = 0.28) compared to when attending to each other's faces in joint attention (M = 0.71; SD = 0.70).

Results for caregivers show a similar trend of looking to faces as for infants across cohorts (see figure 3.36). Caregivers across cohort showed an increase in MLD when looking at faces in joint attention (M = 0.64; SD = 0.23, for caregiver's of 6-months old infants in the UK; M =0.48; SD = 0.76, for caregiver's of 6-month old infant in India, and M = 0.55; SD = 0.53, for caregiver's of 9 month old infants in India) compared to when looking at the infant's face alone (M = 0.28; SD = 0.32, for caregiver's of 6-months old infants in the UK;

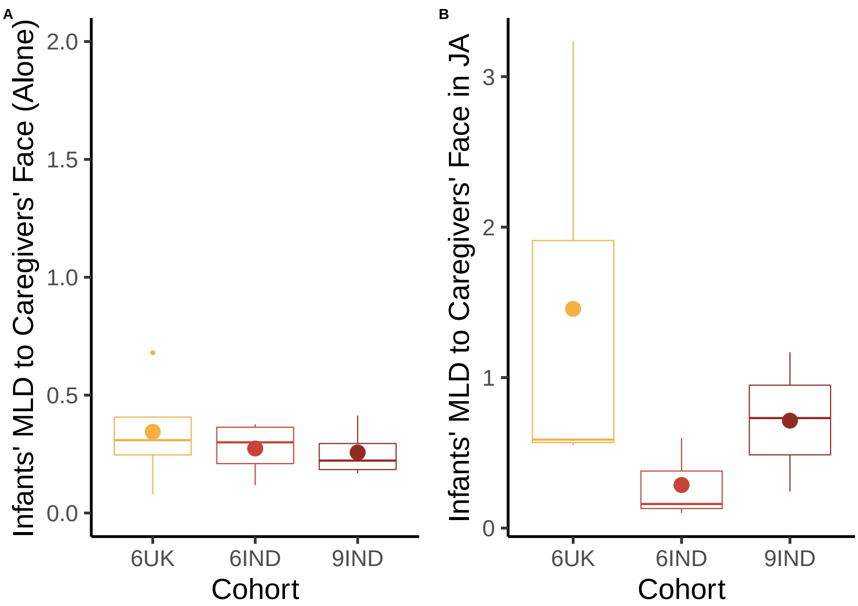


Figure 3.35: Mean Look Duration in seconds. A) Infants' Mean Look Duration to caregivers' faces when caregivers are not looking at infants' faces. B) Infants' Mean Look Duration when both infants and caregivers are looking at each other's faces. Cohort 6UK represents 6-month-old infants from the UK; 6IND represents 6-month-old infants from India; 9IND represents 9-month-old infants in India.

M = 0.25; SD = 0.27, for caregiver's of 6-month old infant in India, and M = 0.21; SD = 0.23, for caregiver's of 9 month old infants in India).

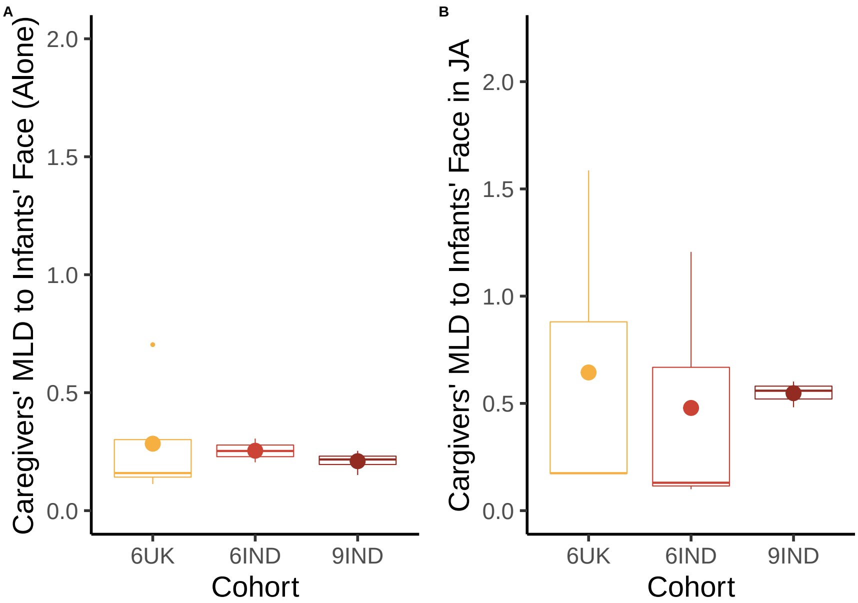


Figure 3.36: Mean Look Duration in seconds. A) Caregivers' Mean Look Duration to infants' faces when infants are not looking at infants' faces. B) Infants' Mean Look Duration when both infants and caregivers are looking at each other's faces. Cohort 6UK represents caregivers of 6-month-old infants from the UK; 6IND represents caregivers of 6-month-old infants from India; 9IND represents caregivers of 9-month-old infants in India.

3.5 Discussion

The current chapter aimed to create a single methodological pipeline to process real-world, dynamic, parent-infants interaction recorded using head-mounted eye-trackers and a camera in a low-resource setting. In this process, we provide solutions to two key issues in developmental science: 1) diversifying our participant pool by going beyond the western, middle-class society and 2) processing parent-infant interaction in a less laborious and efficient way that can be used on large data set (or videos of long periods of time).

In the present chapter, we described a pipeline and accompanying code to process the dynamic interaction between parents and their infants in a non-biased way that is transferable across cultural settings. We, first, demonstrate that parent-infant interaction videos, from a first-person perspective, can be obtained from both high (urban UK) and low (rural India) resource settings. Next, we generated gaze reports for caregivers' eye-trackers and synchronised annotation files using open-source pupil player software. This was accompanied by code that extracted synchronised frames from caregiver-infants interaction videos. Once the synchronised frames were extracted, we used machine learning algorithms to detect faces and recognise toys in the videos.

Note, we focus on face detection and not face recognition as there is only one other face (either infants or their parents) in the videos. Additionally, we used pre-trained face detection models (MTCNN) that have already been labelled on 32,203 images with 393,703 faces labelled, making the process of face detection less laborious. However, using a pre-trained model does not come without limitations. As seen in figure 3.15, the initial accuracy (pre-filtering stage) for our UK and India cohort was 70% and 60% respectively. While adding the filtering step increased the accuracy by 20-28%, we suspect that the WIDER face dataset used by MTCNN to create a pre-trained model consists of fewer infant faces, and mostly of adults and children faces. It would be useful for future work to explore this by running two separate models of equal quality on parent and infant faces and comparing the accuracy. Alternatively, the MTCNN model can be updated by training infant faces (across ethnicity) and adding the weights to the existing pre-trained model.

Next, we showed that a YOLO object recognition algorithm performed remarkably well across cohorts. It would be useful for future work to validate the pipeline using novel objects. We acknowledge that the pipeline is not completely automated and requires the user to make decisions, particularly at the filtering step, after making a qualitative pass on the output obtained from the machine learning algorithm. We expect, however, that the filtering steps applied here will be applicable to other projects.

Given that training the data can be time-consuming, as it requires labelling and annotating a large dataset, setting up a larger database could solve this issue. Training up a network with a large set of commercially available standardised toys, worldwide, could then be used in multiple studies. This would enable researchers across borders to simply use the training weights to run the pipeline without any additional training. Not only would this simplify using the pipeline but also make the process faster.

Future research could also expand the pipeline by looking at other features such as detecting hands (F. Zhang et al., 2020). Work by C. Yu and Smith (2013) in the western setting have noted that the infants' and toddlers' visual field often consists of hands and hands manipulating toys. Similarly, a study by Jayaraman, Fausey, and Smith (2017) have noted an age-related increase in the input of hands in infants' visual view. There, adding a feature of hand can enable us to replicate the findings in low-resource settings. Moreover, it can also further our research in understanding deaf infants and caregiver dyads using sign language (R. Brooks, Singleton, & Meltzoff, 2020).

In terms of data analyses, the TimeVP toolbox successfully enabled us to compute measures of interest such as MLD, shift rate, and caregiver versus infant-led joint attention episodes. Given that the aim of this chapter was to construct a methodological pipeline yielding key measures of interest, we chose to report only the descriptive statistics from a small number of participants. Interestingly, however, this small sample revealed several similarities and differences across cohorts.

Some key highlights from the results indicated that overall infants tend to lead a large proportion of joint attention bouts compared to their caregivers. Specifically, infants from the UK tended to lead a larger proportion of joint attention bouts. This may be indicative of the western child-centred approach where infants are encouraged to lead and guide a play interaction (Lancy, 2014). Similarly, infants in the UK also tend to show longer MLD to targets (e.g. faces) during joint attention bouts than alone. Previous work on infants' sustained attention, with dyads from a western context, have also noted that infants tend to engage in looking longer to targets in a joint attention bout compared to alone (Wass et al., 2018; C. Yu & Smith, 2016). Thus, finding similar results serves as a good validation for our pipeline.

To further evaluate the utility of this pipeline, the following chapter will apply the given pipeline to a larger-data set (sample size >20) across the two resource settings, examining the similarities and differences across the three cohorts. Evaluating the pipeline on a larger data set will enable us to connect the measures of visual cognition obtained from lab and lounge settings in a robust manner.

Chapter 4

Exploring caregiver-infant interaction across cultural settings using the machine learning pipeline

4.1 Introduction

In Chapter 3, we addressed the methodological pipeline created to process and analyse the data recorded during caregiver-infant interactions. We discussed the use of technological advances to quantify the recorded data in a less laborious and more objective manner across both high (urban UK) and low-resource (rural India) settings. Using machine learning algorithms, we successfully extracted key measures of visual cognition related to joint attention from the recorded dynamic caregiver-infant interactions. Measures included caregiver and infants' Mean Look Duration (MLD) to objects and to faces, Switch Rate (SR) between targets of interest, and episodes of caregiver-and infant-led joint attention. Chapter 4 expands on Chapter 3 by implementing the machine learning pipeline on a larger data set from the high and low resource contexts. The goal of the chapter is to explore and understand the similarities and differences between the measures of visual cognition related to joint attention across socio-cultural contexts (India and UK) as well as between age cohorts (6- and 9- months old infants) for the Indian participants.

Joint attention has been conceptualised and operationalised in different manners across studies, time and socio-cultural contexts (Bard et al., 2021; Siposova & Carpenter, 2019). While there is a general consensus that the infant's ability to engage in joint attention lays the groundwork for developmental advances such as language learning (Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; Mundy & Gomes, 1998), social cognition(Mundy & Newell, 2007), and theory of mind (Nelson, Adamson, & Bakeman, 2008), characteristics of joint attention in social interaction are complex and can emerge via many pathways. These include initiating and responding to joint attention, maintaining attention to the common target of interest, ending the joint attention bout as well as shifting to or disengaging from an object. These instances of coordinated joint attention, in turn, predict infants' engagement in sustained attention (C. Yu & Smith, 2016) and vocabulary development (Abney, Smith, & Yu, 2017). For instance, a study by C. Yu and Smith (2016) explored the influence of social context on parent-infant interaction during a free-flowing play session. They found that when the parent visually attended to the same object to which the infant was attending, infants attended to that object for longer than in the case when the parent was attending to a different object. Caregivers' labelling an object in such instances also predicted infants' later vocabulary.

As argued in Chapter 1, infant development takes place in interaction with others, and therefore, specific aspects of said interactions should be expected to be of relevance. Characteristics of attention constitute the ability to engage, maintain, disengage, and shift focus (Rose, Feldman, & Jankowski, 2005). Holding attention for longer or shorter periods of time on a given object or face, the rate at which visual attention shifts between targets, and the different ways in which joint attention is initiated, sustained, and ended, not only can be indicative of the development of the infant but also may shape development.

Thus, the current study examines these characteristics in a large sample of infants in high- and low-resource settings. This comparison is important as most of our knowledge about joint attention comes from research conducted with infants from what Henrich et al. (2010) described as the WEIRD (i.e., Western, Educated, Industrialised, Rich, and Democratic) setting (also see Bard et al., 2021). Therefore, one cannot simply claim a universal joint attention behaviour from a small sample of participants. Indeed Nielsen, Haun, Kärtner, and Legare (2017) showed that less than 3% of research in developmental psychology comes from countries with approximately 85% of the world's population. Therefore, the current chapter sought to explore and understand the real-life ecologies of visual cognition during caregiver-infant interaction in a "non-WEIRD" and low resource context. To do so, we use the objective measures of visual cognition extracted using the machine learning pipeline in Chapter 3 and understand them through the lens of the eco-cultural model of parenting (Keller et al., 2005; Keller, 2007).

In the eco-cultural model of parenting, Keller (2007) identified two key parenting styles, distal and proximal. Distal parenting style involves exclusive focus on face-to-face interaction, object simulation, and child-centred responsiveness as well as emphasising "positive" affect during caregiver-infant interaction. This parenting style has typically been characteristic of the urban, middle-class families from western cultures. The approach entails a pedagogical way of playing with the child, i.e., with the aim of teaching the child (Lancy, 2010). On the other hand, the proximal parenting style constitutes modalities such as body contact, tactile simulation, focus on calming and soothing the infant as well as a directive, adult-centred interaction, which has been typically characteristic of rural, subsistence farming families of traditional villages (e.g. Abels et al., 2017, work in rural Gujarat in India). Proximal parenting style takes on the approach of responding to their infant's distress signals, almost always by breast-feeding, or doing something to calm down the infant, i.e., *"a quiet baby is a healthy baby"* (Lancy, 2007, p.275).

Other research examining the cultural similarities and differences in maternal parenting have also looked at parenting style and conversational patterns between mothers and their 3-month-old infants in Delhi and Berlin using cultural models of autonomy and relatedness. This work found that mothers in Delhi shaped the interactions with their 3-months infants by leading and defining the structure of the play (Keller et al., 2010). By contrast, infants in Berlin tended to take on an active role in leading the interaction by directing their mothers' attention to, e.g., a toy. The autonomous model, an extension of the distal parenting style, involves caregivers addressing their infants as someone with an agency and emphasising the development of autonomy, wherein, infants actively directed and initiated interactions. On the other hand, the parenting style in urban Delhi consisted of a combination of autonomous-relatedness, i.e., mothers having higher formal education (linked to the autonomous style) as well as a bias towards traditional family ties and kinship (linked to relatedness). This was further indicated in the maternal parenting style. That is, while the mothers' showed a bias for proximal caregiving style (i.e., more body contact), they did not significantly differ in the use use of both proximal and distal parenting style during the play. Thus, they used body contact, tactical stimulation as well as playing with the object. By contrast, mothers in Berlin showed a clear bias towards distal (and autonomous) parenting style.

It is important to note that the goal of this study is not to make inferences regarding what parenting style is better or worse. Instead, we focus on evaluating the similarities as well the differences in how parent and infants deploy their visual attention during instances of social interactive play. Therefore, we include both the UK and India samples in our study to understand the infant and their caregiver's deployment of visual attention within its social context, hence broadening our understanding of different populations. Additionally, in line with the work by C. Yu and Smith (2016), in the present research, we use the term "joint attention" to refer to a process in which caregiver and their infants focus their visual attention, together, on a common object or each others' faces at the same time (also see C. Yu & Smith, 2013). As mentioned in Chapter 3, joint attention includes any looking behaviour for a minimum of three frames (i.e., 90msec).

Despite our exploratory focus, the evidence of developmental trajectories and existing differences in social interaction across cultural settings, allow considering some specific predictions. First, in regards to switch rate and MLD measures among infants, the more developed visual cognitive system among older infants is expected to lead to the 9-month-old group displaying shorter MLD and a higher switch rate overall. This is likely to be sustained across targets, particularly when looking at the target alone. We further explore whether the pattern is sustained during episodes of joint attention. For the parent measures, given that the Indian cohorts are defined based on the infants' age, there is no reason to expect any differences between them based on parental characteristics. However, variation may appear as a result of differences in infant characteristics. That is, greater mobility and agency among older infants may lead caregivers to deploy their cognitive resources to a larger degree, leading to greater switch rate and lower MLD. Furthermore, differences are explored across cultural groups given the variation in parenting styles. Differences are specifically expected during episodes of joint attention. Indeed, the deployment of visual cognition (switch rate and MLD) among parents and infants may differ depending on who initiated the episode of joint attention and the degree to which the initiation fits the parenting style (e.g., parent-led bouts in the more directive Indian setting). Similarly, regarding the initiation of joint attention, the distal style of parenting expected in the UK is predicted to result in a greater proportion of joint attention bouts being initiated by infants, the more directive proximal style common to India resulting in a greater proportion of joint attention being initiated by caregivers.

4.2 Study 3: What are the similarities & differences in caregivers & their infants' visual exploratory across cultural contexts?

4.3 Method

4.3.1 Participants

Data for the current study are reported from a sub-sample of the data reported in Chapter 2 and Chapter 3. For the UK sample, we present data from 25 dyads of 6-month-old infants (15 females) recruited by the Developmental

Dynamics Lab at the University of East Anglia, the UK through the same procedure as mentioned in chapter 3. The data for the UK sample was collected between 2017 and 2020. Parents were informed of the experiment's aim and procedure, and written consent was obtained. Remuneration comprised of 20 pounds, travel expenses, a t-shirt and a toy for each participant.

For the Indian sample, we report data from 31 six-month-old infants ±15 days (17 females) and 37 nine-month-old ±15 days (15 females) infants' and their caregivers. Characteristics of the sample from both India and the UK are summarised in table 4.1. The procedure and protocol for recruitment of participants were the same as Chapter 3. The data for the India sample was collected between 2016-2018.

The studies in India and UK were supported by the Bill & Melinda Gates Foundation Grant No. OPP1164153 and the NIH Grant No. R01HD083287, both were awarded to Prof. John P. Spencer. The data reported here is part of a larger study examining infant brain health in India and probing the neural basis of visual working memory in early development in the UK.

4.3.2 Materials and Stimuli

The stimuli and the setup were identical to the study in Chapter 3 for both UK and Indian cohorts.

4.3.3 Procedure

The procedure for capturing caregiver-infant interaction was the same as the study in Chapter 3.

4.3.4 Data processing

We processed the data, step-by-step, through the pipeline mentioned in Chapter 3, except for the visualisation and validation through BORIS manual coding. That is, each dyad's video was processed through pupil player software wherein the synchronised frames were annotated for onset and offset of caregiver-infant interaction. Here, we also processed caregivers' eye-tracker to generate gaze reports. Caregiver and infants' synchronised videos were extracted as synchronised frames excluding any crying event for more than a minute. This step included generating new gaze position files that correspond to each extracted frame. Next, we submitted a bash script to run each dyad's synchronised frames through MTCNN face detection and YOLO object recognition algorithm using the university's HPC. Once the MTCNN and YOLO predictions were generated, the data went through a filtering stage using the same filtering steps mentioned in Chapter 3. Object and gaze coordinates were overlapped followed by generating an events file that consists of onset and offset for each event (e.g., toy or face look). The final step included processing the data through the TimeVP toolbox (also see C. Yu & Smith, 2016) that computed measures of Mean Look Duration (MLD), Switch Rate (SR) and episodes of caregiver and infant-led joint attention.

4.3.5 Analytic strategy

The analyses are divided into three sections. In the first section, we look at scores of the overall mean look duration and switch rate per minute, i.e., including a focus on both faces and toys for both caregivers and their infants across the three cohorts (6UK, 6IND, 9IND). In section 2, we shift the focus to exploring the episodes of joint attention, addressing differences between the three groups in terms of the total number of episodes of joint attention, the

Table 4.1: Summary of key demographic features of our participants from UK and India.

	6UK		6IND	9IND
N	25 (15 females)	**N**	31 (17 females)	34 (15 females)
Mother Education		**Mother Education**		
Left school	1	Primary pass	15	16
Up to A levels	4	Secondary and higher education	16	18
Bachelor's Degree	9	**Father Education**		
Masters Degree	7	Primary pass	15	16
Doctorate or Professional Degree	4	Secondary and higher pass	16	18
Father Education		**Caste**		
Left school	0	General	2	1
Up to A levels	6	Other Backward Class (OBC)	9	14
Bachelor's Degree	11	Schedule caste/ tribe (SC/ST)	20	19
Masters Degree	5	**Family Income in INR**		
Doctorate or Professional Degree	3	<100000	24	25
Ethnic Group		>= 100000	7	9
African	0			
Asian	1			
Mixed	2			
White	22			
Family Income Median in GBP				
<20000	3			
>= 20000 & <40000	5			
<= 40000	17			

Note. 6UK denotes dyads with 6-month-old infants from the UK, 6IND denotes dyads with 6-month-old infants from India and 9IND denotes dyads with 9-month-old infants from India.

proportion of the said total in which the joint attention was led by the infants, and the proportion which was terminated by the infants. Across the first two sections, we performed Welch two-sample t-tests to compare the means of cohorts. In section 3, we analyse the differences in infants' mean look duration towards toys and faces across three conditions: when the infant was looking at toys or caregivers' faces by themselves, i.e., without the caregiver looking at the same target; when the infant initiated the joint attention episode, and when the caregiver initiated the joint attention episode. The same analyses on mean look duration towards toys and faces were then repeated for the caregivers. As with natural behaviour, the mean-looking duration for participants was such that most were very brief and some very long. Therefore, to compare the MLD between cohorts we used the Mann-Whitney U test due to the skewed distribution of looking behaviour rather than Welch's two-sample t-test that assumes normality (Gibbons & Chakraborti, 2011; Yuan, Xu, Yu, & Smith, 2019).

4.4 Results

Figure 4.1 illustrates the raw data, that is, the coupled data streams from caregiver-infant interactions across the three cohorts (6UK, 6IND, 9IND). As can be seen in the figure, there are some interesting patterns across cohorts. First, there is more face looking (see dark blue colour) in the 6-month cohorts (both 6UK and 6IND) with longer bouts of sustained looking to faces. Next, there is more white space in the 6UK time series. This may reflect differences in the context of the interactions. Recall that UK dyads were tested in the home, while Indian dyads were tested in a lab with few objects in the surroundings. Thus, it is likely that UK infants and caregivers looked at other objects in their surroundings more often – objects which were not captured by the machine learning approach. Finally, it looks like the 9-month cohort in India has more distributed looking patterns with shorter bouts of sustained looking. This may reflect a developmental shift in looking patterns from 6 to 9 months.

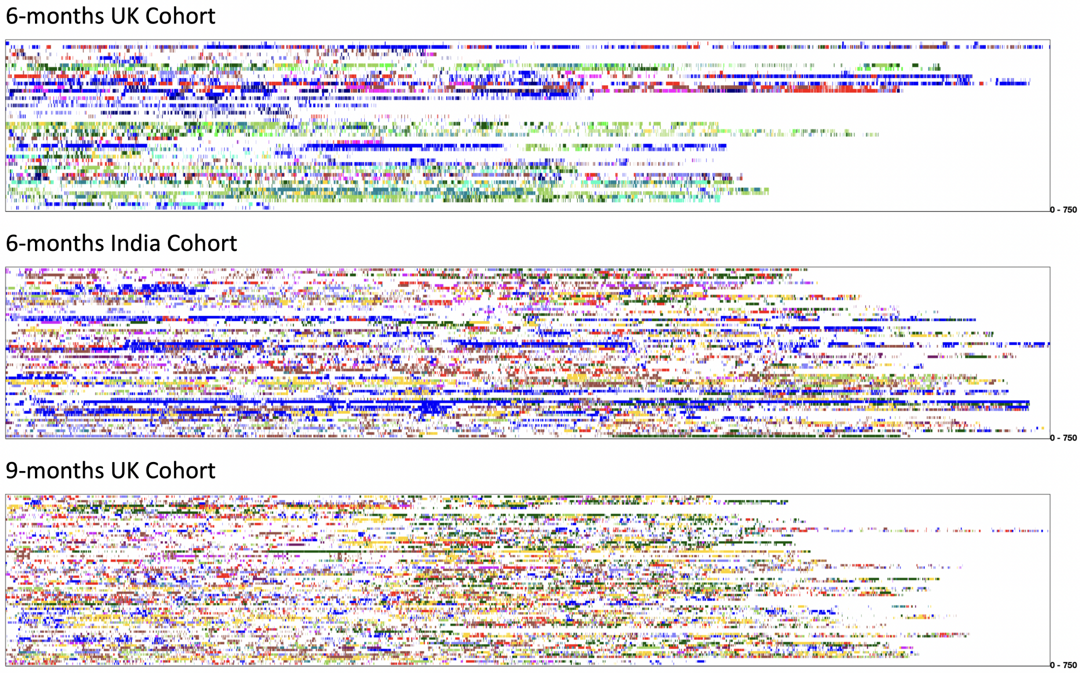


Figure 4.1: Visualisation of the coupled data stream and joint attention for each dyad across three cohorts (6 months UK, 6 months India and 9 months India) using the TimeVP toolkit. For each cohort, the first row indicates data for infants followed by their caregivers' (second row). Each colour indicates a different object or social partner's face. Dark blue indicates the social partner's face. White space indicates looking at a non-target.

4.4.1 Overall MLD & SR

Scores of the overall mean look duration and switch rate per minute for infants and caregivers across the three cohorts are shown in figure 4.2A and B (also see Table 4.2). The Welch two sample t-tests on the the MLD of infants across the three cohorts found no significant differences ($t_{6IND,9IND}(42.79)$ =

Table 4.2: Overall means and standard deviation for caregiver and infants Mean Look Duration and Switch Rate.

	Infant		Caregiver	
		MLD		
Cohort	**Mean**	**SD**	**Mean**	**SD**
6UK	0.46	0.23	0.30	0.12
6IND	0.47	0.22	0.38	0.07
9IND	0.53	0.11	0.35	0.05
Cohort		**SR**		
6UK	4.54	1.06	4.67	0.96
6IND	7.35	1.4	7.97	1.0
9IND	8.25	1.02	8.32	0.78

Note. MLD stands for Mean Look Duration (in seconds). SD stands for Standard Deviation (in seconds). 6UK denotes dyads with 6-month-old infants from the UK, 6IND denotes dyads with 6-month-old infants from India and 9IND denotes dyads with 9-month-old infants from India.

-1.38, p = 0.18; $t_{9IND,6UK}$ (30.53) = 1.46, p = 0.15; $t_{6IND,6UK}$ (49.08)= 0.20, p = 0.84). On the other hand, the switch rate between targets (e.g., between toys, or between toys and faces) differed significantly across cohorts. Infants in the 9-month-old cohort from India had higher switch rate compared to the 6-month old infants in India ($t_{6IND,9IND}$(54.09) = -2.94, p <0.01). In turn, the 6-month-old infants from India displayed a significantly higher switch rate compared to those from the UK($t_{6IND,6UK}$(52.98) = 8.46, p <.001). Note that the low switch rate for the 6UK infants is consistent with the observation above that visualisations from the UK cohort had more white space. More white space would create longer gaps between events, thereby lowering the shift rate per minute.

Analyses for caregivers' MLD revealed that the caregivers' of 6-month-old infants in India tended to have longer MLD than caregivers of 6-month-old infants in the UK ($t_{6UK,6IND}$(34.74) = 2.85, p <.01). However, there were no differences between the two Indian cohorts ($t_{6IND,9IND}$(56.94) = 1.82, p = .074; also see figure 4.2C) nor between the MLD of caregivers from the 6-month cohort in the UK and 9-month old cohort in India ($t_{6UK,9IND}$(28.88) = -1.95, p

= .061).

In term of switch rate, caregivers of 9-month old infants in India switched at a significantly higher rate than caregivers in the UK ($t_{6UK,9IND}(42.04)$ = -15.56, p <.001; see figure 4.2D). Caregivers in the 6-month-old infant cohorts also differed in their switch rate ($t_{6UK,\,6IND}(50.65)$ = 12.49, p <.001) such that caregivers of the Indian cohort showed significantly higher switch rate. However, there was no difference between the two cohorts in India ($t_{6IND,9IND}(58.49)$ = -1.63, p = .11). Again, this is consistent with the observation of greater white space in the raw data visualisation in the UK cohort which would lead to lower switch rates in the UK caregivers.

4.4.2 Proportion of Joint Attention Episodes

Table 4.3 shows the descriptive statistics for the total bouts of joint attention, and the proportion of joint attention bouts led and terminated by infants. Results revealed no significant differences between the cohorts in regards to overall joint attention episodes ($t_{6IND,9IND}(50.05)$ = 1.84, p = .07; $t_{9IND,6UK}$ (33.91) = 1.91, p = .07; $t_{6IND,6UK}$ (29.89)= 0.50, p = .62; figure 4.3A). However, the proportion of infant-led joint attention differed significantly across the three cohorts, with 6-months-old infants from India leading a significantly smaller proportion of joint attention episodes than their 9-month-old counterparts ($t_{6IND,9IND}(40.99)$ = -2.24, p <.05). In turn, 9-month-old infants from India led a significantly smaller proportion of joint attention episodes compared to the UK cohort ($t_{6UK,9IND}(30.70)$ = 2.53, p <.05; also see figure 4.3B). Lastly, the proportion of joint attention bouts terminated by infants did not differ by cohort($t_{6UK,6IND}(44.84)$ = -0.23, p = .82; $t_{6IND,9IND}(44.41)$ = 0.97, p = .34); $t_{6UK,9IND}(28.60)$ = 0.47, p = .64).

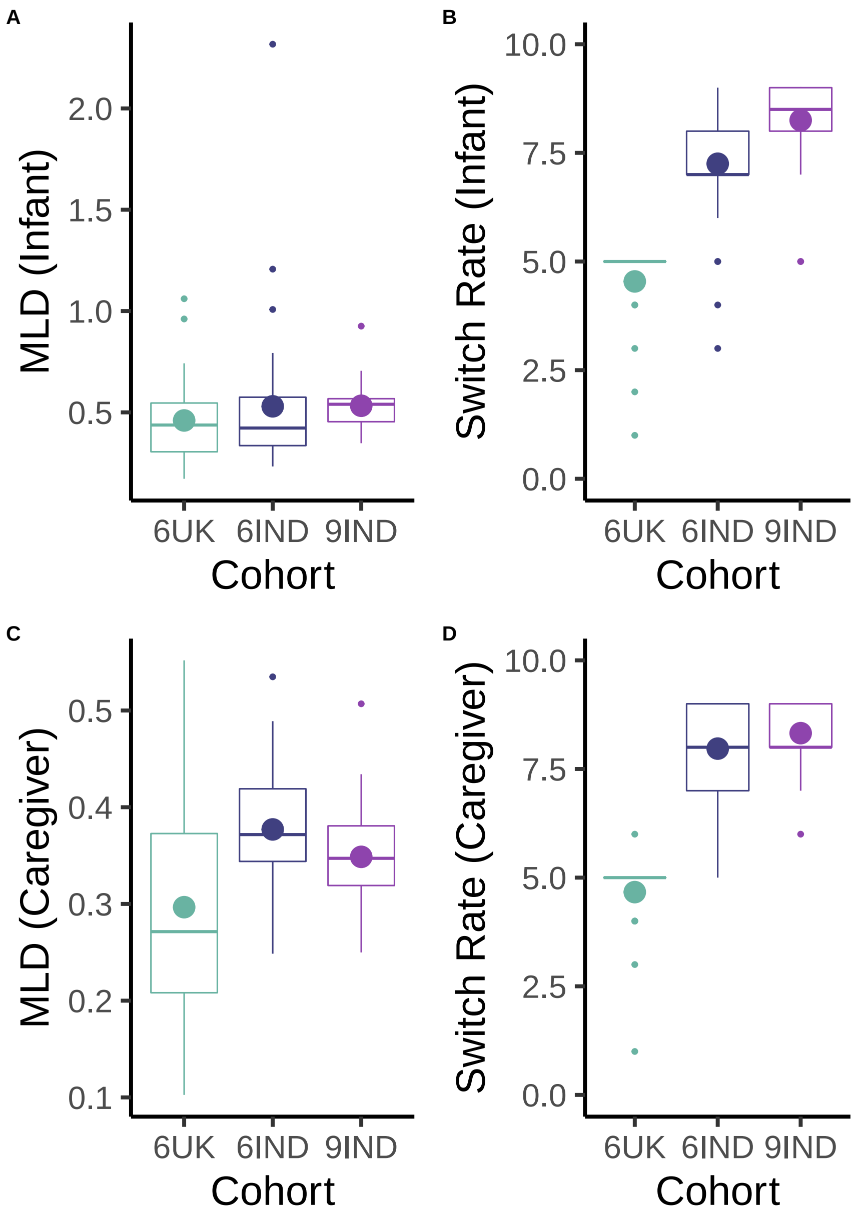


Figure 4.2: Overall MLD in seconds and switch rater per minute across cohort for caregivers and infants.

Table 4.3: Overall means and standard deviation for the total number of joint attention episodes, and the proportion of infant-led and terminated joint attention episodes in percentage.

	Total no. of JA Episodes		Infant-led JA episodes		Infant-terminated JA Episodes	
Cohort	M	SD	M	SD	M	SD
6UK	118	89.4	68.6	14.9	45.7	18.6
6IND	145	99.41	52.8	17.4	46.8	15.8
9IND	115	50.7	60.3	7.45	43.8	7.91

Note. JA stands for Joint Attention. M stands for Mean. SD stands for Standard Deviation. 6UK denotes dyads with 6-month-old infants from the UK, 6IND denotes dyads with 6-month-old infants from India and 9IND denotes dyads with 9-month-old infants from India.

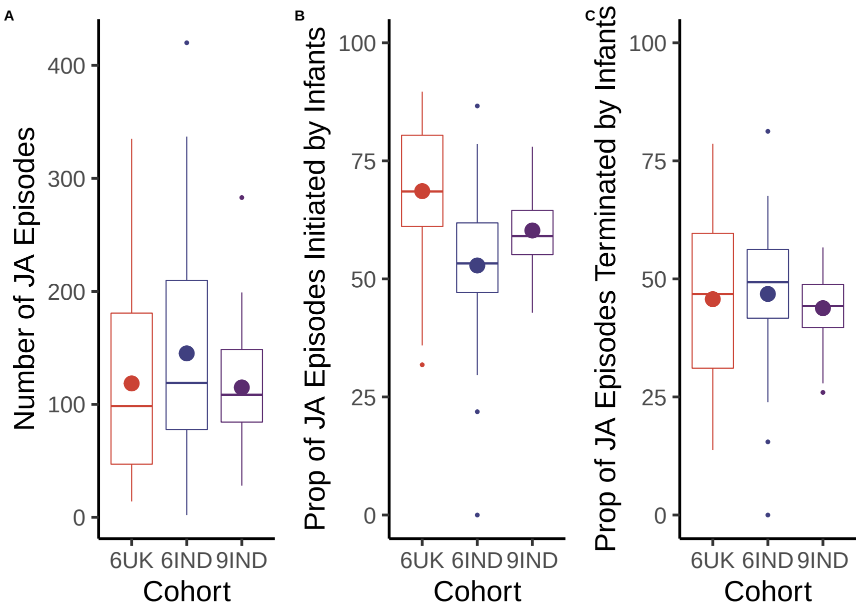


Figure 4.3: Box plots indicate A) the total number of joint attention episodes in each cohort, B) the proportion of joint attention episodes imitated by infants in each cohort (in percentage) and C) the proportion of joint attention episodes terminated by infants in each cohort (in percentage). Each large circle in the box plot indicates means and each horizontal line within the box plot indicates the median.

4.4.3 Infants' MLD for Faces and Toys

The descriptive statistics for the overall scores of Infant MLD towards toys and towards the caregivers' face split across contexts of attention (alone, infant-led joint attention, parent-led joint attention) are summarised in table 4.4 (also see Appendix O.1 for infants' median scores split by target, context and cohort). We discuss analyses of each context in turn below.

Table 4.4: Overall means, standard deviation, median and range for infants' MLD in various categories of caregiver-infant interaction.

	Infant			
	Mean	**SD**	**Median**	**Range**
Toys				
Infant MLD (Alone)	0.38	0.54	0.17	0.06-5.00
Infant MLD in Infant-led JA	1.56	1.83	0.93	0.06-14.33
Infant MLD in Caregiver-led JA	0.5	0.81	0.2	0.06-11.90
Caregivers' face				
Infant MLD (Alone)	0.31	0.49	0.13	0.06-5.00
Infant MLD in Infant-led JA	3.00	3.26	1.67	0.06-14.86
Infant MLD in Caregiver-led JA	0.63	1.19	0.23	0.06-14.86

Note. MLD stands for Mean Look Duration (in seconds). SD stands for Standard Deviation. JA stands for Joint Attention.

Infant MLD to toys when looking Alone. As shown in figure 4.4A, infants' MLD differed significantly across cohorts in the 'looking alone' context. When looking by themselves, 9-month-old infants from India showed the longest MLD directed at the toys and 6-month-old infants from India showed the shortest MLD when looking at the toys, with the UK sample falling between both. Differences between all groups were significant.

Infant MLD to toys in Infant-led joint attention episodes. Results showed that when infants initiated the joint attention episodes, 9-month-old infants in India showed significantly longer MLD to the toys compared to 6-month-old infants in both India and the UK. Among the 6-month-old infants, those from India showed significantly shorter MLD than their UK counterparts (see figure 4.4B). These findings mimic the results from the looking alone context, although overall, mean look durations were longer in the infant-led joint attention context.

Infant MLD to toys in Caregiver-led joint attention episode. With regard to infants' MLD in caregiver-led joint attention episodes, as shown in figure 4.4C, the 9-month-old infants once again displayed longer MLD than the two 6-month-old cohorts. Moreover, the 6-month-old infants in India displayed shorter MLDs than the UK group. Once again, these cohort differences are similar to the other contexts. Note, however, that MLDs overall were shortest in the 'looking alone' context, longer during parent-led joint attention episodes, and longer still during child-led joint attention episodes.

Infant MLD to caregivers' face when looking Alone. When looking at the caregiver's faces by themselves, the 6-month-old infants from India showed significantly longer MLD compared to the 6-month-old infants in the UK and the 9-month-old infants in India. Results indicated no significant difference in the MLD between the latter two cohorts (see figure 4.4D).

Infant MLD to caregivers' face in Infant-led joint attention episode. As shown in figure 4.4E, during episodes of joint attention initiated by the infants, 6-month-old infants from India spent a significantly longer time looking at their caregiver's faces compared to the two other groups. Moreover, 6-month-old infants from the UK spent significantly longer MLD focused on their caregivers' faces compared to the 9-month Indian infants.

Infant MLD to caregivers' faces in Caregiver-led joint attention episode. As in the previous analysis, when caregivers led the joint attention bouts, 6-month-old infants from India tended to look for longer at their caregivers' faces than infants in the other two cohorts (see figure 4.4F). In contrast to the analysis of infant-led joint attention, the MLD of 6-month-old infants from the UK was significantly shorter than that of 9-month-old infants in India.

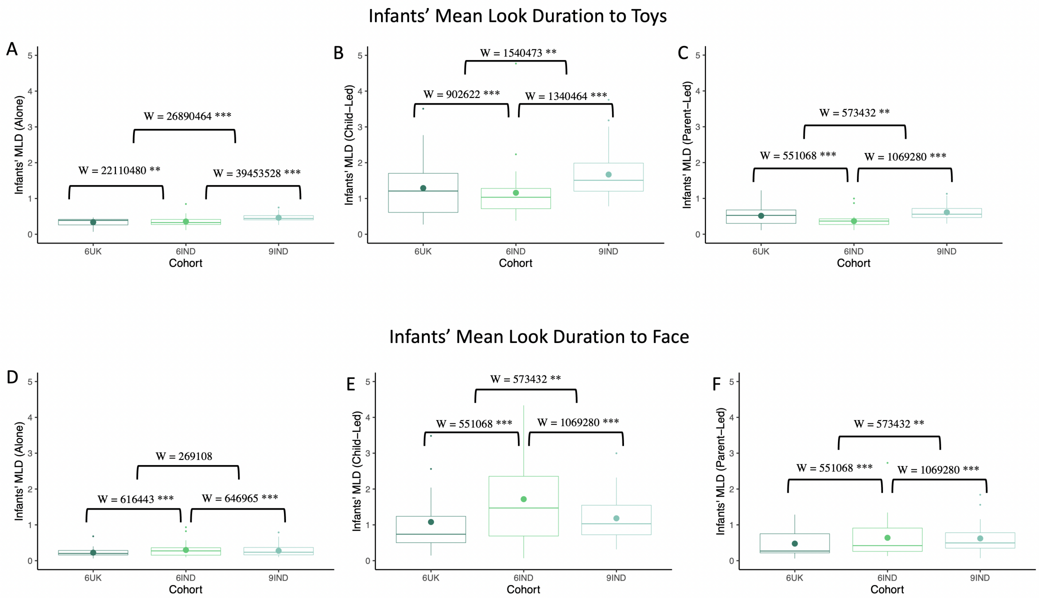


Figure 4.4: Comparing infants MLD using Mann-Whitney-U test across the three cohorts. The top row depicts infants' Mean Look Duration (MLD) to toys when A) looking alone B) looking with a caregiver in an infant-led joint attention episode C) looking with a caregiver in a caregiver-led joint attention episode. The bottom row depicts infants' Mean look Duration to caregivers' faces when D) looking alone E) looking at each others' faces in infant-led joint attention episodes F) looking at each others' faces in caregiver-led joint attention episodes. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001

Table 4.5: Overall means, standard deviation, median and range for caregivers MLD in various categories of caregiver-infant interaction.

Caregiver				
	Mean	**SD**	**Median**	**Range**
Toys				
Caregiver MLD (Alone)	0.31	0.43	0.17	0.60-10.97
Caregiver MLD in Infant-led JA	0.38	0.53	0.20	0.06-9.03
Caregiver MLD in Caregiver-led JA	1.04	1.29	0.60	0.06-11.24
Caregivers' face				
Caregiver MLD (Alone)	0.29	0.40	0.16	0.06-7.23
Caregiver MLD in Infant-led JA	0.38	0.51	0.20	0.06-6.37
Caregiver MLD in Caregiver-led JA	1.12	1.22	0.70	0.06-7.20

Note. MLD stands for Mean Look Duration (in seconds). SD stands for Standard Deviation. JA stands for Joint Attention.

4.4.4 Caregivers MLD for Faces and Toys

The overall caregivers' MLD for each target across the three categories (alone, infant-led, caregiver-led) is summarised in table 4.5 (also see Appendix O.2 for median across the target, categories and cohorts). We discuss each result below in turn.

Caregiver MLD to toys when looking Alone. Results showed that caregivers' of 6-month-old infants in the UK had significantly shorter MLD, compared to caregivers of Indian cohorts (see figure 4.5A). However, no significant difference in MLD to toys was found between the two Indian cohorts.

Caregiver MLD to toys in infant-led joint attention. With regards to caregivers' MLD to toys in joint attention episodes led by infants, as shown in figure 4.5B, caregivers in the UK cohort once again displayed significantly shorter MLD than those in both Indian cohorts. There was also no significant difference in the MLD of caregivers between the two Indian cohorts.

Caregiver MLD to toys in caregiver-led joint attention. Here as well, the MLD of caregivers' from the UK was significantly shorter than the MLD of 6- and 9-month caregivers in India. However, unlike in the previous results, caregivers of 6-month-old infants in India displayed longer MLD than those in the 9-

month cohort (see figure 4.5 C).

Caregiver MLD to infants' face Alone. No significant differences were found on caregivers' MLD toward their infants' faces outside of episodes of joint attention (see figure 4.5 D).

Caregiver MLD to infants' face in infant-led joint attention. In joint attention episodes initiated by infants, caregivers of the UK group showed significantly shorter MLD to their infants' faces compared to caregivers in the two Indian groups. Caregiver MLD to their infants' faces did not differ between the two Indian cohorts (see figure 4.5 E).

Caregiver MLD to infants' face in caregiver-led joint attention. As shown in figure 4.5 F, during caregiver-led joint attention, caregivers' from the UK displayed shorter MLD towards their infants' faces compared to the Indian caregivers. Once again, caregivers from the two Indian cohorts did not significantly differ in MLD.

4.5 Discussion

Chapter 4 has yielded a rich array of findings using the pipeline described in Chapter 3, which allows us to 1) validate the proposed pipeline with a larger data set and 2) explore how infants and their caregivers deploy their visual attention during a dyadic interaction playing with toys across different settings (the urban UK and rural India) and cohorts (dyads with 6-month-old infants from the UK and India, and dyads with 9-month old infants in India). Specifically, the current study aimed to address the similarities and differences across settings and cohorts between measures associated with visual cognition, such as mean look duration (MLD) and shift rate (SR). Overall, findings indicated consistent trends in the deployment of visual attention by

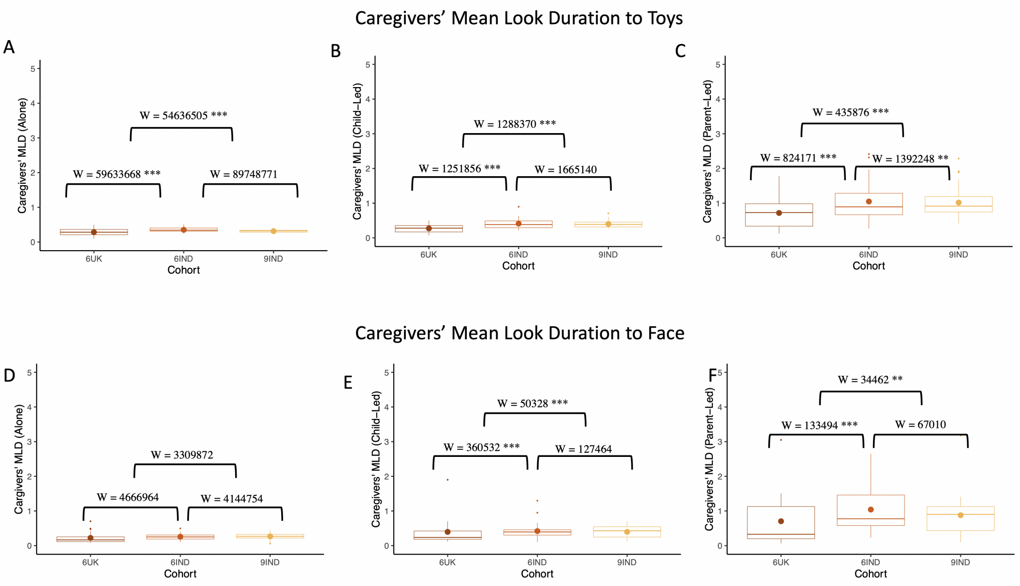


Figure 4.5: Comparing caregivers’ MLD using Mann-Whitney-U test across the three cohorts. The top row depicts Caregivers’ Mean Look Duration (MLD) to toys when A) looking alone, B) looking with caregiver in infant-led joint attention episode, and C) looking with caregiver in caregiver lead joint attention episode. The bottom row depicts infants’ Mean look Duration to caregivers’ faces when, D) looking alone, E) looking at each others’ faces in infant-led joint attention episodes, and F) looking at each others’ faces in caregiver-led joint attention episodes. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001

caregivers and their infants that reveal a complex interplay of influences on caregiver-infant interaction. Henceforth, I discuss findings through the lens of the eco-cultural model of parenting (Keller et al., 2005; Keller, 2007) and establish links to the visual cognition literature which enables us to connect this chapter to prior studies and sets the base for the final empirical chapter in this thesis (5).

In regards to the overall MLD, which indicates the sustained looking towards the target of fixation, we found no overall differences for infants across groups suggesting that they deployed their visual attention similarly. On the other hand, analyses on caregiver MLD showed no differences between the two Indian cohorts and revealed that the caregivers in the UK looked for a shorter amount of time towards targets compared to parents of 6-month-olds in India. Interestingly, this difference did not hold when comparing the caregivers from the UK to the caregivers of the older (9-month-old) Indian cohort. Together, the findings suggest a cultural element of similarity between the two groups of Indian caregivers as well a degree of difference when comparing the British caregivers to the Indian caregivers of younger infants but not those of the older ones.

When looking at the overall switch rate, caregivers and infants displayed a similar pattern. In the cases of both caregivers and infants, Indian cohorts switched more between targets of fixation than the UK cohort. When looking at differences between the Indian groups, older infants displayed larger switch rates but no difference was found for caregivers. Apparent cultural differences in the deployment of visual attention, however, need to be interpreted with caution. Typically, MLD and switch rate are interrelated such that longer MLD is related to fewer switches (or slower disengagement) (Colombo et al., 1991). In contrast in this study, we found that caregivers with higher

MLD also tended to have a higher switch rate.

A key point to consider here is that although both Indian groups completed the play session in the same controlled environment, British families carried out the study in their own homes. This is of particular importance because the measure of switch rate only takes into account switches between targets set in the TimeVP toolbox. Thus, if participants switched between the coded targets, the switch was computed. However, if participants shifted from a coded target to a non-coded target, no shift was recorded. Given that the rooms in which British families completed the study had many more objects and decorations in their surroundings (e.g., TV, sofa, photographs), which were not included in the TimeVP as targets of interest, multiple instances of switches of attention are likely to have been missed. Indeed, looking back at the raw data, visualised in figure 4.1, it suggested more consistent shifts to non-targets (i.e., white spaces in the figure) for the dyads in the UK than the Indian cohorts.

Differences between the two Indian cohorts, however, can be assessed with confidence as both cohorts were tested in the same, fully controlled environment. Here, we found higher switch rates for 9-month-old infants in India. This suggests there are changes in looking dynamics as the child grows. Such differences may stem from changes in parental engagement, changes in the capabilities of the infant, or a mix of both. Here it is important to note that no differences were found between the two Indian groups in terms of caregiver MLD or their switch rate. With the equivalent overall deployment of visual attention among caregivers, differences in switch rates among infants may be best explained by their own development and the age differences between the cohorts. Infant developmental influences would include a more developed visual cognitive system with an expected faster rate of switching for older infants (Rose et al., 2001, 2002). There are also likely to be improvements in motor development which enhances the exploratory capabilities and field of vision of older infants.

When exploring the joint attention episodes, which are considered of great importance and shape infant development (Carpenter et al., 1998; Mundy & Newell, 2007; Nelson et al., 2008), we found a similar number of total episodes of joint attention in all cohorts, showing consistency across samples, both in terms of cultural setting and differences in infants' age. Similarities extended to the proportion of joint attention bouts which were terminated by infants, but not to the episodes initiated by them. UK infants led a greater proportion of joint attention episodes during caregiver-infant interaction than the two Indian groups. Indian infants of 9 months of age also led a greater proportion of joint attention than their 6-month-old counterparts. Differences can be interpreted as revealing a cultural element of difference in parenting styles. That is, UK parents, are more likely to follow child-centred distal parenting styles and be less directive (Keller, 2007), thus providing greater opportunities for their infants to lead the interactions (Keller et al., 2010). In contrast, the proximal parenting style, which is more common in rural India, is more directive, thus resulting in parents leading a greater proportion of joint attention bouts (Keller et al., 2010). These apparent cultural differences are further nuanced by the infants' age. Perhaps, greater mobility and cognitive skills allow the 9-month infants in India to create further opportunities for interaction with their caregivers and engage more actively in leading joint attention.

Considering the context of attention (e.g., joint vs alone) further served to reveal that the overall MLD similarities found among infants of all groups should not be taken at face value. Indeed, findings indicated differences between the groups of infants when MLD was considered across targets and in the context of joint attention. When infants directed their attention at toys, 9-month-old infants consistently showed longer MLD than the other two groups regardless of whether they were attending on their own, during joint attention led by them, or joint attention led by their caregivers. Among the 6-month-old cohorts, MLD across all contexts was greater for the British infants. Given the sustained attention element of MLD, results indicate greater sustained attention, potentially due to greater interest and engagement among the oldest infants. The finding of sustained attention by older infants, compared to those led by younger infants, fits long-standing research on the developmental trajectory of infants in the context of coordinated attention in interactions with objects (Bakeman & Adamson, 1984).

Regarding attention toward the caregiver's face, the 6-month-old infants from India consistently had longer MLD regardless of whether they were attending to the target on their own or in joint attention episodes. The other two groups displayed similar MLD towards their caregiver's face when looking on their own but displayed interesting differences during bouts of joint attention. When joint attention was led by infants themselves, British infants sustained their attention on their caregivers' faces for longer than the Indian 9-month infants. However, when joint attention was led by caregivers, the opposite was found. That is, the group of older Indian infants sustained their look towards their caregiver's face for longer than the British infants.

Findings showing that 6-month-old infants in India displayed lower sustained attention towards toys but greater focus on faces, when compared to the other groups, may be partly explained by how they were positioned as well as by their mobility (Fausey et al., 2016; Soska & Adolph, 2014). Previous research has indicated that the infant-perspective field of view changes with age-related development due to changes in their motor abilities, skills, as well as caretaking needs (Fausey et al., 2016; Jayaraman et al., 2017). Not only do younger infants have more input from faces in their daily lives (Fausey et al., 2016), but they also have fewer opportunities to manually and visually explore objects when in supine and prone positions (Soska & Adolph, 2014). In the case of our findings, both infants' positioning and motor capabilities would interact with parental practices, as the younger Indian infants tended to be placed on their backs, directly in front of their parents. Furthermore, caregivers directed the interaction, moving the toys in and out of the infants' range of vision. Less directive and more child-centred practices in the UK would shift the world that is accessible to infants. Last, for the older Indian infants, their greater motor and cognitive development (hence, capabilities) would allow them to explore their surroundings in a more proactive way, despite directive caregiving.

Accounting for cultural differences in parenting styles also serves to explain differences in MLD directed at caregiver faces between the 9-month Indian infants and the UK cohort. The 9-month-old Indian infants have grown in a setting of more directive parenting, with interactions being more adult-centred whereas the British infants are accustomed to more child-centred interactions (Lancy, 2014). Each group, thus, engages with their parents differently, with Indian infants being more responsive to parental-led attention and British infants being more proactive in generating joint attention and sustaining their focus.

Regarding the caregiver's deployment of visual attention, UK caregivers displayed shorter MLD towards toys across all contexts of attention and towards the face of their infant in both joint attention contexts. When looking at the Indian groups, differences in MLD only appeared when the targets were toys and during joint attention episodes led by the caregiver. That is, caregivers of the 9-month cohort displayed shorter MLD directed at toys when they led the joint attention episodes but deployed their visual attention comparatively to the caregivers of the younger cohort in all other cases. Caregivers in all groups displayed comparable MLD towards their infant's face when outside of joint attention. Given that caregivers should be expected to be familiar with their infants' faces, this similarity may be an indication of short bouts of attention that allow caregivers to check up on the infant.

Once again, differences between groups can be seen in the context of parenting styles as well as the caregivers' understanding of their infant's capabilities and their expectations of what the interaction should be like. In the UK, where child-centred parenting fosters the view of children as a subject with a degree of agency since early infancy, parents are more active and engaged in play interactions (Keller et al., 2010). In India, infants are seen to a larger degree as passive, which shapes parental engagement in the interaction. In the older Indian cohort, in which infants are more developed and have greater mobility, caregivers may be seeing them to a lower degree as passive and shift their behavioural patterns. Indeed, the fact that differences between the Indian caregivers only appeared in caregiver-led joint attention episodes suggests that said differences do not result from varying capabilities between the two Indian groups and are more likely to stem from shifts in the engagement of caregivers.

Imagine a caregiver shaking a rattle. The caregiver may shift their focus to the infant to see their response (is the child interested?), and back to the rattle more frequently, thus sustaining MLD for less time, when there is an expectation that the infant may reach out to the object or move to grab it. The same caregiver may sustain their attention on the rattle for longer (even if only slightly) if the child is not expected to have the capability to approach, reach out, or grab the rattle. In a way, the parent is more on guard when the infant can move than when they are seen as passive. Given the directive/proximal parenting approach in India, it is logical for this shift in the parental deployment of attention to be particularly noticeable in joint attention episodes led by the caregiver. Critically, as shown by our findings on infants' attention towards their caregiver's face when caregivers initiate joint attention, their infants also tend to respond to the parental direction.

Although the goal of this chapter is not to make claims about Visual Working Memory (VWM), the trend of shorter MLD among caregivers in the UK invites considering differences beyond parenting style, including potential differences in the speed of visual information processing of the caregivers. Indeed, it is likely that differences in the deployment of visual attention among caregivers are indicative of differences in VWM, with greater VWM allowing British caregivers to process information faster and reducing the need for sustained attention on the targets of interest. It is important, however, not to oversimplify results as solely displaying differences in VWM between the British and Indian cohorts given the differences found in the deployment of attention of Indian caregivers of infants of different ages. That is, differences between caregivers' characteristics, including VWM, are likely to interact with cultural values and parenting styles, leading to a range of differences in parental engagement with their infants.

Interestingly, in the case of infants, those with greater VWM (the older cohort; see Chapter 2 Study 2) sustained their attention towards toys for longer than the other two groups. In contrast, they displayed shorter MLD towards their caregiver's face in episodes of caregiver-led joint attention. An important nuance to consider here entails the difference in VWM capabilities between parents and infants. That is, whereas processing the visual information related to familiar faces and simple objects is easy and within the expected capabilities of parents, the same should not be expected for infants.

Infants will have had massive exposure to their caregivers' faces. As the facial visual information is processed on multiple occasions it becomes more familiar and easier to process. Thus, shorter sustained attention toward the caregiver's face may be indicative of better VWM. On the other hand, the infants' exposure to the specific toys used in this research was much lower. Moreover, the toys can be in motion and infants need to process the visual information from multiple perspectives. To make things more complicated for the infants, they may be holding the objects at some times but not others, with additional sensory information (e.g., tactile) providing further input. Thus, infants are just learning about the novel objects and longer sustained looking is expected to support VWM. Thus, infants with longer MLDs might learn more about the objects and, more generally, show better VWM abilities.

Two key limitations in this study should be considered in order to inform future research. The first entails the different degrees of control across play sessions addressed when interpreting differences in switch rates. The first point of lack of control results from British dyads completing the play session in their own homes, each with their own environments and distractors. This lack of control is inevitable when carrying out research in naturalistic settings and should be seen as a methodological characteristic rather than a key limitation. However, the fact that all Indian dyads completed the play session in a highly controlled environment with almost no distractors or additional objects is a key limitation. On one hand, it reduces the ecological validity sought from naturalistic research. On the other, the fact that some groups completed the task in the artificial laboratory setting creates a confound that hinders the interpretation of findings when comparing the different groups.

The methodology followed in this study was an inevitable consequence of the limited access to the home environment of participants in rural India. While it is tempting to state that future research should go into homes in rural India, such an endeavour goes hand-in-hand with technological demands. Given the lack of constant electricity in infra-structurally underdeveloped contexts, the battery life of available portable equipment does not allow for collecting data from more than two participants per day. In cases in which travelling from the data collection points to locations with access to electricity is time-consuming, moving back and forth to recharge the equipment becomes costly in time and money. For our own research, the limited time spent in India in data collection rounds, limitations in available technology, lack of electricity while in the field, and the distance and time required for travelling to recharge equipment resulted in the decision to sacrifice the ecological validity and invite families to the make-shift lab for the play sessions. Bringing the UK families to the lab as well required a further sacrifice of ecological validity and was disregarded due to the value of validating the pipeline in a fully naturalistic setting. For these reasons, the interpretation of analyses has focused on MLD and switch rates have been interpreted with caution. Future technological advances may enable future research that reduces limitations of this type.

The current chapter has served to understand the deployment of visual attention in a naturalistic setting across socio-cultural groups. Results indicate consistent patterns in caregiver and infants' visual exploratory behaviour such that infants from the UK tend to initiate a higher proportion of joint attention compared to the other groups. This fits the eco-cultural model of parenting wherein caregivers in western middle-class families tend to take a child-centric approach (Keller, 2017; Keller et al., 2005). In comparison, infants in India, particularly the younger cohort, tend to lead fewer joint attention episodes. However, both age groups in India tended to sustain their attention to caregivers' faces in caregiver-led joint attention episodes, thus responding to their caregiver's attention, compared to the infants in the UK who showed sustained attention in a self-led bout of joint attention. In terms of visual information processing, consistent with research on visual cognition, 6-month-infants in India have shown longer MLDs across contexts suggesting slower information processing speed compared to the 9-month-old infant in India. Similarly, caregivers in the UK have shown shorter MLD compared to the caregivers in India indicating faster visual information processing. To, further, our understanding of how these real-life measures of visual cognition relate to the visual cognition measured in the laboratory-based task, the next chapter focuses on addressing the connection between caregivers and their infants' visual exploratory behaviour in the naturalistic setting to their VWM in a standardised lab-based task.

Chapter 5

Exploring the relationship between measures of visual cognition from lab and lounge

5.1 Introduction

Throughout this thesis, studies have served to provide insights into the Visual Working Memory (VWM) of infants and their caregivers. Through laboratory tasks, Studies 1 and 2 (Chapter 2) linked the VWM of Indian mothers to their socio-emotional context and the VWM abilities of their infants. Study 3 (Chapter 4) focused on the deployment of visual cognition by caregivers and infants from the UK and India in the context of naturalistic dyadic interactions during play. Here, we used available portable technology and developed a machine learning pipeline to measure key aspects of dyadic interactions across cultures including quantifying joint attention episodes. In Chapter 5, we use these same measures to link the visual cognition of infants and caregivers in the lounge to measures of VWM in the lab (as seen in chapter 2).

Specifically, we explore whether and how infants' and caregivers' visual experience when interacting in the real world relates to their performance in the standardised visual cognition task measured in the laboratory.

Visual cognition is important for gathering information from our surroundings, allowing us to attend to, and hold information about, objects and people. It also is a key tool that enables infants to learn from their surroundings. Contemporary knowledge of infants' visual cognition and its development has risen from two key traditions, one focused on researching development in the lab and the other on the lounge. Laboratory-based research, involving highly controlled environments and stimuli, has highlighted the importance of measuring individual differences in visual cognition (Jankowski et al., 2001). For instance, visual cognition assessed in infancy has been found to be predictive of long-term outcomes such as academic achievement and IQ at 11 years of age (Rose et al., 2012). More specifically, within the umbrella of visual cognition, the development of VWM has been argued to play a pivotal role in infants' cognitive development, with increases in VWM underlying improvements in speed of processing among other cognitive processes. Indeed, as noted by Spencer (2020), the development of VWM is conceptualised not just as an influence but as foundational in cognitive development.

In empirical research, laboratory measures of VWM have been found to be related to learning and academic outcomes (S. F. Ahmed, Tang, Waters, & Davis-Kean, 2019; Allen, Higgins, & Adams, 2019). For instance, the development of mathematical skills is frequently associated with working memory (Bull & Scerif, 2001; Bull et al., 2008; Holmes & Adams, 2006; Holmes, Adams, & Hamilton, 2008). A recent study by Fanari, Meloni, and Massidda (2019) conducted a longitudinal study with young children to assess visuospatial working memory and early mathematics skills measured using standardised mathematics measures for primary school children. A battery of laboratory-based tasks was administered to test working memory in the school setting. The results showed that performance on the visuospatial working memory test at 6 years was predictive of mathematics achievements at 7-8 years of age.

Despite the wealth of laboratory research and its many contributions, a key challenge is to relate lab-based measures to how visual cognition is deployed in the real world. In the real-world, infants encounter more complex visual stimuli than what has been used in the highly controlled lab setting (e.g., the screen-based stimuli used in a preferential-looking task). Similarly, the real-world is much more complex, with more opportunities for distraction. Of course, as with all other laboratory research, this limitation can be raised to challenge the degree to which findings can be applied in real-world settings. These gaps are covered by the second tradition of interest to this thesis, which has focused on researching the visual exploratory behaviours and deployment of attention of infants in naturalistic settings.

As with lab-based approaches, naturalistic developmental research conceptualises visual cognition in general, and VWM more specifically, as essential to infants' socio-cognitive development (C. Yu et al., 2019). Infant development, however, is observed and studied in the "messy" real world, with its richer stimuli and, perhaps more importantly, the influence of social partners on infants' experiences of the real-world. Social partners influence how infants deploy their attention, from what infants prioritise attending to, to the time spent holding their attention. This, in turn, shapes the development of VWM and its multiple outcomes (e.g., word learning). For instance, C. Yu et al. (2019) have argued that VWM may be critical to the visual information processing of an object, thus enabling the infant to sustain their attention towards it, which is essential to understanding the object and mapping it in order to learn its name. Joint attention episodes with a caregiver serve to prolong the sustained attention on the target and thus facilitate the understanding of the object and learning of the new word (Vlach & Johnson, 2013; C. Yu et al., 2019). More specifically, caregivers who "follow-in" on their infants' attention and sustain it have been found to promote early world learning (C. Yu & Smith, 2016; C. Yu et al., 2019). Thus, it is not surprising that being able to coordinate attention with a social partner on an object of interest is considered to be an important developmental milestone (Moore & Dunham, 1995; Scaife & Bruner, 1975).

Although naturalistic research has advantages, it also has limitations. For many years, research in naturalistic settings has used hand-held or tripod-based cameras to record caregiver-infant interactions, with researchers manually coding the dyads' visual exploratory behaviour as perceived from a third-person perspective. However, recent technological advances (e.g., mobile head-mounted cameras and eye-trackers) have revealed a different "visual" story. Work by Yoshida and Smith (2008) has shown that infants' and their caregivers' visual field are different compared to what is perceived and coded from a third-person perspective (also see L. B. Smith et al., 2011). Infants' visual scenes consist of one or two objects, which are placed close to their heads and faces, thus taking up most of their visual field. Additionally, infants hardly look at their parents' faces and instead mostly look at hands (their own or of their caregivers) manipulating objects. This has been found when infants play with toys together with their caregivers regardless of positioning, such as when playing on the floor (Franchak, Kretch, & Adolph, 2018) and on the table (C. Yu & Smith, 2013).

The problems stemming from limitations such as this one have been reduced by technological advances, but the higher ecological validity of naturalistic observations still comes at a cost. The very fact that stimuli in the real world are more complex and messy, means that the exercise of understanding the development of visual cognition becomes equally messy. Indeed, complex environments and reduced control preclude establishing clear-cut or straightforward explanations on the individual's experience of the world or the effects (individual or interactive) of the colour, shape, movement, or accompanying sound of different stimuli on the infants experience, their deployment of attention, and the outcomes these have for the development of VWM.

As well as having their own limitations, each tradition makes its own assumptions. As addressed in the discussion of findings from Study 2, a key assumption of laboratory research is that its findings have real-world implications. Similarly, research focused on naturalistic settings and the observation of interactions assumes that observed behaviour and the deployment of attention are representative of how VWM works. Both traditions have strong theoretical and empirical support but their connection remains unaddressed. Exploring the connections between both, and integrating the knowledge constructed through them, allows us to reduce their limitations and establish clearer findings with less dependency on the traditional assumptions. That is, rather than assuming the real-world outcomes relate to lab-based VWM measures, the integration of methodologies allows us to test the actual associations between the lab and lounge (Dahl, 2017).

To our knowledge, only one research has addressed the connections between measures of visual cognition in laboratory screen-based tasks and semi-naturalistic free play tasks. More specifically, the study by Wass (2014) focused on peak look duration towards novel stimuli across the two settings. The author found relationships between the screen-based scores (including static and non-static stimuli), but not between the screen and semi-naturalistic paradigms. Moreover, an exploratory factor analysis revealed two underlying factors, with three out of four screen-based measures (again including static and dynamic images) loading on the first factor and the free play scores loading on the second. Interestingly, a single screen-based measure (using static images of complex scenes) cross-loaded on both factors. Its primary load was on the second factor (together with semi-naturalistic tasks) but it loaded in the opposite direction to the more naturalistic measures. The authors considered a range of influences which could explain findings, ranging from the field of vision of infants (with screens taking up a greater proportion of it) to the contrast in luminescence (between the targets and their surroundings as well as between the edge of the screen and its surroundings), additional auditory stimuli present in the screen-based tasks, and autonomic arousal linked to changes in luminance in the screen tasks. By focusing on different measures and using different tasks, our research expands on that of Wass (2014) and builds upon our knowledge on whether measures of visual cognition are equivalent, or at least related, across contexts.

Drawing from laboratory and naturalistic research we can identify good indicators of VWM across methodological settings. Measures of mean look duration (MLD) as well switch rates (SR) in the lab and lounge, and the change preference scores as measured by VWM laboratory tasks all provide robust indices of aspects of VWM. Additionally, episodes of joint attention in real-world interactions are critical given that they not only serve to indicate the development of infant VWM but also foster it. More specifically, factors such as initiating and responding to bouts of joint attention have been shown to have unique associations with word learning at different stages of infancy (Mundy et al., 2007) and significant impairment in initiating joint attention, in clinical research, is indicative of autism (Mundy et al., 2003; Mundy & Newell, 2007). Thus, we draw from the above measures, as obtained in previous studies in this Thesis, to develop a more comprehensive picture of the links between the lab and lounge.

Henceforth, Study 4 connects the laboratory (also see Chapter 2, Studies 1 & 2) and naturalistic (see Chapter 4, Study 3) measures of MLD and switch rate of both infants and caregivers. We further explore the links between naturalistic measures of MLD and of infant/caregiver-led joint attention and laboratory VWM-PL scores for both infants and caregivers. Last, we tackle the unresolved inverse relationship between laboratory VWM-PL scores of mothers and their infants by testing our mismatch hypothesis in connection to naturalistic measures.

5.2 Study 4: Examining the relationship between measures of VWM from the lounge to the lab.

5.3 Method

5.3.1 Participants

The VWM-PL task, as noted in chapter 2, was completed by 228 infants in Year 1 and 186 in Year 2. In terms of dyadic interaction, 69 dyads contributed to the data. Of the 69 dyads, 4 dyads did not have SES data and one infant did not have VWM data. Therefore, the analyses looking at the relationship between infants' measures of visual cognition in the lab and lounge consist of 64 participants. Similarly, 125 caregivers completed the VWM-PL task, however, when combined with the dyadic data, final analyses for caregivers' visual cog-

Table 5.1: Means and standard deviation of Infant and Caregivers' age and SES Score for lounge-to-lab analyses

	N = 64		N = 46
	6-month Infants	9-month Infants	Caregiver (Y2)
N	28	36	-
Infant Gender	16 Female	12 Females	-
$M_{Age\ (Y1/Y2)}$	6.04/ 18.12	8.91/ 20.86	25.54
$SD_{Age\ (Y1/Y2)}$	0.36/ 0.56	0.34/ 0.55	3.95
$M_{SES\ Score}$	10.38	10.07	10.26
$SD_{SES\ Score}$	3.85	3.74	3.66

Note. M denotes Means; SD denotes Standard Deviation; Y1 denotes age at Year 1 of testing; Y2 denotes age at Year 2 of testing. Missing one 6-month-old and five 9-month-old infants' demographic data at Year 2.

Table 5.2: Means and standard deviation of Infant and Caregivers' age and SES Score for Mismatch analyses.

	N = 37			
	Dyads 6 months		Dyads 9 months	
N	16		19	
Infant Gender	9 Females		7 Females	
$M_{SES\ Score}$	10.94		9.88	
$SD_{SES\ Score}$	3.7		4.08	
	Infant	Caregiver	Infant	Caregiver
$M_{Age\ (Y1/Y2)}$	5.99/18.01	26.25	8.91/20.94	25.94
$SD_{Age\ (Y1/Y2)}$	0.35/0.54	5.09	0.37/0.55	2.7

Note. M denotes Means; SD denotes Standard Deviation; Y1 denotes age at Year 1 of testing; Y2 denotes age at Year 2 of testing. Missing one 6-month-old and five 9-month-old infants' demographic data at Year 2.

nition measures from the lab to the lounge included 46 caregivers. Means and standard deviations of infants and caregivers' age and socio-economic scores are indicated in Table 5.1. Lastly, of the 120 caregivers and infants that completed the VWM-PL task, and 69 dyads who contributed to the dyadic data, the final 'mismatch' analyses consisted of 37 dyads with VWM-PL measures from both infant and caregiver as well as dyadic data (see table 5.2 for participant age and SES Score summary).

5.3.2 Materials and Procedure

The materials, apparatus, and procedure for the laboratory data are the same as in Studies 1 and 2 in Chapter 2 (for caregivers and infants respectively). The materials and the procedure for the dyadic interaction data are the same as in Chapters 3 and 4.

5.3.3 Measure Construction and Methods of Analysis

Measures from the Lounge MLD measures from the dyadic interaction were created for the infants and for the caregivers. As in the previous study, measures were taken from three different types of episodes: 1) when looking at a target (toy, face) outside of joint attention episodes (henceforth, "MLD Alone"), 2) during infant-led joint attention episodes, and 3) during caregiver-led joint attention episodes.

Infant switch rate (SR) measures from the dyadic interaction included the number of times per minute that the infant switched their attention from one target to another (e.g., from face to toy, between toys).

The proportion of infant-led joint attention episodes was calculated by dividing the number of infant-led joint attention episodes by the total number of joint attention episodes. For the caregiver analyses, the proportion of joint attention measure was re-computed to display the proportion of joint attention episodes led by caregivers (1 - the proportion of infant-led joint attention).

Measures from the Lab Infant MLD in the VWM-PL is the length of each fixation within the target box containing the array of squares. Given that the score of MLD of infants was skewed, we carried out a logarithmic transformation (see Appendix P.1 for a histogram of non-transformed and transformed scores).

Infant Switch-rate consisted of the number of times per second that the infant switched between the change and no-change sides of the display.

VWM-PL scores indicated the proportion of time spent looking at the changing side of the preferential-looking task when the first look was directed at the no-change side. Two VWM-PL measures were computed, one for the infants and the other for the caregivers.

A mismatch Ratio was calculated in order to address the mismatch hypothesis proposed to explain the inverse relationship between maternal and infant VWM found in Study 2. The measure was computed by dividing the infants' VWM-PL score by their caregivers' VWM-PL score, thus summarising the relationship in a single value. High ratio values indicated an infant with high VWM scores paired with a caregiver with low VWM scores. By contrast, low ratio values indicated infants with low VWM scores paired with caregivers with high VWM scores. It must be noted that even caregivers with the lowest VWM capabilities would still surpass the VWM of their infants (even those with the highest VWM).

Demographic Measures Demographic measures of mothers' age, infants' age, and SES scores of the family were taken from Chapter 2 studies 1 and 2 (also see table 5.1 and table 5.2).

5.3.3.1 Analytic Strategy

In order to explore the relationship between VWM as displayed in the lab and lounge, we divided analyses across four stages. Given that we lacked data for all participants when accounting for all tasks, we deemed the sample size insufficient for running complex models with multiple variables.

The four stages were as follows: 1) understanding the relationship between infants' measures of MLD and switch rate between the lab and lounge; 2) understanding the association between infants' naturalistic measures of MLD and proportion of infant-led joint-attention episodes and their lab VWM-PL performance; 3) exploring the same relationships as in (2) but focusing on caregiver measures, and 4) examining the correlations between measures of caregiver-infant interaction from the real-world setting and the mismatch between caregiver and infant's VWM noted in Chapter 2. At each stage, the maximum number of participants possible was included in the analyses.

In the first stage, we start by creating two baseline models, each addressing infant lab measures of MLD and SR as outcomes. Predictors in both models were the same as those included in the infants' VWM base model from Study 2, that is, year of study (1 or 2), load (low, medium, or high), SES Score (centred), and age cohort (cohort 1 = 6 and 18 months, cohort 2 = 9 and 21 months). Year and age cohorts were difference-coded, the load was input as a factor, and the model includes the main effects only. No caregiver-infant interaction measures were added at this point. We then added measures from the dyadic interaction as predictors to the baseline model. Given that we had three measures of MLD, three separate linear mixed-effect models were computed, one including MLD Alone, another including MLD in infant-led joint attention, and the last with MLD in caregiver-led joint attention. In addition to main effects of the baseline model and MLD (centred) measures, each model included a two-way interaction between MLD and Load.

The same approach was taken for the model on laboratory SR, with SR from the dyadic interaction data being added as a predictor to the base model. Given that there is only one switch rate measure, a single model was computed. Once again, a two-way interaction between task load and SR was included.

In Stage 2, four models were computed in order to address the relationship between infants' dyadic measures of MLD and infant-led episodes of joint attention and their VWM-PL scores. Each model is built on the infant VWM baseline model from study 2 in chapter 2 (computed as in Stage 1). As in the previous stage, three separate models were created to address the association between MLD in the lounge and VWM in the lab, one for each MLD measure. Thus, models 1, 2, and 3 include a main effect of the relevant measure of infants' MLD as well as a two-way interaction effect between MLD and Load (from the VWM-PL task). Model 4 shifted from the focus on MLD and instead included the proportion of infant-led joint attention episodes as a main effect to the baseline VWM model. Once again, the two-way interaction between the proportion of infant-led joint attention and load was included as a predictor.

In stage 3, we repeated the same steps from Stage 2 but focusing on the caregiver measures. Thus, we ran four mixed-effect models on the VWM-PL scores of caregivers, building upon the caregivers' VWM baseline model from Study 1 in Chapter 2 (i.e., main effects of Mother's age, centred SES Scores, and Load, as well as the two-way interaction between load and mothers' age). As in Stage 2, three models were computed with the caregiver MLD measures from the lounge as predictors. Within each model, the two-way interaction between the relevant MLD measure and load was included. Model 4 built on the baseline by including the proportion of caregiver-led joint attention episodes as a main effect and interacting with load.

Each model in stages 1, 2, and 3 included a random intercept for each participant to allow for individual differences. All models were assessed for fit based on a Q-Q plot of the residuals and using the R package DHARMa (Hartig, 2021).

The fourth and final stage addressed the inverse relationship between the

VWM-PL performance of caregivers and their infants found in chapter 2. We used the mismatch ratio computed with the infants' and caregivers' VWM scores and explored possible correlations with the dyadic interaction measures of visual cognition of both caregivers and their infants.

5.4 Results

5.4.1 Infants Mean Look Duration and Switch Rate in lounge and lab

Figure 5.1 shows correlations between MLD measures from VWM-PL task to A) MLD Alone, B) MLD in infant-led JA, C) MLD in caregiver-led JA, and, D) Switch rate measure between VWM-PL task and SR measure in the lounge. Furthermore, a scatterplot showing the correlations between lab measures, i.e., MLD, SR, and VWM-PL scores can be found in Figure Q.1 in Appendix.

MLD Results of the MLD baseline model indicated a main effect of load and a main effect of year on infants' MLD as measured in the VWM-PL task (see table 5.3 for Wald Chi-squared tests), with infants' MLD increasing as the task became more difficult. Moreover, as shown in figure 5.2 infants' MLD significantly increased from Year 1 to Year 2. When considering the effects of the year of testing on MLD, it must be noted that task difficulty was increased and more items needed to be consolidated in working memory in the Year 2 VWM-PL task (set sizes 1, 2, 3 versus 2, 4, 6).

Next, we added the infants' MLD measures (alone, infant-led joint attention, caregiver-led joint attention) from the caregiver-infant interaction to the 'base' MLD model. Consistent with our base model, results from all three models indicated a main effect of load and a main effect of year. However, no

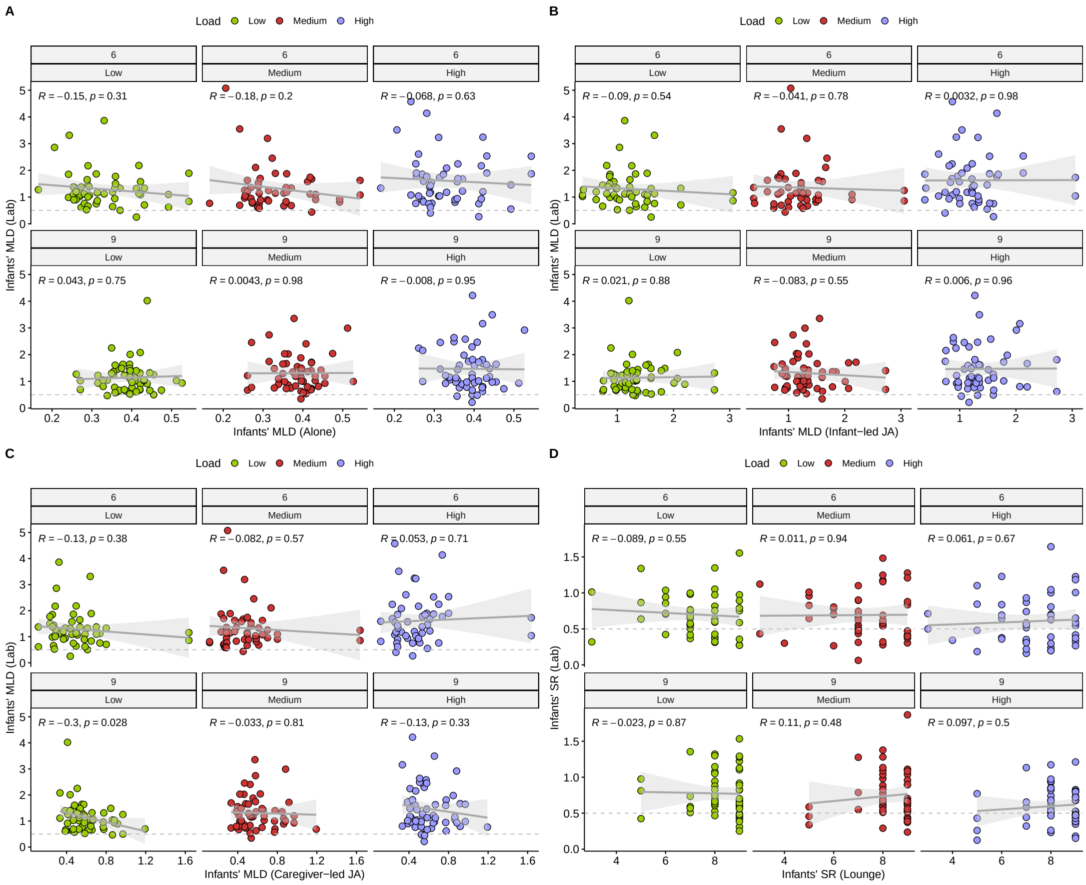


Figure 5.1: Scatterplot for MLD lab and MLD lounge measures (A-C) and SR lab and SR lounge measures (D) across low, medium, and high load conditions (from VWM-PL task). 6 indicates 6-month-old and 9 indicates 9-month-old infants' age cohort.

Table 5.3: Infant MLD baseline regression results with Load, Year, SES Score and Age as criterion

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	9.20	1	<.01 **
Load	102.74	2	<.001 ***
Year_s	274.27	1	<.001 ***
SESScore_c	2.03	1	.15
Age_s	2.62	1	.11

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

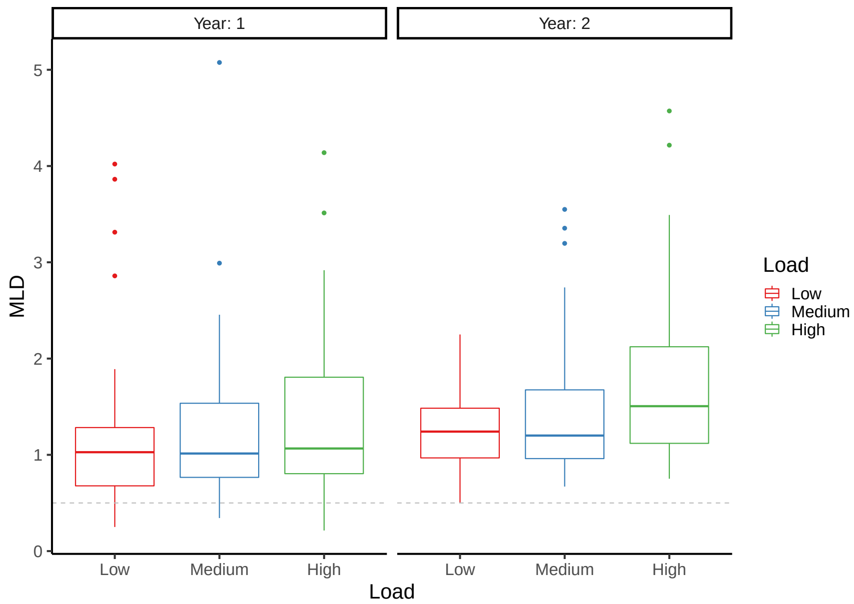


Figure 5.2: Boxplots show the regression results for infants' Mean Look Duration in VWM-PL task (baseline model).

main effects were found for any of the three MLD scores taken from the dyadic interaction. No two-way interaction effects between infants' MLD measures and load on the MLD lab score were found (see Appendix R.1 for the table with MLD alone, Appendix R.2 for the table with MLD in infant-led joint-attention, and Appendix R.3 for the table with MLD in caregiver-led joint-attention).

Switch Rate As shown in figure 5.3, results of the baseline model once again indicated a main effect of load on infants' SR in the VWM-PL task such that the SR decreased as the task difficulty increased (also see table 5.4 for Wald chi-square tests). There were no other main effects.

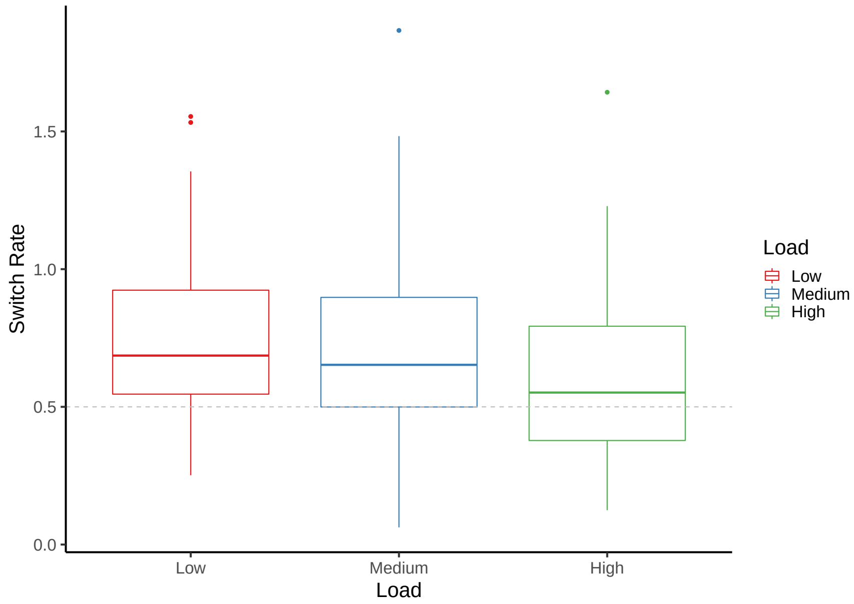


Figure 5.3: Switch Rate in VWM-PL task.

Table 5.4: Infant Switch Rate baseline regression results with Load, Year, SES Score and Age as criterion

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	318.73	1	<.001 ***
Load	14.75	2	<.001 ***
Year_s	0.15	1	.70
SESScore_c	1.01	1	.31
Age_s	0.69	1	.40

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

In terms of the relationship between infants' SR during caregiver-infant interaction as predictors of SR in the VWM-PL task, results revealed no significant main or interaction effects of infants' SR from the dyadic setting (see table in Appendix S.1). The main effect of load on infants' SR as measured in the lab remained significant.

5.4.2 Infants' measures of visual cognition from the lounge related to their VWM score in the lab

Figure 5.4 shows a scatterplot with correlations between Infants' VWM-PL scores in lab to A) MLD Alone, B) MLD in infant-led JA, C) MLD in caregiver-led JA, and, D) proportion of infant-led JA episodes measures in the lounge setting.

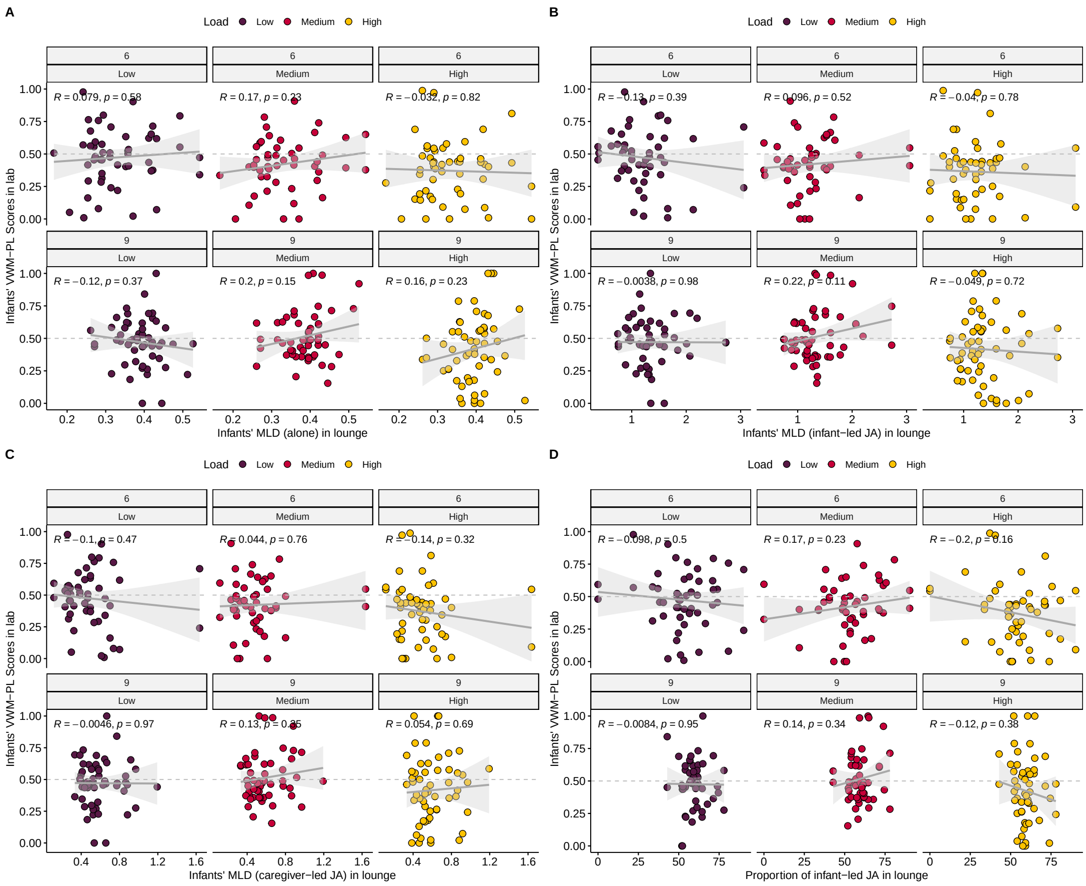


Figure 5.4: Scatterplot for MLD lab and MLD lounge measures (A-C) and SR lab and SR lounge measures (D) across low, medium, and high load conditions (from VWM-PL task). 6 indicates 6-month-old and 9 indicates 9-month-old infants' age cohort.

MLD lounge to VWM lab As the first step in Stage 2 of analysis, we computed the VWM baseline model with the smaller sub-sample from this study.

Results in table 5.5 show that the main effects of load and age found in Study 2 remained significant in our subsample. The following three models further included the infants' MLD while looking at a target (face and toys) as predictors of their VWM performance on the lab-based task. Regarding the association between infants' MLD (alone) and VWM-PL, results from Model 1 indicated that the main effect of load remained significant after controlling for the MLD (alone) measure. Moreover, a two-way interaction was found between load and infants' MLD (alone) as measured in the caregiver-infant interaction (see table 5.6). No main effect of MLD (alone) was found. As can be seen in figure 5.5, the longer the MLD when infants looked at targets outside of joint attention, the better their performance on the VWM-PL lab-based tasks in middle and high load conditions.

In the case of infants' MLD in infant and caregiver-led joint attention episodes on their VWM performance, results from Model 2 (for table with infant-led joint attention see Appendix T.1) and Model 3 (for table with caregiver-led joint attention see Appendix T.2) revealed no significant main or interaction effects of the infants' MLD on the VWM-PL scores. Once again, consistent with the baseline model, the main effect of load remained significant across both models.

Table 5.5: Results for infants' VWM baseline model.

Variable	χ^2	Df	p-value
(Intercept)	839.54	1	<.001 ***
Load	101.99	2	<.001 ***
Year_s	3.46	1	.06 .
SESScore_c	0.23	1	.63
Age_s	3.85	1	.05 *

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

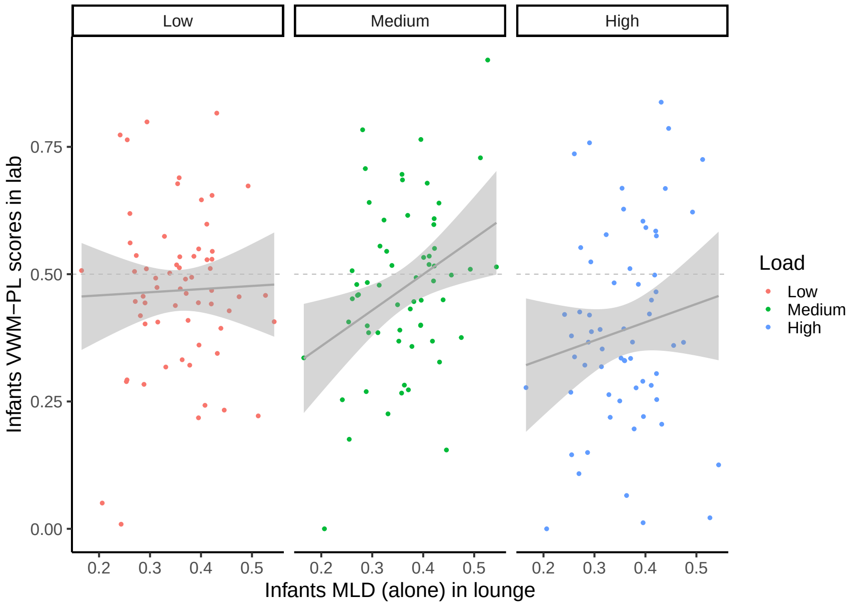


Figure 5.5: Relationship between infants' visual working memory score assessed in the laboratory setting and the MLD when looking at the target alone in the naturalistic setting.

Table 5.6: Results for infants' VWM scores from VMW-PL task with infants MLD (alone) in dyadic interaction as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	836.64	1	<.001 ***
Load	102.39	2	<.001***
Year_s	3.48	1	.06 .
SESScore_c	0.23	1	.63
Age_s	3.84	1	.05 .
ChAloneDur_mean_c	2.48	1	.12
Load:ChAloneDur_mean_c	6.33	2	.04 *

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Proportion of infant-led JA in the lounge on VWM from the lab Within Stage 2 of our analysis, we further explored whether the proportion of infant-led joint attention episodes predicted their performance in the VWM-PL task. For the difference between the 6- and 9-month-old infant-led proportion of JA see figure U.1 in Appendix.

We found no main effect of the proportion of infant-led joint attention episodes on their VWM-PL performance. Once again, the main effect of load remained significant in the model. Moreover, there was a significant two-way interaction effect of load and proportion of infant-led joint attention on the VWM outcome (see table 5.7 for Wald-chi square test). As shown in figure 5.6, in the medium load condition, the higher the proportion of infant-led joint attention episodes the higher the working memory scores. This pattern was inverted in the high load condition, although there was quite a bit of variability in the VWM scores in this condition.

Table 5.7: Results for infants' VWM scores from VMW-PL task with the proportion of infant-led joint attention episodes as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	348.86	1	<.001 ***
Load	9.34	2	<.01 **
Year_s	0.02	1	.88
SESScore_c	0.03	1	.86
Age_s	3.71	1	.05 .
prop_ch_led_JA_c	2.99	1	.08 .
Load:prop_ch_led_JA_c	6.57	2	<.05 *

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

5.4.3 Caregivers' measures of visual cognition in caregiver-infant interaction and their VWM score in the lab

Figure 5.4 shows a scatterplot with correlations between Catregivers' VWM-PL scores in the lab to A) MLD Alone, B) MLD in infant-led JA, C) MLD in caregiver-led JA, and, D) proportion of caregiver-led JA episodes measures in the lounge setting.

MLD lounge to VWM lab Results for the caregiver baseline model in Stage 3 were consistent with findings from Chapter 2 Study 1. Analyses revealed a main effect of load on the VWM-PL performance of caregivers in our subsample (see Table 5.8). However, the two-way interaction between load and

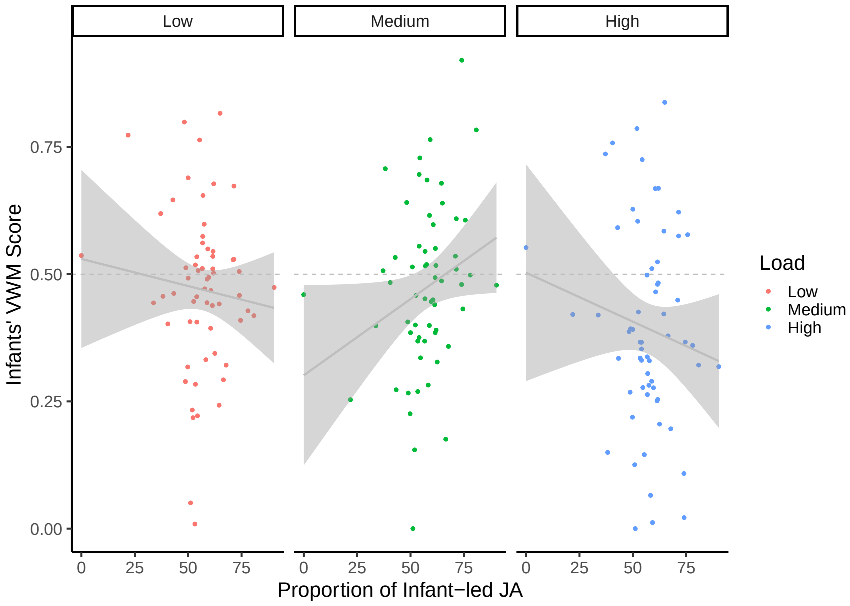


Figure 5.6: Relationship between infants' visual working memory score assessed in the laboratory setting and the proportion of infant-led joint attention episodes in the naturalistic setting.

mothers' age on the VWM-PL performance failed to reach significance. The shift in results may stem from the reduced sample size and resulting lack of power. Indeed, the pattern of relationships in the interaction remained stable. The visual representation of results in Figure V.1 (in Appendix) indicates a sustained trend of scores on VWM performance decreasing the older the participants were, but only in the medium and high load conditions.

Table 5.8: Results from caregivers VWM baseline model.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
Intercept	237.09	1	<.001 ***
Load	10.97	2	<.01 **
SESScore_s	0.26	1	.61
MotherAge_s	0.01	1	.93
Load:MotherAge_s	3.70	2	.16

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

We then computed the three models with the caregivers' MLD scores while looking at a target (face and toys) during interaction with their infants on the

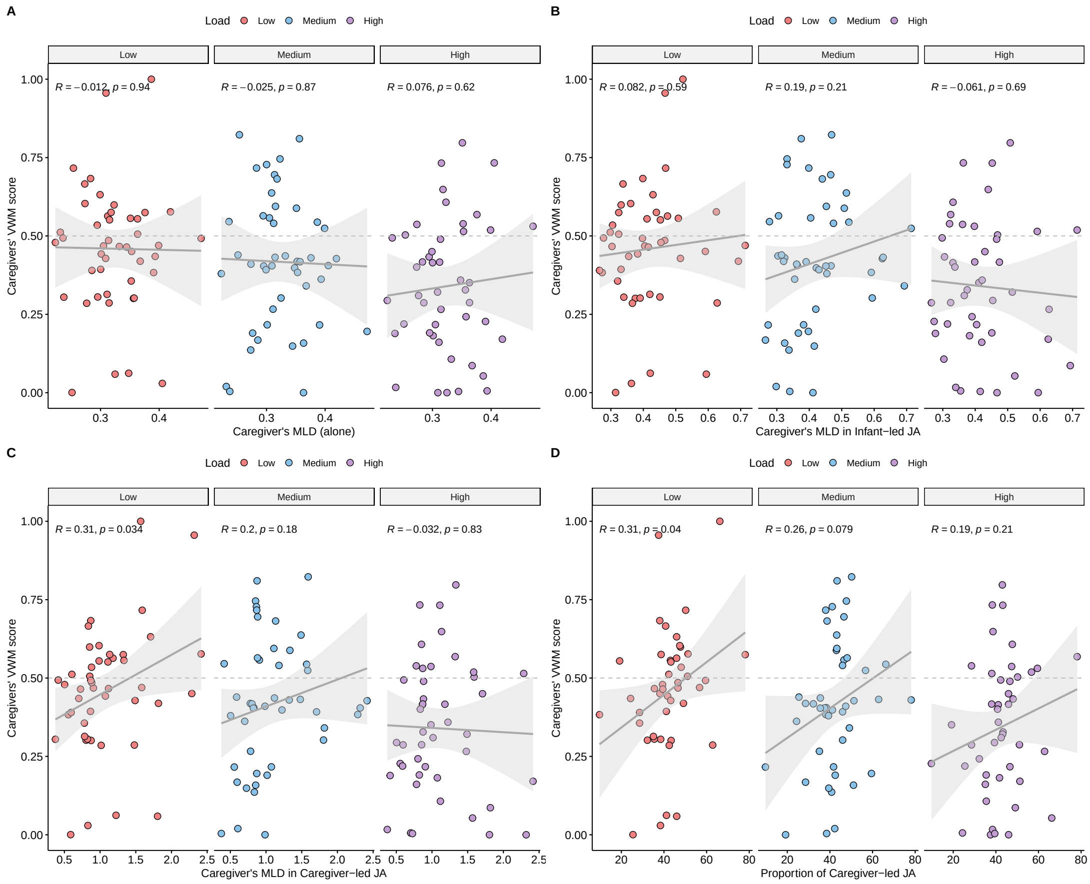


Figure 5.7: Scatterplot with correlations between Caregivers' VWM-PL scores from lab and A) MLD alone, B) MLD in infant-led JA, C) MLD in caregiver-led JA, and D) proportion of caregiver-led joint attention episodes in the lounge.

VWM lab score. Model 1 (MLD Alone) revealed no significant main or interaction effects for MLD but the main effect of load remained significant (see table W.1 in Appendix). Model 2 (MLD in infant-led joint attention; see table W.2 in Appendix) and Model 3 (MLD caregiver-led joint attention; see table W.3 in Appendix) also did not indicate main or interaction effect of caregivers' MLD from dyadic interaction on their VWM-PL scores. However, the main effect of load remained significant in both the latter models.

The model with the proportion of caregiver-led joint attention as a predictor of VWM performance indicated a main effect of the proportion of caregiver-led joint attention such that the greater the proportion of joint attention led by caregivers the better their working memory (see table 5.9 for the Wald chi-squared tests). In other words, caregivers with poorer working memory tended to follow in on their infants' attention rather than initiating joint attention episodes (see figure 5.8A). This led us to run post-hoc analyses to explore the association between caregivers terminating joint attention episodes and their VWM performance. As shown in figure 5.8B, the higher the working memory of caregivers the fewer the joint attention episodes they terminated (see table 5.10 for the Wald chi-square test). Thus, caregivers with high VWM scores tended to initiate caregiver-led joint attention episodes more often and they seemed to maintain these episodes long enough that the episodes were more likely to be terminated by the infant.

5.4.4 Testing the Mismatch Hypothesis

To explore the inverse relationship between the VWM scores of caregivers and those of their infants' revealed in chapter 2, we conducted a one-tail correlation with the single mismatch ratio score from the VWM task with variables from the caregiver-infants interaction and multiple MLD measures from

Table 5.9: Results for caregivers' VWM scores from VMW-PL task with the proportion of caregiver-led joint attention episodes as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	4.17	1	<.05 *
Load	9.94	2	<.01 **
SESScore_s	0	1	.99
MotherAge_s	1.29	1	.26
prop_par_led_JA	6.97	1	<.01 **
Load:MotherAge_s	3.52	2	.17

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table 5.10: Results for caregivers VWM scores from VMW-PL task with the proportion of caregiver-terminated joint attention episodes as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	32.47	1	<.001 ***
Load	9.94	2	<.01 **
SESScore_s	0.04	1	.84
MotherAge_s	1.07	1	.30
prop_par_term_JA	4.99	1	<.05 *
Load:MotherAge_s	3.52	2	.17

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

dyadic interaction. A higher mismatch ratio indicates infants with higher VWM and caregivers with lower VWM scores. A lower ratio indicates caregivers with higher VWM scores and infants with lower VWM scores.

As can be seen in Table 5.11 results indicated a significant correlation between the VWM ratio and infants' MLD alone from the lounge as well as the proportion of infant-terminated joint attention episodes (inverse of the proportion of joint attention terminated by caregivers). That is, the higher the VWM ratio (i.e., the greater the infants' VWM score), the longer the infants' MLD outside of joint attention episode and the lower the proportion of joint attention episodes terminated by the infant. Figure 5.9 shows the directions of the two correlations. Trends in findings fit the interpretation from Study 2 that a mismatch between parental and infant VWM would be expected to have implications in real-life interactions, which, in turn, may hinder the develop-

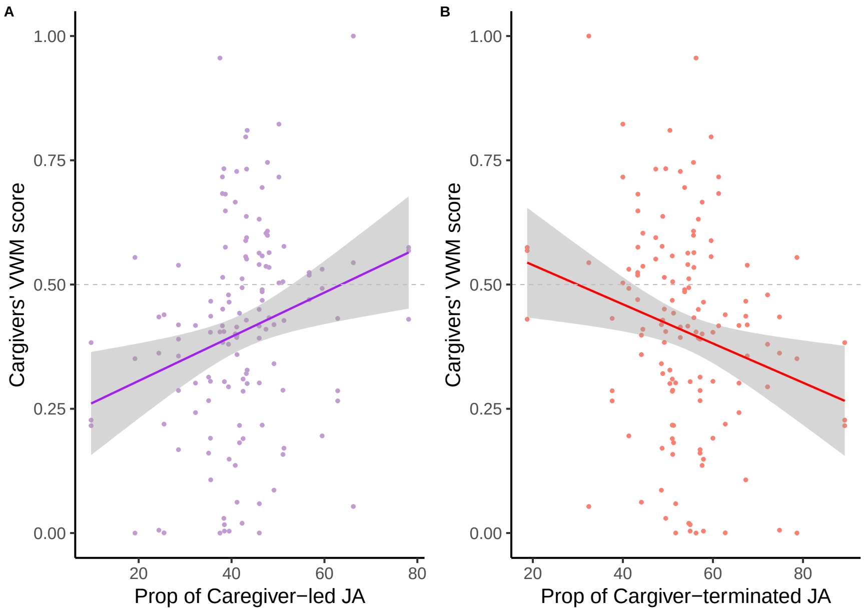


Figure 5.8: Relationship between caregivers' visual working memory score assessed in the laboratory setting and the proportion of A) caregiver-led joint attention episodes and B) joint attention episodes terminated by the caregiver in the naturalistic setting.

ment of infant VWM. No other significant correlations were found between the VWM mismatch ratio and any of the variables of interest.

Table 5.11: Results for correlation between the ratio of caregiver and infants VWM scores and measures from caregiver-infant interaction.

	VWM Ratio
M (SD)	1.35 (0.63)
Infant	
MLD Alone	.27*
Infant-led MLD	.24
Caregiver-led MLD	.17
Caregiver	
MLD Alone	-.12
Infant-led MLD	-.07
Caregiver-led MLD	-.16
Both	
Prop of Infant-led JA episode	.22
Prop of Infant terminated JA episode	-.27*

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

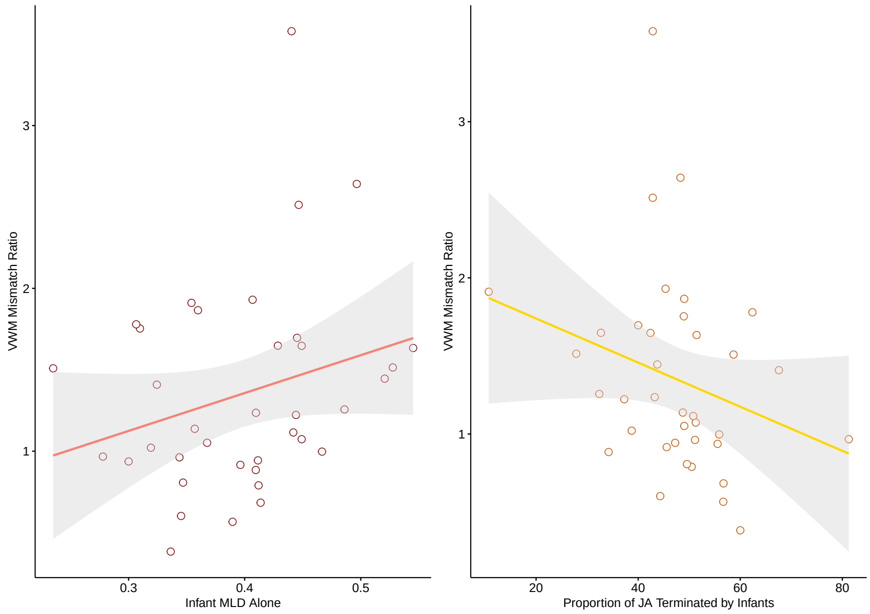


Figure 5.9: Correlation between caregiver-infant VWM ratio, and A) infants' mean look duration when looking at a target alone (left panel), B) joint attention terminated by caregivers (right panel). Note, that a higher VWM ratio indicates infants with higher VWM scores and caregivers with lower VWM scores. A lower VWM ratio indicates caregivers with higher VWM scores and infants with lower VWM scores.

5.5 Discussion

Prior research has noted the importance of studying visual cognition in the laboratory as well as in the naturalistic settings (Rose et al., 2012; L. B. Smith et al., 2015), however, the connection between the two remains largely unaddressed. The present study sought to relate measures of visual cognition from the lounge and the lab from participants in rural India. We used measures from the VWM-PL lab-based task (as elaborated in Chapter 2) and from naturalistic dyadic interactions (as elaborated in Chapter 3 and 4) to explore the links between the laboratory and naturalistic research traditions. Our results show that there are overlaps between specific measures of VWM across settings for both caregivers and their infants. Additionally, our results indicated the relationship between the mismatch ratio (as proposed in Study 2) and dyadic measures, thus addressing the larger theme of this thesis. Henceforth, we discuss our findings in connection to previous studies in this thesis and the broader literature.

Findings of the current study, for infants, indicated that MLD alone measure (i.e., sustained focus on a target outside of joint attention episodes) from the dyadic interaction was related to the VWM assessed in the lab, such that infants with longer MLD alone tended to have better VWM. However, we did not find a relationship between the VWM of infants, and their caregivers, and their measures of MLD in episodes of joint attention. An explanation is that the joint attention episodes involve more variability than the solo deployment of attention. That is, being interactive, joint attention is affected by a broader range of behavioural input (e.g., pointing, vocalisation). Joint attention also draws from the influence of the cognitive capabilities and other characteristics of both the infant and the caregiver. With bouts of joint attention being more context-sensitive and related to the social partner, the measures of solo deployment of attention may be more indicative of individual differences in cognitive capabilities.

From the (Perone & Spencer, 2013a) work we know that when an infant model is presented with a preferential-looking task display, it looks at a visual stimulus. In order to shift to another stimulus, the infant model needs to consolidate the information from the first target in the working memory and then release the fixation. A key element of this cycle entails the time it takes to consolidate the stimulus in the working memory. Infants with robust working memory should consolidate the features of the stimuli and move to another target faster than infants with less robust working memory. However, the speed of this cycle is affected by how salient the item is.

More recent work by Bhat, Spencer, and Samuelson (2022) involved the construction of a WOLVES model (Word-Object Learning via Visual Exploration in Space), which is embedded within DFT and uses visual and word feature binding fields to understand the contribution of attention and memory in cross-situational word learning. Based on the WOLVES model, we see that in addition of the infant model cognitive capacities, factors such as the labelling of a visual stimulus can also influence the duration of the look. When a visual stimulus (e.g. toy in the real world) is labelled, top-down attention comes into play and keeps the fixation on the same stimulus, thus extending the look duration (also see C. Yu et al., 2019; C. Yu & Smith, 2016). Thus, when caregivers label an object (or point towards it, shake it, etc) during real-world interactions, the infant's fixation on that object is extended (also see C. Yu et al., 2019; C. Yu & Smith, 2016). Similarly, if a feature of the visual stimulus is salient (e.g., colour), it keeps the attention to the object for longer, making it harder to release fixation as there is more activation to overcome. The consequence in this context may be that the working memory peak is sustained for longer, thus leaving a strong memory trace.

Taken together, the literature supports the interpretation that variations in caregiver input (e.g. labelling and pointing) may extend the measure of MLD in joint attention beyond only the consolidation of objects in the working memory and the infants' individual characteristics. Thus, our results indicating a significant association between infants' MLD outside joint attention and VWM performance may result from this specific measure of MLD allowing the individual differences in infants' visual cognition to stand out and the cross-setting relationships to appear.

Drawing from the literature and findings from Study 3 (Chapter 4) further provides a plausible explanation for the positive direction in the association between MLD (alone) in the dyadic interaction and VWM-PL. As shown in Study 3 (Chapter 4), younger Indian infants displayed longer MLD towards faces but shorter MLD towards toys than their older counterparts. This was interpreted in terms of motor and cognitive development, with the greater VWM of older infants (and enhanced mobility) enabling their sustained attentional engagement with less familiar targets which were seen from multiple perspectives (i.e., the toys).

Infants tend to direct their attention more towards objects and hands than towards faces (Yoshida & Smith, 2008). Moreover, as they age, their focus towards objects increases (Bakeman & Adamson, 1984). Thus, general measures of attentional focus from an infant (i.e., measures including all targets), should draw from more data stemming from fixations on toys than data from fixations on faces. This suggests that our general MLD measures should be more indicative of attention towards toys than towards faces. However, this imbalance may not apply to the same degree during joint attention.

Imagine a caregiver and infant engaging in joint attention. Even if the focus of the interaction is placed on a toy (e.g., a rattle), the attention of both infant and caregiver will shift between the object (or hands manipulating it) and each others' faces (which enables them to receive feedback from their playing partner). Thus, even if a greater focus of attention remains on the toy, a measure of MLD will have relatively greater input from attention directed at faces compared to when looking at the target alone. In the context of our research and findings, it may be that the MLD measure outside of joint attention was more indicative of attention towards toys and thus allowed revealing a positive association in line with findings from Study 3.

In our research, the exploratory approach and sample size limitations resulted in the computation of multiple models. Further splitting the MLD measures for toys and faces would have inflated the number of analyses even further, essentially doubling the models required for addressing the dyadic measures of MLD. Thus, the decision was made to not separate the MLD scores. Using our findings and interpretations for reference, future research on lab-to-lounge connections should consider more targeted analyses. More specifically, future research with larger samples and a narrower focus may find more nuanced results on the links between MLD during joint attention bouts and VWM-PL if distinguishing between the attention targets (face versus toys).

It is important to note that the association between MLD from the dyadic interaction and the VWM-PL task was not replicated for the caregivers of infants. The lack of a relationship may be down to a lack of variability in the parental MLD outside joint attention (see Figure 4.5, in Chapter 4). It is worth noting that for caregivers, who have a more developed VWM, processing visual information from their infants' faces and from simple toys should be a relatively easy thing to do (unlike for infants). If all caregivers meet the required threshold to process the information comfortably, variations in MLD would be small and not related to the variation in VWM.

Another key finding in this study stems from results showing that greater VWM capabilities as measured by the laboratory preferential-looking task (VWM-PL) were positively associated with leading a greater proportion of joint attention bouts. That is, the greater the proportion of joint attention episodes led by infants, the higher their VWM-PL scores, and the greater the proportion of joint attention led by caregivers, the higher their own VWM-PL scores. Critically, as in the case of the relationship between infants' dyadic MLD alone and their VWM-PL scores, the relationship between the proportion of infant-led joint attention and infant VWM-PL was strongest in the medium load condition of the VWM-PL task. This suggests that the medium load condition provides us with the "sweet spot" where the task is hard enough to pull out the individual differences but not so hard so as to lead to large variability. This resonates with interpretations of results on the relationship between maternal age and VWM-PL in Chapter 2.

It is important to note that the empirical associations found in this study preclude establishing causality. Indeed, if looking at the data from infants, it may be that having their caregivers follow in on their attention served to foster the development of their VWM. Similarly, findings could be interpreted as higher VWM capacities enabling infants to create more opportunities to lead in their visual exploration of the world. Previous findings from research on interventions (Landry et al., 2006, 2008), however, invite considerations of causality. Indeed, our findings fit research showing that changes in parental behaviour (i.e., when trained to follow in on their infants) foster the development of visual cognitive capabilities (see Landry et al., 2006; Perone & Spencer, 2013a). Thus, although causality cannot be established from our own research, our findings do support the causal link established in other research.

Another point worth considering here is that the proportion of joint attention led by infants does not solely depend on them and their capabilities. It also depends on their caregivers following their lead. Thus, framing a causal relationship in the opposite direction (from VWM-PL as measured in the lab to the real-world deployment of attention) requires considering not only whether the infants have the capability to lead but also whether the caregivers follow in on their attention. This suggests that the infants' VWM capabilities in isolation are insufficient to explain the variation in the dyadic interaction. This point is of particular importance within our sample and population of interest.

Interestingly, we also found that caregivers' with high VWM-PL tended to terminate a lesser proportion of joint attention. Together, findings suggest that caregivers with high VWM in our sample tend to lead joint attention to a greater degree and also tend to sustain it, enabling the infant to terminate the episode. This pattern may also reflect interactions in which infants are redirected by their caregivers to join in on their attention but then shift back to their own targets of interest. Imagine an infant playing with a toy elephant, the caregiver directs their attention away from the elephant and towards a rattle. Even if the infant looks to the rattle and the caregiver sustains that focus of attention, the infant may end the joint attention episode and turn back to the elephant. It is possible that this indicates a redirecting pattern among the caregiver with high VWM.

Findings on the VWM mismatch ratio also revealed a correlation between it and the proportion of joint attention episodes terminated by the caregivers (inverse of the proportion of joint attention terminated by the infants). That is, the greater the mismatch between the VWM of caregivers and of infants (i.e., lower ratio score), the lower the proportion of joint attention that was terminated by the caregiver instead of by the infant. Correlations also revealed a link between the VWM mismatch and infants' MLD outside of joint attention. That is, the greater the mismatch (i.e., the lower the ratio) the shorter the infants' MLD duration outside of joint attention. This means that when there is a mismatch between maternal and infant VWM (high maternal VWM and low infant VWM), the infants focus for a shorter amount of time on the objects on which they focused on their own. This finding matches an interpretation of parental input redirecting the infants' own attentional focus. Building on our interpretations, overall findings suggest that the pattern of redirection of attention by caregivers, which is then terminated by the infants, is best reflected when the caregiver has higher VWM capacity and the infant lower VWM.

In Chapter 2, we argued that maternal VWM would be expected to influence their interactions with their children. Given that VWM can be fostered through social interactions (C. Yu & Smith, 2016), it logically followed that shifts in interactions resulting from maternal VWM capacity would in turn shape the development of their infants. Moreover, we proposed that a mismatch between the VWM capacities of caregivers and their infants would particularly be of relevance to understanding their interactions. The trends of findings in the present study (i.e., positive relationships between VWM-PL and leading joint attention episodes for both caregivers and infants), together with findings on termination of joint attention in connection to maternal VWM as well as to VWM mismatch provide support for our prior speculation. Future research should build upon these findings by testing mediational chains of causality longitudinally, from the VWM mismatch, to the social interaction, and from this to the changes in infant VWM over time. Moreover, future research in western settings, where parenting styles are more child-centred, should test whether this picture holds true across socio-cultural settings.

In addition to relating the laboratory measure of VWM-PL to a range of measures of visual deployment of attention during dyadic interaction, we further extended the findings of this thesis by exploring the links between concrete measures (MLD and switch rate) across methodological paradigms. When considering the use of a variety of measures that conceptually address the same construct, scores should be expected to relate to one another. Thus, intuitively, findings using one measure could be expected to be transferable to other research paradigms. Within our study, the simplest and most direct expectation when relating comparable measures between the preferential looking task and dyadic interaction settings (i.e., MLD from lab and MLD lounge, and SR across the two settings) would be for them to be associated. However, our study repeatedly found that this was not the case, with neither MLD nor SR being associated between research approaches.

An important point to consider here is that the settings, stimuli, and activities carried out are very different in a VWM-PL task as carried out in a laboratory and in a real-life interaction. The positioning of the infant and their visual field, the method of presentation of stimuli (in a screen versus tangible objects) and its novelty, as well as the additional sensory stimulation and the input from the social partner, in this case, the caregiver, are all different across settings. Said differences should be expected to reduce the overlap between measures and may be sufficient to completely separate them. Indeed, although focusing on different measures, our findings match those of Wass (2014), which revealed the lack of relationship between measures of peak look duration when stimuli were presented on a screen versus when the setting was semi-naturalistic. Thus, we build upon the literature by extending the insights gained on peak look duration to other measures of visual attention, with our findings corroborating that intuitively equivalent measures can be, in fact, substantially different.

Findings from the current study provide valuable insights into understanding the relationship between measures of visual cognition taken from dyadic interactions and VWM-PL tasks. For infants, the results indicated an overlap between MLD outside of joint attention in the lounge and VWM performance in the lab. The longer that infants sustained their attention to a target on their own during dyadic interaction, the better their VWM-PL performance. We suspect that infants' MLD alone during the dyadic interaction captures the individual differences that the MLDs in other settings do not, due to the latter being more context-sensitive. In terms of caregiver and infant-led joint attention episodes, both show a positive association between their VWM performance and the proportion of self-led joint attention episodes. This links into the inverse VWM relationship found in Chapter 2, suggesting that caregivers with high VWM capabilities tend to lead more joint attention episodes and have infants with lower VWM capabilities. Conversely, infants with high VWM capabilities are not only leading more joint attention episodes but also have caregivers following in on their attention. The latter fits the intervention and modelling work by Landry et al. (2006) and Perone and Spencer (2013a). Having gone full circle in this thesis, the next chapter integrates the information and findings from different chapters to develop a holistic picture of caregiver-infants interaction and its connection to VWM development in rural India.

Chapter 6

General Discussion

6.1 Summary and Integration of Findings

Throughout this thesis, we have aimed to develop a holistic picture of infant Visual Working Memory (VWM) development that addresses its connections to parental influences and characteristics, as well as to the broader sociocultural setting of our population of interest in rural India. We further developed a machine learning pipeline that enabled us to extract key measures of visual cognition from recordings of caregiver-infant interactions. The complementary research in this project, thus, served to tackle the key goal of relating the measures of visual cognition from the real world to those obtained from the standardised laboratory-based preferential looking task. In the present chapter, we review the work carried out across the previous chapters and integrate findings in order to address our contributions to the field. We further consider challenges and limitations within this project as well as the standing of the researcher in relation to the research carried out and the checks and balances resulting from collaborative input. We finish our discussion by considering the real-life implications and applications of the insights obtained through this project. Throughout the different sections of the discussion, we consider future research that can be built upon our work in order to further enrich a holistic understanding of infant visual cognition.

In Chapter 1, we reviewed the literature from two traditions of research on visual cognition – in the lab and lounge. The rich literature available consistently presents VWM as essential to human daily life (Spencer, 2020), connecting its development, and individual differences in it, to multiple outcomes, including reading, mathematics and general intelligence (Bull et al., 2008; Holmes & Adams, 2006; Gathercole et al., 2004). Although most research on visual cognition has been conducted in WEIRD settings, studies on caregiver-infant interaction with cultural elements (Bard et al., 2021; Keller et al., 2005; Keller, 2007), together with the existing literature connecting real-life interactions to the deployment of attention (C. Yu & Smith, 2016; C. Yu et al., 2019), provided insights into what infant development is like in non-western settings and, thus, what development in our population of interest might look like. Cautious towards the dangers of uncritically applying a biased body of knowledge and research from the west to LMIC settings (see Morelli et al., 2018; Weber, Diop, Gillespie, Ratsifandrihamanana, & Darmstadt, 2021), we decided to use the literature to guide our questions but take an exploratory stance in the following chapters.

Given the ambitious scope of our project and its many components, we broke down our ultimate goal into separate steps. Thus, Chapter 2 set the empirical foundation for the lab-based portion of the thesis. In Study 1, we developed a comprehensive picture of our understudied population and explored the potential influences of maternal socio-emotional and demographic characteristics on their VWM as measured by the laboratory-based preferential looking task (VWM-PL). Study 2 used the same laboratory paradigm to assess the VWM of infants and its relationships to the maternal data. Findings of both studies fit the available research on the developmental trajectory of VWM (Ross-sheehy et al., 2003; Simmering, 2012), with greater age predicting higher VWM-PL scores for infants but lower scores for their mothers (Brockmole & Logie, 2013).

Results in Studies 1 and 2 also adhered to the literature on the preferential looking task by revealing the consistent role of load (i.e., task difficulty) as a predictor of the proportion of looking to the changing side (Delgado Reyes et al., 2020). Moreover, in Study 1, load acted as a moderator, allowing the relationships between SES scores, maternal age, experiences of IPV, and the maternal VWM-PL outcome to appear when addressing task performance at specific difficulties. A particularly striking finding entailed the interaction between load, IPV, and maternal age, such that at the medium load condition young mothers showed decreasing working memory the more abuse they had reported but the relationship was reversed for older mothers (more IPV relating to more VWM). We discussed findings in relation to the potential hipervigilance of older mothers who reported experiencing IPV and are more likely to have experienced it for longer periods of time, making them hypersensitive to detecting changes in the environment.

Our work also contributes to the literature on the preferential looking task by revealing the influence of measuring VWM-PL specifically when the first look was directed at the no-change side. Like in the case of load, the first look allowed analyses to display greater variability which served to display interesting interactions. For instance, the time-course analysis in Study 1 revealed that at appropriate difficulty levels (medium and high loads), there was a significant relationship between SES and VWM-PL but only when the first look was directed at the no-change side. That is, in medium and high load conditions, when the first look was directed at the no-change side, the higher the SES, the more robustly that participants released their fixation and redirected their attention to the changing side.

The SES scores of families also interacted with the sense of empowerment of mothers and the gender of infants in a three-way interaction. Specifically for mothers of female infants, at higher SES scores a positive relationship was found between empowerment and maternal VWM-PL. However, at lower SES scores, the relationship was reversed, with more empowerment predicting less VWM-PL. Findings reveal the complex interplay of socio-economic, socio-emotional, and socio-cultural (e.g., gendered norms) variables.

Although several socio-emotional measures were related to maternal VWM-PL performance, they did not predict the VWM-PL scores of infants. Instead, caregivers' own VWM predicted that of their infants, such that the higher the VWM of the caregivers, the lower the VWM of their infants. The inverse relationship was interpreted in terms of the potential effect of a VWM mismatch on parent-infant interactions, with greater mismatch resulting in interactions that are less supportive of infant VWM development. Taken together, findings invite further research that takes into account the socio-economic and socio-emotional characteristics of infants and their families. Our findings further provide a base for future confirmatory research on mediational chains that extend from the broader social context to the development of the infant through their caregivers. Although this is of interest across socio-cultural contexts, it is of particular importance when researching understudied populations.

Having set an initial picture of the VWM of our sample in the lab setting, we then turned our attention towards researching the real-life deployment of visual exploration in dyadic interactions between infants and their caregivers (in our sample, this was their mothers). Gathering and analysing real dyadic data presents a range of difficulties. Regardless of the setting, coding data from recordings of full dyadic interactions from a large sample of participants is highly time-consuming. Moreover, within our setting, access to technology and electricity on the field was severely limited, so the process of data collection faced additional difficulties. Chapter 3, thus, elaborated our methodological approach to resolving these difficulties. In it, we addressed technological advances which enabled the collection of data in both low and high-resource settings. More specifically, portable equipment that we could easily transport with us to India included head-mounted cameras and eye-trackers. The main contribution of Chapter 3, however, was the development and validation of a machine learning pipeline that allowed us to extract objective measures of visual cognition from the recordings of caregiver-infant interactions in an efficient manner. The pipeline was trained and validated not only with dyadic data from our Indian participants but also with data from dyads in the UK, thus assuring its validity for use with different infrastructure and resource settings.

Having developed our machine learning pipeline, in Chapter 4 we turned our attention toward the evaluation of the visual dynamics in caregiver-infant interactions within distinct cultural settings. Although the focus of this thesis is directed at our Indian sample, the inclusion of British dyads served to establish a point of connection with the broader literature. Moreover, the British participants provided a point of comparison that enabled us to explore the deployment of visual attention from an eco-cultural perspective (Keller et al., 2005; Keller, 2007), considering similarities and differences across cultural groups and putting findings in the context of relevant parenting practices.

Results revealed consistent trends across groups, with the younger Indian infants spending more time than the other groups attending to their caregivers' faces and the older infants from India spending more time directing their attention towards toys. The older group of Indian infants also led a greater proportion of joint attention episodes than their younger counterparts. Findings also revealed cultural differences, with British dyads engaging in the greatest proportion of joint attention led by infants. Further cultural differences were found when comparing the attention towards faces during bouts of joint attention by British infants and the older Indian cohort. British infants sustained their attention towards their caregivers' faces more than the Indian group when the infant led the joint attention bout. In contrast, the Indian 9-month sample sustained their attention towards their caregivers' faces for longer than the British infants when joint attention was led by caregivers. Findings were interpreted by considering proximal and distal parenting styles, the former (more common in India) being more directive and the latter (more common in the UK) being more child-centric. The eco-cultural lens put the greater proportion of caregiver-led joint attention in India in the context of the directive parenting style. Similarly, it put the infant-led joint attention in the context of child-centred parenting. Critically, in both settings, infants seemed to respond to the parenting style.

Chapter 5 elaborated the final study within this thesis which connected the laboratory measures of VWM from Studies 1 and 2 to those taken from the dyadic task in Study 3. Findings relating specific, and intuitively comparable, measures from the lab and dyadic settings (i.e., MLD and switch rate) consistently revealed the lack of links between them, raising caution against the assumption that the measures are equivalent in different contexts (also see Wass, 2014). However, analyses linking measures of MLD from the caregiver-infant interaction to the VWM-PL scores obtained in the laboratory indicated that infants' MLD outside the episodes of joint attention significantly predicted their

VWM-PL performance. Moreover, the links between the proportion of joint attention led by infants (versus caregivers) and the VWM-PL scores revealed consistent associations, with both caregivers and infants leading a greater proportion of joint attention episodes the greater their VWM-PL scores. Findings were consistent with research on interventions which foster infant-led interactions (Landry et al., 2006, 2008). In addition, results revealed the consistent direct association between preferential looking task load and VWM-PL, with the load also interacting with MLD and with the proportion of infant-led joint attention in the infants' models. More specifically, we found that relationships were stronger in the medium-load condition, which seems to have optimally taxed infants' working memory abilities, revealing robust individual differences.

Taken together, findings across the four studies in this project can be summarised along three lines of contribution to the literature. The first line relates to the contributions made to the literature built upon the preferential-looking task. Our findings served to indicate the value of evaluating VWM-PL when looks are initially directed at the no-change side (Study 1). Moreover, findings contribute to the extensive body of research by providing insights into the "Goldilocks" spots of difficulty provided by the medium load condition, which allows the relationships between other variables of interest and the VWM-PL score of caregivers as well as infants to become apparent (Studies 1 & 4). Contributions within this line also include the exploration of relationships between the laboratory measures of visual cognition and those from dyadic interactions. Not only does Study 4 serve to caution future research against assuming the equivalence of measures across research settings (see also Wass, 2014), but also served to provide initial data upon which future research can be built. For instance, as discussed in Chapter 5, future research may want to consider the relationships between more nuanced measures of MLD from the lab and lounge that control for the target of attention (e.g., faces versus toys).

The second line of contribution to the field stems from the development of a machine learning pipeline which was trained and validated cross-culturally. Although the process of developing such pipelines is time-consuming, once developed, they allow the objective computation of measures of interest from large quantities of data across high and low resource settings. A collaborative project between researchers across multiple sites would serve to create a larger database with common, everyday toys that are easily (and commercially) available to researchers across socio-cultural contexts. Once the machine learning algorithm is trained with a larger database of accessible toys, the shared open pipeline will facilitate future research across multiple settings. A collaborative endeavour of this type will allow us, researchers, to diversify our pool of participants and will enable the efficient and objective coding of video-recorded observations used for quantitative analyses.

The third and main line of contributions entails the findings connecting the VWM of infants to their caregivers across methodological paradigms and contextualised within socio-cultural elements. Contributions start with findings on maternal VWM-PL (Study 1) being associated with her SES, experience of domestic violence, empowerment, and infants' gender but not with depression. We also found an unexpected inverse relationship between maternal VWM-PL and that of their infants (Study 2). The exploration of cultural similarities and differences in dyadic interactions (study 3) further served to put the deployment of visual attention in the context of child-centred and directive parenting practices, indicating that infants respond to said practices. Findings on Study 4 closed the loop by indicating that infant VWM-PL scores were predicted by the proportion of infant-led joint attention whereas VWM-PL scores of caregivers were predicted by the proportion of joint attention episodes that they led.

We further found support for our mismatch interpretation in Study 2, with greater mismatch between the VWM of caregivers and infants being related to shorter sustained attention by the infants. Moreover, greater mismatch was also related to a higher proportion of infant-terminated episodes of joint attention. The greater maternal VWM being related to greater caregiver-led initiation of joint attention, together with the relationship between VWM mismatch and infant termination of episodes, was interpreted in term of parental redirection of attention which the infants did not sustain. Findings, thus, suggested that for mothers with higher VWM in our Indian sample, behavioural patterns during interactions were not supportive of infant VWM development.

Overall, the research project succeeded in developing an eco-cultural picture of the development of VWM among Indian infants that accounts for maternal characteristics and socio-economic influences, and which connects laboratory and real-world research traditions. Our work therefore provides an empirical base upon which future confirmatory research can build in order to elaborate our understanding of the connections between the laboratory and real-world measures of visual cognition across socio-cultural contexts. Despite caution in the design, robust methods, and a relatively large cross-cultural sample on an understudied population (particularly pre-attrition), the project faced some challenges and has some limitations that need to be considered.

6.2 Challenges, Limitations, and Strengths

The context of running studies in Shivgarh, India involved working with up to four-to-five families in a day. Families travelled to the lab once a year, thus providing us with only one shot at good quality data collection. In addition, working in rural areas with restricted infrastructure came with key challenges. Therefore, despite access to a large sample and being able to test a large number of dyadic interactions in our first visit, we lost a significant amount of data due to power cuts forcing us to stop sessions or restricting us from checking data quality. Similarly, overheating of the equipment in the summer led to complications in the process of data collection. The problem with the power cuts was fixed by bringing in generators to provide continuous electricity, whereas setting up air conditioning helped prevent equipment from overheating. Nevertheless, there was an inevitable loss of data from dyadic interactions in the first round of our data collection.

The loss of data due to technological difficulties was compounded by attrition and loss of participants across tasks. This research was developed within a larger project in collaboration with a local partner organisation in India. This came with the advantage of access to participants and the possibility of carrying out multiple studies with the same families. At the same time, the large scale of the project went hand in hand with the difficulty of gathering all data for all studies from all participants (as shown by dwindling samples across tasks). Thus, we were able to study the same sample across settings (lab versus lounge) and using multiple measures (caregiver and infants' VWM-PL, caregivers' socio-emotional measures, dyadic interaction, demographic data, etc). This strength was inseparable from the limitations on sample size, due to attrition and loss of data. Nevertheless, initial results from this project will allow future research to address more targeted questions which do not demand having access to the same participants and completing as many measures and tasks over long periods of time.

Given the western bias in the majority of the developmental psychological literature, there was a large distance between the available literature and our population of interest. Focusing on caregiver-infant interaction during play allowed us to address an activity that our participants could carry out and which could be connected to existing research on caregiver-infant interactions. However, there are some caveats to consider. The first point to consider entails the cultural variations in the understanding of play. Here, it is worth noting the work of anthropologist Lancy (2016, 2007), who notes that adult-infant play is most commonly found in Euro-American and some foraging communities (Lancy, 2007). In contrast, in other communities, such as rural Guinea-Bissau where the infant mortality rate is high, mothers are expected not to play with their infants to avoid forming an attachment with them. Although the practices in Guinea-Bissau are not comparable to what was expected from our research setting, it must be noted that cultural variations in the purpose of playing and the form that ”play” takes, should still be expected.

A critical methodological detail that serves to tackle this issue entails instructing participants to ”play as you usually would” during the free play session. Instructing parents to ”play as you usually would” does not entail imposing a culturally biased ”goal” or concrete task in an explicit manner. Thus, it accommodates variations in the understanding of play and serves to capture typical everyday interactions between caregivers and their infants (Abels et al., 2017). Although this instruction enables studying whatever ”play” is seen to be in a given social context, the cultural variations that it enables come at the expense of control over the tasks/activities. Indeed, if the goal of play (e.g., distracting the infant versus soothing the infant versus. fostering the infant's development) and the activity understood as "play" change, then key behavioural patterns and some of the outcomes of play should be expected to change as well. On balance, for our research, the advantages provided by fostering cultural variations in behaviour and enhanced ecological validity outweighed the limitations of loss of control. However, the limitation still deserves recognition as it means that our findings addressing play activities (as those in other research) may not be easily connected across cultures. Qualitative input from future research on the understanding of the play by participants and the degree to which they engage in playing activities in their typical day-to-day life (and for what purposes) will enhance our insights into this culturally varied and complex activity. This qualitative input would be particularly useful when integrated in observational quantitative research.

Future research should also consider other forms of caregiver-infant interaction such as infant massage, which is a common traditional practice in India (Chaturvedi et al., 2020; Joseph et al., 2013). Research on the prevalence and perception of massage in central and western India has noted not only that mothers perceive infant massage as beneficial for infants' physical health (e.g. bone strength, better sleep, growth) but that they also believe that massaging babies makes them active, playful, smile more, and less irritable (Chaturvedi et al., 2020). Usually while massaging infants, mothers lay their infants face-up on their stretched legs, and move back and forward to massage them (as well as give them some light stretches). Understanding caregiver-infant interaction during massage can tell us more about face-to-face interaction between mothers and infants in communities where it is an everyday practice such as countries in Asia and Africa (Falle et al., 2009; Fikree, Ali, Durocher, & Rahbar, 2005).

The partial disconnection between the literature and our understudied population also resulted in our exploratory stance. This exploratory stance still set questions in line with the literature but relied on an inductive approach to developing and reporting models. That is, although variables included were informed by the literature, the inclusion and exclusion of interaction effects and, thus, the selection of the models reported in our studies were data-driven. Exploratory research and data-driven models play an important part in the development of research and can help inform later studies as well as theory development. However, they have their own limitations, which include a greater risk of type 1 errors when carrying out multiple analyses. In our research, this limitation was compounded by the sample limitations, which required us to simplify our models and constrain the use of multiple comparisons. Therefore, although our findings reveal overall consistent trends which can be explained through available theory and research, it is important for future confirmatory research to replicate and expand upon our findings.

A final key point worth highlighting entails a key strength of this research project, which stems from its collaborative and international approach. Although this is my own thesis, as an individual, the work carried out in it drew from the input of multiple colleagues across organisations and cultural settings. My own standing and life experiences, as a woman born and raised in India, meant that some of my expectations related to the research (and what we would find) were informed by my own background as well as by my knowledge of the literature. Input from western colleagues and the supervisor of this project balanced my input and served to keep biases at bay when setting goals as well as when interpreting results. On the other hand, input from our collaborators in India (CEL) was invaluable to adjust our research, measures and tasks to the norms and expectations of the community with whom we were working. Although collaborative engagement (with supervisors, lab colleagues, etc) is to be expected and valuable in any doctoral research, it was particularly the case in this project and served to balance many of the limitations previously addressed. Collaborative contributions served to outweigh the challenges of working within a larger project, ensured that our tasks were appropriate and ecologically valid for research with this specific population (e.g., using stainless steel plates and spoons as toys), and kept biased post-hoc rationalisation of results to a minimum (cross-cultural input being of particular value on this point). Taken together, although limitations need to be borne in mind when considering our findings and the future research that our work informs, they are balanced, if not outweighed, by the strengths of our research approach, particularly the checks and balances gained from collaborative elements.

6.3 Real-Life Implications and Applications

A central tenet in this thesis has been that understanding the relationships between laboratory and real-world research traditions in developmental psychology is of great importance to developing a valid and reliable body of knowledge. We have further developed our work under the guiding principle that the knowledge that we construct needs to be contextualised so that applications grounded on our findings adhere to the complex and multifaceted realities of human life. These two points are of special importance in the context of developing interventions.

In this thesis, we have consistently borne in mind that applying western interventions in other socio-cultural settings can be problematic if not harmful (Morelli et al., 2018; Scheidecker et al., 2021; Weber et al., 2021). As an example, an intervention aiming at fostering language development which encourages caregivers to speak more to their infants may be problematic in settings in which scholastic achievement is less important than motor skills and other developmental outcomes. Thus, rather than focusing on a concrete developmental outcome of interest in the west-informed literature, we decided to focus on an underlying cognitive processing capacity with multiple daily-life outcomes. Indeed, the use of VWM for detecting change and processing visual information can be of relevance for tasks as varied as crossing the road, driving, and detecting threats or dangers in busy environments (see Luck & Vogel, 1997; Vogel et al., 2001). Moreover, as mentioned in the introduction, we know that VWM can be assessed early in life and is open to intervention. Therefore, our research and findings can be applied to the real world through informing a range of targeted interventions of value to the community with whom we conducted our research.

With our focus being placed on infant VWM development as the central outcome, the core applications of our findings can inform two key points of intervention. First, findings in our research suggested that some behavioural patterns in dyadic interactions in the setting of rural India are not supportive of VWM. That is, greater proportion of caregiver-led joint attention, which we interpreted as fitting directive parenting styles, was related to higher maternal VWM but lower infant VWM. These findings not only support prior research on interventions carried out in western settings (Landry et al., 2006, 2008) but also identify our population as being among those likely to benefit from interventions in which parents are instructed to follow up on the attention of their infants (see Aboud & Yousafzai, 2015; L. Zhang et al., 2021). Moreover, our findings identify caregivers with higher VWM as a particularly important point of intervention for they tended to lead a greater proportion of joint attention episodes.

The second point of intervention relates to findings on the differences in parental engagement with older and younger infants. These findings were interpreted in relation to the expectations of what children are capable of doing at different ages (see Jeong et al., 2018). Our findings suggest that interventions in the setting of Shivgarh should not be limited to encouraging VWM-supportive practices but should also foster shifts in parental engagement so that said practices occur at an earlier point in time. This may entail addressing the expectations of caregivers as well as their understanding of what infants are capable of doing and what is beneficial for them, thus fostering a view of infants as being active instead of passive from younger ages. Indeed, enhanced knowledge of early development is among the elements targeted and positively affected by successful interventions in low-and-middle-income countries (Jeong et al., 2018, 2021).

6.4 Review and Conclusion

The work developed throughout this thesis has served to establish connections from social elements to the characteristics of adults and through them to the VWM of their infants. Moreover, it has served to establish links between laboratory and naturalistic research traditions and put finding through an eco-cultural lens that takes into account cultural similarities and differences in caregiver-infant interactions. Our work has made valuable contributions to the literature ranging from providing methodological insights of relevance to laboratory and naturalistic research (e.g., related to task difficulty and to the differences between intuitively equivalent measures), to identifying interesting patterns in the relationship between caregiver and infant VWM (e.g., the inverse relationship found between both). Our work further provides valuable insights that can aid the development and implementation of interventions aimed at fostering infant VWM development in rural India.

Although not free from limitations, our work has met the challenges of an ambitious line of research within a larger cross-cultural project and succeeded in developing a picture of the complex interplay between key cultural, contextual, and individual factors of relevance to infant cognitive development. Collaborative input with colleagues across organisations (i.e., UEA and CEL) has also served to maximise objectivity and ensure that our work remained of relevance to the community with whom we were researching. Our work, therefore, serves to open the door for future academic and applied activities by providing a robust base upon which confirmatory research and interventions can build.

References

Abels, M., & Hutman, T. (2015, August). Infants' behavioral styles in joint attention situations and parents' socio-economic status. *Infant Behavior and Development*, *40*, 139–150. doi: 10.1016/j.infbeh.2015.05.004

Abels, M., Papaligoura, Z., Lamm, B., & Yovsi, R. D. (2017, September). How Usual Is "Play As You Usually Would"? A Comparison of Naturalistic Mother-Infant Interactions with Videorecorded Play Sessions in Three Cultural Communities. *Child Development Research*, *2017*, 1–8. doi: 10.1155/2017/7842030

Abney, D. H., Smith, L. B., & Yu, C. (2017). It's time: Quantifying the relevant timescales for joint attention. In *Cogsci.*

Abney, D. H., Suanda, S. H., Smith, L. B., & Yu, C. (2020, November). What are the building blocks of parent–infant coordinated attention in free-flowing interaction? *Infancy*, *25*(6), 871–887. doi: 10.1111/infa.12365

Aboud, F. E., Moore, A. C., & Akhter, S. (2008, October). Effectiveness of a community-based responsive feeding programme in rural Bangladesh: a cluster randomized field trial. *Maternal & Child Nutrition*, *4*(4), 275–286. doi: 10.1111/j.1740-8709.2008.00146.x

Aboud, F. E., & Yousafzai, A. K. (2015). Global Health and Development in Early Childhood. *Annual Review of Psychology*, *66*(1), 433–457. doi: 10.1146/annurev-psych-010814-015128

Ackerson, L. K., & Subramanian, S. (2008, September). State Gender Inequality, Socioeconomic Status and Intimate Partner Violence (IPV) in India: A Multilevel Analysis. *Australian Journal of Social Issues*, *43*(1), 81–102. doi: 10.1002/j.1839-4655.2008.tb00091.x

Ahmed, I., Ali, S. M., Amenga-Etego, S., Ariff, S., Bahl, R., Baqui, A. H., ... Zaidi, A. (2018, December). Population-based rates, timing, and causes of maternal deaths, stillbirths, and neonatal deaths in south Asia and sub-Saharan Africa: a multi-country prospective cohort study. *The Lancet Global Health*, *6*(12), e1297–e1308. doi: 10.1016/S2214-109X(18)30385-1

Ahmed, S. F., Tang, S., Waters, N. E., & Davis-Kean, P. (2019). Executive function and academic achievement: Longitudinal relations from early childhood to adolescence. *Journal of Educational Psychology*, *111*(3), 446–458. doi: 10.1037/edu0000296

Allen, K., Higgins, S., & Adams, J. (2019, September). The Relationship between Visuospatial Working Memory and Mathematical Performance in School-Aged Children: a Systematic Review. *Educational Psychology Review*, *31*(3), 509–531. doi: 10.1007/s10648-019-09470-8

Andrew, A., Attanasio, O., Augsburg, B., Day, M., Grantham-McGregor, S., Meghir, C., ... Rubio-Codina, M. (2020, June). Effects of a scalable home-visiting intervention on child development in slums of urban India: evidence from a randomised controlled trial. *Journal of Child Psychology and Psychiatry*, *61*(6), 644–652. doi: 10.1111/jcpp.13171

Aslin, R. N. (2009, June). How Infants View Natural Scenes Gathered From a Head-Mounted Camera. *Optometry and Vision Science*, *86*(6), 561–565. doi: 10.1097/OPX.0b013e3181a76e96

Attanasio, O. P., Fernandez, C., Fitzsimons, E. O. A., Grantham-McGregor,

S. M., Meghir, C., & Rubio-Codina, M. (2014, September). Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial. *BMJ*, *349*(sep29 5), g5785–g5785. doi: 10.1136/bmj.g5785

Baddeley, A. D. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, *63*(1), 1–29. doi: 10.1146/annurev-psych-120710-100422

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology*, *8*(4), 485–493. doi: 10.1037/0894-4105.8.4.485

Bakeman, R., & Adamson, L. B. (1984, August). Coordinating Attention to People and Objects in Mother-Infant and Peer-Infant Interaction. *Child Development*, *55*(4), 1278. doi: 10.2307/1129997

Baldwin, D. A., & Markman, E. M. (1989, April). Establishing Word-Object Relations: A First Step. *Child Development*, *60*(2), 381. doi: 10.2307/1130984

Bard, K. A., Keller, H., Ross, K. M., Hewlett, B., Butler, L., Boysen, S. T., & Matsuzawa, T. (2021, December). Joint Attention in Human and Chimpanzee Infants in Varied Socio-Ecological Contexts. *Monographs of the Society for Research in Child Development*, *86*(4), 7–217. doi: 10.1111/mono.12435

Barrett, J., & Fleming, A. S. (2011, April). Annual Research Review: All mothers are not created equal: neural and psychobiological perspectives on mothering and the importance of individual differences: Neurobiological perspectives on mothering. *Journal of Child Psychology and Psychiatry*, *52*(4), 368–397. doi: 10.1111/j.1469-7610.2010.02306.x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Benjamin, D., Chandramohan, A., Annie, I., Prasad, J., & Jacob, K. (2005). Validation of the tamil version of edinburgh post-partum depression scale. *J Obstet Gynecol India*, *55*(3), 241–243.

Bhat, A. A., Spencer, J. P., & Samuelson, L. K. (2022, July). Word-Object Learning via Visual Exploration in Space (WOLVES): A neural process model of cross-situational word learning. *Psychological Review*, *129*(4), 640–695. doi: 10.1037/rev0000313

Black, M. M., Walker, S. P., Fernald, L. C. H., Andersen, C. T., DiGirolamo, A. M., Lu, C., ... Grantham-McGregor, S. (2017, January). Early childhood development coming of age: science through the life course. *The Lancet*, *389*(10064), 77–90. doi: 10.1016/S0140-6736(16)31389-7

Bliznashka, L., Udo, I. E., Sudfeld, C. R., Fawzi, W. W., & Yousafzai, A. K. (2021, September). Associations between women's empowerment and child development, growth, and nurturing care practices in sub-Saharan Africa: A cross-sectional analysis of demographic and health survey data. *PLOS Medicine*, *18*(9), e1003781. doi: 10.1371/journal.pmed.1003781

Bloom, S. S., Wypij, D., & Das Gupta, M. (2001, February). Dimensions of women's autonomy and the influence on maternal health care utilization in a North Indian city. *Demography*, *38*(1), 67–78. doi: 10.1353/dem.2001.0001

Boukerche, A., & Hou, Z. (2022, March). Object Detection Using Deep Learning Methods in Traffic Scenarios. *ACM Computing Surveys*, *54*(2), 1–35. doi: 10.1145/3434398

Britto, P. R., Ponguta, L., Reyes, C., & Karnati, R. (2015). A systematic review of parenting programmes for young children in low-and middle-income countries. *New York, NY: United Nations Children's Fund*, 363–394.

Brockmole, J. R., & Logie, R. H. (2013). Age-Related Change in Visual Working Memory: A Study of 55,753 Participants Aged 8–75. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00012

Broesch, T., Rochat, P., Olah, K., Broesch, J., & Henrich, J. (2016, May). Similarities and Differences in Maternal Responsiveness in Three Societies: Evidence From Fiji, Kenya, and the United States. *Child Development*, *87*(3), 700–711. doi: 10.1111/cdev.12501

Brookman, R., Kalashnikova, M., Conti, J., Xu Rattanasone, N., Grant, K., Demuth, K., & Burnham, D. (2020, November). Depression and Anxiety in the Postnatal Period: An Examination of Infants' Home Language Environment, Vocalizations, and Expressive Language Abilities. *Child Development*, *91*(6). doi: 10.1111/cdev.13421

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., ... Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400. Retrieved from `https://journal.r-project.org/archive/2017/RJ-2017-066/index.html`

Brooks, R., Singleton, J. L., & Meltzoff, A. N. (2020, March). Enhanced gaze-following behavior in Deaf infants of Deaf parents. *Developmental Science*, *23*(2). doi: 10.1111/desc.12900

Bull, R., Espy, K. A., & Wiebe, S. A. (2008, April). Short-Term Memory, Working Memory, and Executive Functioning in Preschoolers: Longitudinal Predictors of Mathematical Achievement at Age 7 Years. *Developmental Neuropsychology*, *33*(3), 205–228. doi: 10.1080/87565640801982312

Bull, R., & Scerif, G. (2001, June). Executive Functioning as a Predictor of Children's Mathematics Ability: Inhibition, Switching, and Working Memory. *Developmental Neuropsychology*, *19*(3), 273–293. doi: 10.1207/S15326942DN1903_3

Buss, A. T., Ross-Sheehy, S., & Reynolds, G. D. (2018, October). Visual working memory in early development: a developmental cognitive neuroscience perspective. *Journal of Neurophysiology*, *120*(4), 1472–1483. doi: 10.1152/jn.00087.2018

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development*, *63*(4), i. doi: 10.2307/1166214

Chandra, P. S., Satyanarayana, V. A., & Carey, M. P. (2009, August). Women reporting intimate partner violence in India: associations with PTSD and depressive symptoms. *Archives of Women's Mental Health*, *12*(4), 203–209. doi: 10.1007/s00737-009-0065-6

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000, October). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development*, *15*(4), 481–498. Retrieved 2022-07-20, from https://linkinghub.elsevier.com/retrieve/pii/S0885201401000375 doi: 10.1016/S0885-2014(01)00037-5

Chaturvedi, S., Randive, B., Pathak, A., Agarkhedkar, S., Tillu, G., Darmstadt, G. L., & Patwardhan, B. (2020, December). Prevalence and perceptions of infant massage in India: study from Maharashtra and Madhya Pradesh states. *BMC Pediatrics*, *20*(1), 512. doi: 10.1186/s12887-020-02416-y

Chavajay, P., & Rogoff, B. (1999). Cultural variation in management of attention by children and their caregivers. *Developmental Psychology*, *35*(4), 1079 - 1090. Retrieved from https://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=psyh&AN=1999-05934-018&authtype=sso&custid=s8993828&site=eds-live&scope=site

Check, C. b. (2016). *Colblindor.* Google Play Store. Retrieved from https://play.google.com/store/apps/details?id=ch.colblindor.colorblindcheck&hl=en_GB&gl=US

Clay, C. E. C., & Seehusen, M. D. A. (2004, February). A Review of Postpartum Depression for the Primary Care Physician:. *Southern Medical Journal*, *97*(2), 157–161. Retrieved 2023-05-19, from http://sma.org/southern-medical-journal/article/a-review-of-postpartum-depression-for-the-primary-care-physician doi: 10.1097/01.SMJ.0000091029.34773.33

Cohn, J. F., Matias, R., Tronick, E. Z., Connell, D., & Lyons-Ruth, K. (1986). Face-to-face interactions of depressed mothers and their infants. *New Directions for Child and Adolescent Development*, *1986*(34), 31–45. doi: 10.1002/cd.23219863405

Colombo, J., Mitchell, D. W., Coldren, J. T., & Freeseman, L. J. (1991, December). Individual differences in infant visual attention: are short lookers faster processors or feature processors? *Child Development*, *62*(6), 1247–1257.

Colombo, J., Mitchell, D. W., & Horowitz, F. D. (1988, October). Infant Visual Attention in the Paired-Comparison Paradigm: Test-Retest and Attention-Performance Relations. *Child Development*, *59*(5), 1198. doi: 10.2307/1130483

Costello, M. C., & Buss, A. T. (2018, October). Age-related Decline of Visual

Working Memory: Behavioral Results Simulated with a Dynamic Neural Field Model. *Journal of Cognitive Neuroscience*, *30*(10), 1532–1548. doi: 10.1162/jocn_a_01293

Cox, J., Holden, J., & Sagovsky, R. (1987). Detection of Postnatal Depression. *The British Journal of Psychiatry*, *150*, 5.

Cuevas, K., Deater-Deckard, K., Kim-Spoon, J., Wang, Z., Morasch, K. C., & Bell, M. A. (2014, March). A longitudinal intergenerational analysis of executive functions during early childhood. *British Journal of Developmental Psychology*, *32*(1), 50–64. Retrieved 2023-05-31, from https://onlinelibrary.wiley.com/doi/10.1111/bjdp.12021 doi: 10.1111/bjdp.12021

Cunningham, K., Ruel, M., Ferguson, E., & Uauy, R. (2015, January). Women's empowerment and child nutritional status in South Asia: a synthesis of the literature: Women's empowerment and child nutrition: South Asia. *Maternal & Child Nutrition*, *11*(1), 1–19. doi: 10.1111/mcn.12125

Dahl, A. (2017, June). Ecological Commitments: Why Developmental Science Needs Naturalistic Methods. *Child Development Perspectives*, *11*(2), 79–84. doi: 10.1111/cdep.12217

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early Social Attention Impairments in Autism: Social Orienting, Joint Attention, and Attention to Distress. *Developmental Psychology*, *40*(2), 271–283. doi: 10.1037/0012-1649.40.2.271

Delgado Reyes, L., Wijeakumar, S., Magnotta, V. A., Forbes, S. H., & Spencer, J. P. (2020, October). The functional brain networks that underlie visual working memory in the first two years of life. *NeuroImage*, *219*, 116971. doi: 10.1016/j.neuroimage.2020.116971

Dettrick, Z., Jimenez-Soto, E., & Hodge, A. (2014, May). Socioeconomic and Geographical Disparities in Under-Five and Neonatal Mortality in Uttar Pradesh, India. *Maternal and Child Health Journal*, *18*(4), 960–969. Retrieved 2023-05-18, from http://link.springer.com/10.1007/s10995-013-1324-8 doi: 10.1007/s10995-013-1324-8

DHS Program. (2020). *The DHS Program - Quality information to plan, monitor and improve population, health, and nutrition programs.* Retrieved 2022-01-01, from https://dhsprogram.com/

Dink, J., & Ferguson, B. (2016). *eyetrackingR: An R library for eye-tracking data analysis.* Retrieved from http://www.eyetracking-r.com

Dutton, M. A., & Goodman, L. A. (2005, June). Coercion in Intimate Partner Violence: Toward a New Conceptualization. *Sex Roles*, *52*(11-12), 743–756. doi: 10.1007/s11199-005-4196-6

Eysenck, M. W. (1997). *Anxiety and cognition: a unified theory.* Hove, East Sussex, UK: Psychology Press.

Falle, T. Y., Mullany, L. C., Thatte, N., Khatry, S. K., LeClerq, S. C., Darmstadt, G. L., ... Tielsch, J. M. (2009, September). Potential Role of Traditional Birth Attendants in Neonatal Healthcare in Rural Southern Nepal. *Journal of Health, Population and Nutrition*, *27*(1), 53–61. doi: 10.3329/jhpn.v27i1.3317

Fanari, R., Meloni, C., & Massidda, D. (2019, November). Visual and Spatial Working Memory Abilities Predict Early Math Skills: A Longitudinal Study. *Frontiers in Psychology*, *10*, 2460. doi: 10.3389/fpsyg.2019.02460

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016, July). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107. doi: 10.1016/j.cognition.2016.03.005

Feldman, R., Granat, A., Pariente, C., Kanety, H., Kuint, J., & Gilboa-Schechtman, E. (2009, September). Maternal Depression and Anxiety Across the Postpartum Year and Infant Social Engagement, Fear Regulation, and Stress Reactivity. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(9), 919–927. doi: 10.1097/CHI.0b013e3181b21651

Fikree, F. F., Ali, T. S., Durocher, J. M., & Rahbar, M. H. (2005, March). Newborn care practices in low socioeconomic settlements of Karachi, Pakistan. *Social Science & Medicine*, *60*(5), 911–921. doi: 10.1016/j.socscimed.2004.06.034

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third ed.). Thousand Oaks CA: Sage. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Franchak, J. M., Kretch, K. S., & Adolph, K. E. (2018, July). See and be seen: Infant–caregiver social looking during locomotor free play. *Developmental Science*, *21*(4). doi: 10.1111/desc.12626

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011, November). Head-Mounted Eye Tracking: A New Method to Describe Infant Looking: Head-Mounted Eye Tracking. *Child Development*, *82*(6), 1738–1750. doi: 10.1111/j.1467-8624.2011.01670.x

Friard, O., & Gamba, M. (2016, November). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*(11), 1325–1330. doi: 10.1111/2041-210X.12584

Frick, J. E., Colombo, J., & Allen, J. R. (2000, July). Temporal Sequence of Global-Local Processing in 3-Month-Old Infants. *Infancy*, *1*(3), 375–386. doi: 10.1207/S15327078IN0103_6

Fry, A. F., & Hale, S. (1996, July). Processing Speed, Working Memory, and Fluid Intelligence: Evidence for a Developmental Cascade. *Psychological*

Science, *7*(4), 237–241. doi: 10.1111/j.1467-9280.1996.tb00366.x

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004, January). Working memory skills and educational attainment: evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*(1), 1–16. Retrieved 2022-08-01, from https://onlinelibrary.wiley.com/doi/10.1002/acp.934 doi: 10.1002/acp.934

Gathercole, S. E., Tiffany, C., Briscoe, J., Thorn, A., & The ALSPAC team2. (2005, June). Developmental consequences of poor phonological short-term memory function in childhood: a longitudinal study. *Journal of Child Psychology and Psychiatry*, *46*(6), 598–611. doi: 10.1111/j.1469-7610.2004.00379.x

Gibbons, J. D., & Chakraborti, S. (2011). Nonparametric Statistical Inference. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 977–979). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-04898-2_420

Goncu, A. E. (1999). Children's engagement in the world: Sociocultural perspectives.

Gonzalez, A., Jenkins, J. M., Steiner, M., & Fleming, A. S. (2012, July). Maternal Early Life Experiences and Parenting: The Mediating Role of Cortisol and Executive Function. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(7), 673–682. doi: 10.1016/j.jaac.2012.04.003

Gotlib, I. H., & Joormann, J. (2010, March). Cognition and Depression: Current Status and Future Directions. *Annual Review of Clinical Psychology*, *6*(1), 285–312. doi: 10.1146/annurev.clinpsy.121208.131305

Grantham-McGregor, S., Adya, A., Attanasio, O., Augsburg, B., Behrman, J., Caeyers, B., . . . Vats, K. (2020, December). Group Sessions or Home Visits for Early Childhood Development in India: A Cluster RCT. *Pediatrics*, *146*(6), e2020002725. doi: 10.1542/peds.2020-002725

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007, January). Developmental potential in the first 5 years for children in developing countries. *The Lancet*, *369*(9555), 60–70. doi: 10.1016/S0140-6736(07)60032-4

Gustafsson, H. C., Coffman, J. L., & Cox, M. J. (2015). Intimate partner violence, maternal sensitive parenting behaviors, and children's executive functioning. *Psychology of Violence*, *5*(3), 266–274. doi: 10.1037/a0037971

Hamberger, L. K., Larsen, S. E., & Lehrner, A. (2017, November). Coercive control in intimate partner violence. *Aggression and Violent Behavior*, *37*, 1–11. doi: 10.1016/j.avb.2017.08.003

Harkness, S., & Super, C. M. (2020, September). Culture and human development: Where did it go? And where is it going? *New Directions for Child and Adolescent Development*, *2020*(173), 101–119. doi: 10.1002/cad.20378

Hartig, F. (2021). Dharma: Residual diagnostics for hierarchical (multi-level / mixed) regression models [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=DHARMa` (R package version 0.4.4)

Harvey, P., Le Bastard, G., Pochon, J., Levy, R., Allilaire, J., Dubois, B., & Fossati, P. (2004, November). Executive functions and updating of the contents of working memory in unipolar depression. *Journal of Psychiatric Research*, *38*(6), 567–576. doi: 10.1016/j.jpsychires.2004.03.003

Henrich, J., Heine, S. J., & Norenzayan, A. (2010, June). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61–83. doi: 10.1017/

S0140525X0999152X

Holmes, J., & Adams, J. W. (2006, June). Working Memory and Children's Mathematical Skills: Implications for mathematical development and mathematics curricula. *Educational Psychology*, *26*(3), 339–366. doi: 10.1080/01443410500341056

Holmes, J., Adams, J. W., & Hamilton, C. J. (2008, March). The relationship between visuospatial sketchpad capacity and children's mathematical skills. *European Journal of Cognitive Psychology*, *20*(2), 272–289. doi: 10.1080/09541440701612702

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009, July). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), F9–F15. doi: 10.1111/j.1467-7687.2009.00848.x

Home Office. (2018). *Domestic abuse: how to get help.* Retrieved from `https://www.gov.uk/guidance/domestic-abuse-how-to-get-help`

Jankowski, J. J., Rose, S. A., & Feldman, J. F. (2001, March). Modifying the Distribution of Attention in Infants. *Child Development*, *72*(2), 339–351. doi: 10.1111/1467-8624.00282

Jarvis, H. L., & Gathercole, S. E. (2003). Verbal and non-verbal working memory and achievements on national curriculum tests at 11 and 14 years of age. *Educational and Child Psychology*, *20*(3), 123–140.

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2017). Why are faces denser in the visual experiences of younger than older infants? *Developmental Psychology*, *53*(1), 38–49. doi: 10.1037/dev0000230

Jensen, S. K. G., Berens, A. E., & Nelson, C. A. (2017, November). Effects of poverty on interacting biological systems underlying child development. *The Lancet. Child & Adolescent Health*, *1*(3), 225–239. doi: 10.1016/S2352

-4642(17)30024-X

Jeong, J., Franchett, E. E., Ramos de Oliveira, C. V., Rehmani, K., & Yousafzai, A. K. (2021, May). Parenting interventions to promote early child development in the first three years of life: A global systematic review and meta-analysis. *PLOS Medicine*, *18*(5), e1003602. doi: 10.1371/journal.pmed.1003602

Jeong, J., Pitchik, H. O., & Yousafzai, A. K. (2018, April). Stimulation Interventions and Parenting in Low- and Middle-Income Countries: A Meta-analysis. *Pediatrics*, *141*(4), e20173510. doi: 10.1542/peds.2017-3510

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., ... Minh, M. T. (2022, February). *ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference.* Zenodo. Retrieved 2022-08-02, from `https://zenodo.org/record/6222936` doi: 10.5281/ZENODO.6222936

Joormann, J., & Gotlib, I. H. (2008, February). Updating the contents of working memory in depression: Interference from irrelevant negative material. *Journal of Abnormal Psychology*, *117*(1), 182–192. doi: 10.1037/0021-843X.117.1.182

Joseph, N., Unnikrishnan, B., Naik, V., Mahantshetti, N., Mallapur, M., Kotian, S., & Nelliyanil, M. (2013). Infant rearing practices in south India: A longitudinal study. *Journal of Family Medicine and Primary Care*, *2*(1), 37. doi: 10.4103/2249-4863.109942

Joshi, U., Lyngdoh, T., & Shidhaye, R. (2020, February). Validation of hindi version of Edinburgh postnatal depression scale as a screening tool for antenatal depression. *Asian Journal of Psychiatry*, *48*, 101919. doi: 10.1016/j.ajp.2019.101919

Kagitcibasi, C. (2005, July). Autonomy and Relatedness in Cultural Context:

Implications for Self and Family. *Journal of Cross-Cultural Psychology*, *36*(4), 403–422. doi: 10.1177/0022022105275959

Kail, R., & Salthouse, T. A. (1994, August). Processing speed as a mental capacity. *Acta Psychologica*, *86*(2-3), 199–225. doi: 10.1016/0001-6918(94)90003-5

Karasik, L. B., Tamis-LeMonda, C. S., Adolph, K. E., & Bornstein, M. H. (2015, September). Places and Postures: A Cross-Cultural Comparison of Sitting in 5-Month-Olds. *Journal of Cross-Cultural Psychology*, *46*(8), 1023–1038. doi: 10.1177/0022022115593803

Kassner, M., Patera, W., & Bulling, A. (2014, September). Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (pp. 1151–1160). Seattle Washington: ACM. doi: 10.1145/2638728.2641695

Katus, L., Hayes, N. J., Mason, L., Blasi, A., McCann, S., Darboe, M. K., ... Elwell, C. E. (2019). Implementing neuroimaging and eye tracking methods to assess neurocognitive development of young infants in low- and middle-income countries. *Gates Open Research*, *3*, 1113. doi: 10.12688/gatesopenres.12951.2

Keller, H. (2007). *Cultures of infancy.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers. (Pages: xix, 318)

Keller, H. (2017, September). Culture and Development: A Systematic Relationship. *Perspectives on Psychological Science*, *12*(5), 833–840. doi: 10.1177/1745691617704097

Keller, H. (2018, December). Parenting and socioemotional development in infancy and early childhood. *Developmental Review*, *50*, 31–41. doi: 10.1016/j.dr.2018.03.001

Keller, H., Abels, M., Lamm, B., Yovsi, R. D., Voelker, S., & Lakhani, A. (2005, December). Ecocultural Effects on Early Infant Care: A Study in Cameroon, India, and Germany. *Ethos*, *33*(4), 512–541. doi: 10.1525/eth.2005.33.4.512

Keller, H., Borke, J., Chaudhary, N., Lamm, B., & Kleis, A. (2010, May). Continuity in Parenting Strategies: A Cross-Cultural Comparison. *Journal of Cross-Cultural Psychology*, *41*(3), 391–409. doi: 10.1177/0022022109359690

Keller, H., Borke, J., Lamm, B., Lohaus, A., & Dzeaye Yovsi, R. (2011, May). Developing patterns of parenting in two cultural communities. *International Journal of Behavioral Development*, *35*(3), 233–245. doi: 10.1177/0165025410380652

Keller, H., Borke, J., Staufenbiel, T., Yovsi, R. D., Abels, M., Papaligoura, Z., ... Su, Y. (2009, September). Distal and proximal parenting as alternative parenting strategies during infants' early months of life: A cross-cultural study. *International Journal of Behavioral Development*, *33*(5), 412–420. doi: 10.1177/0165025409338441

Keller, H., & Demuth, C. (2007). The discursive construction of selfhood in chinese and euro-american mother-infant interactions. In *Perspectives and progress in contemporary cross-cultural psychology* (pp. 77–88). China Light Industry Press.

Keller, H., Papaligoura, Z., Kuensemueller, P., Voelker, S., Papaeliou, C., Lohaus, A., ... Mousouli, V. (2003, November). Concepts of Mother-Infant Interaction in Greece and Germany. *Journal of Cross-Cultural Psychology*, *34*(6), 677–689. doi: 10.1177/0022022103257035

Keller, H., Yovsi, R., Borke, J., Kartner, J., Jensen, H., & Papaligoura, Z. (2004, December). Developmental Consequences of Early Parenting

Experiences: Self-Recognition and Self-Regulation in Three Cultural Communities. *Child Development*, *75*(6), 1745–1760. doi: 10.1111/j.1467-8624.2004.00814.x

Kessler, R. C., McLaughlin, K. A., Green, J. G., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M., ... Williams, D. R. (2010, November). Childhood adversities and adult psychopathology in the WHO World Mental Health Surveys. *British Journal of Psychiatry*, *197*(5), 378–385. doi: 10.1192/bjp.bp.110.080499

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., ... Westerberg, H. (2005). Computerized Training of Working Memory in Children With ADHD-A Randomized, Controlled Trial. *J. AM. ACAD. CHILD ADOLESC. PSYCHIATRY*, 10.

Krug, E. G., Mercy, J. A., Dahlberg, L. L., & Zwi, A. B. (2002, October). The world report on violence and health. *The Lancet*, *360*(9339), 1083–1088. doi: 10.1016/S0140-6736(02)11133-0

Kudo, M. F., Lussier, C. M., & Swanson, H. L. (2015, May). Reading disabilities in children: A selective meta-analysis of the cognitive literature. *Research in Developmental Disabilities*, *40*, 51–62. Retrieved 2022-08-01, from https://linkinghub.elsevier.com/retrieve/pii/S0891422215000037 doi: 10.1016/j.ridd.2015.01.002

Kumar, V., Mohanty, S., Kumar, A., Misra, R. P., Santosham, M., Awasthi, S., ... Darmstadt, G. L. (2008, September). Effect of community-based behaviour change management on neonatal mortality in Shivgarh, Uttar Pradesh, India: a cluster-randomised controlled trial. *The Lancet*, *372*(9644), 1151–1162. doi: 10.1016/S0140-6736(08)61483-X

Kärtner, J., Keller, H., Lamm, B., Abels, M., Yovsi, R. D., Chaudhary, N., & Su, Y. (2008, September). Similarities and differences in contingency experiences of 3-month-olds across sociocultural contexts. *Infant Behavior and Development*, *31*(3), 488–500. doi: 10.1016/j.infbeh.2008.01.001

Kärtner, J., Keller, H., & Yovsi, R. D. (2010, March). Mother-Infant Interaction During the First 3 Months: The Emergence of Culture-Specific Contingency Patterns. *Child Development*, *81*(2), 540–554. doi: 10.1111/j.1467-8624.2009.01414.x

LaBore, K., Ahmed, T., Rizwan-ur-Rashid, & Ahmed, R. (2021, July). Prevalence and Predictors of Violence Against Women in Pakistan. *Journal of Interpersonal Violence*, *36*(13-14), NP7246–NP7263. Retrieved 2023-05-18, from `http://journals.sagepub.com/doi/10.1177/0886260518824652` doi: 10.1177/0886260518824652

Lamm, B., Gudi, H., Fassbender, I., Freitag, C., Graf, F., Goertz, C., ... Keller, H. (2015). Rural NSO and German middle-class mothers' interaction with their 3- and 6-month-old infants: A longitudinal cross-cultural analysis. *Journal of Family Psychology*, *29*(4), 649–655. Retrieved 2021-07-10, from `http://doi.apa.org/getdoi.cfm?doi=10.1037/fam0000100` doi: 10.1037/fam0000100

Lamm, B., Gudi, H., Freitag, C., Teubert, M., Graf, F., Fassbender, I., ... Keller, H. (2014, July). Mother–Infant Interactions at Home and in a Laboratory Setting: A Comparative Analysis in Two Cultural Contexts. *Journal of Cross-Cultural Psychology*, *45*(6), 843–852. doi: 10.1177/0022022114532357

Lancy, D. F. (2007, June). Accounting for Variability in Mother?Child Play. *American Anthropologist*, *109*(2), 273–284. doi: 10.1525/aa.2007.109.2.273

Lancy, D. F. (2010, September). Learning 'From Nobody': The Limited Role of Teaching in Folk Models of Children's Development. *Childhood in the*

Past, *3*(1), 79–106. doi: 10.1179/cip.2010.3.1.79

Lancy, D. F. (2014). *The Anthropology of Childhood: Cherubs, Chattel, Changelings* (2nd ed.). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139680530

Lancy, D. F. (2016, May). Playing With Knives: The Socialization of Self-Initiated Learners. *Child Development*, *87*(3), 654–665. doi: 10.1111/cdev.12498

Landry, S. H., Smith, K. E., & Swank, P. R. (2006, July). Responsive parenting: Establishing early foundations for social, communication, and independent problem-solving skills. *Developmental Psychology*, *42*(4), 627–642. doi: 10.1037/0012-1649.42.4.627

Landry, S. H., Smith, K. E., Swank, P. R., & Guttentag, C. (2008). A responsive parenting intervention: The optimal timing across early childhood for impacting maternal behaviors and child outcomes. *Developmental Psychology*, *44*(5), 1335–1353. doi: 10.1037/a0013030

Levens, S. M., & Gotlib, I. H. (2010). Updating positive and negative stimuli in working memory in depression. *Journal of Experimental Psychology: General*, *139*(4), 654–664. doi: 10.1037/a0020283

Little, E. E., Carver, L. J., & Legare, C. H. (2016, July). Cultural Variation in Triadic Infant-Caregiver Object Exploration. *Child Development*, *87*(4), 1130–1145. doi: 10.1111/cdev.12513

Long, B. L., Sanchez, A., Kraus, A. M., Agrawal, K., & Frank, M. C. (2022, January). Automated detections reveal the social information in the changing infant view. *Child Development*, *93*(1), 101–116. doi: 10.1111/cdev.13648

Lovejoy, M., Graczyk, P. A., O'Hare, E., & Neuman, G. (2000, August). Maternal depression and parenting behavior. *Clinical Psychology Review*, *20*(5), 561–592. doi: 10.1016/S0272-7358(98)00100-7

Luck, S. J., & Vogel, E. K. (1997, November). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. doi: 10.1038/36846

Malik, M., Munir, N., Ghani, U., & Ahmad, N. (2020, December). Domestic violence and its relationship with depression, anxiety and quality of life: A hidden dilemma of Pakistani women. *Pakistan Journal of Medical Sciences*, *37*(1). doi: 10.12669/pjms.37.1.2893

Markus, J., Mundy, P., Morales, M., Delgado, C. E. F., & Yale, M. (2000, August). Individual Differences in Infant Skills as Predictors of Child-Caregiver Joint Attention and Language. *Social Development*, *9*(3), 302–315. doi: 10.1111/1467-9507.00127

Masmoudi, M., Friji, H., Ghazzai, H., & Massoud, Y. (2021). A Reinforcement Learning Framework for Video Frame-Based Autonomous Car-Following. *IEEE Open Journal of Intelligent Transportation Systems*, *2*, 111–127. doi: 10.1109/OJITS.2021.3083201

McCall, R. B. (1977, June). Challenges to a Science of Developmental Psychology. *Child Development*, *48*(2), 333. doi: 10.2307/1128626

McCoy, D. C., Peet, E. D., Ezzati, M., Danaei, G., Black, M. M., Sudfeld, C. R., ... Fink, G. (2016, June). Early Childhood Developmental Status in Low- and Middle-Income Countries: National, Regional, and Global Prevalence Estimates Using Predictive Modeling. *PLOS Medicine*, *13*(6), e1002034. doi: 10.1371/journal.pmed.1002034

Milosavljevic, B., Vellekoop, P., Maris, H., Halliday, D., Drammeh, S., Sanyang, L., ... Lloyd-Fox, S. (2019, September). Adaptation of the Mullen Scales of Early Learning for use among infants aged 5- to 24-months in rural Gambia. *Developmental Science*, *22*(5). doi: 10.1111/desc.12808

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. New York: Chapman and Hall/CRC. doi: https://doi.org/10.1201/9781315373218

Mohan, M., & Bhat, J. S. (2022, January). Influence of socio-economic status on play development in toddlers – a mother child dyadic interaction based study. *Neuropsychiatrie de l'Enfance et de l'Adolescence*, S0222961721002233. doi: 10.1016/j.neurenf.2021.11.005

Moore, C. ., & Dunham, P. J. (1995). *Joint attention : its origins and role in development*. Hillsdale, N.J.: Lawrence Erlbaum Associates. (Section: viii, 286 pages : illustrations ; 24 cm)

Morelli, G., Quinn, N., Chaudhary, N., Vicedo, M., Rosabal-Coto, M., Keller, H., ... Takada, A. (2018, January). Ethical Challenges of Parenting Interventions in Low- to Middle-Income Countries. *Journal of Cross-Cultural Psychology*, *49*(1), 5–24. doi: 10.1177/0022022117746241

Moser, D. A., Doucet, G. E., Ing, A., Dima, D., Schumann, G., Bilder, R. M., & Frangou, S. (2018, October). An integrated brain–behavior model for working memory. *Molecular Psychiatry*, *23*(10), 1974–1980. doi: 10.1038/mp.2017.247

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007, May). 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., Goldstein, J., Parlade, M., Hogan, A., Seibert, J., & Mundy, D. (2003, 01). Early social communication scales (escs). , *305*.

Mundy, P., & Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant Behavior and Development*, *21*(3), 469–482. doi: 10.1016/S0163-6383(98)90020-0

Mundy, P., & Newell, L. (2007, October). Attention, Joint Attention, and Social Cognition. *Current Directions in Psychological Science*, *16*(5), 269–274. doi: 10.1111/j.1467-8721.2007.00518.x

Murray, L., & Cooper, P. J. (1996, January). The impact of postpartum depression on child development. *International Review of Psychiatry*, *8*(1), 55–63. doi: 10.3109/09540269609037817

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

Niedźwiecka, A., Ramotowska, S., & Tomalski, P. (2018, November). Mutual Gaze During Early Mother-Infant Interactions Promotes Attention Control Development. *Child Development*, *89*(6), 2230–2244. doi: 10.1111/cdev.12830

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017, October). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38. doi: 10.1016/j.jecp.2017.04.017

Nongrum, R., Thomas, E., Lionel, J., & Jacob, K. S. (2014, April). Domestic Violence as a Risk Factor for Maternal Depression and Neonatal Outcomes: A Hospital-Based Cohort Study. *Indian Journal of Psychological Medicine*, *36*(2), 179–181. doi: 10.4103/0253-7176.130989

Oakes, L. M., Baumgartner, H. A., Barrett, F. S., Messenger, I. M., & Luck, S. J. (2013). Developmental changes in visual short-term memory in infancy: evidence from eye-tracking. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00697

Oakes, L. M., Hurley, K. B., Ross-Sheehy, S., & Luck, S. J. (2011, March).

Developmental changes in infants' visual short-term memory for location. *Cognition*, *118*(3), 293–305. Retrieved 2023-06-05, from https://linkinghub.elsevier.com/retrieve/pii/S0010027710002660 doi: 10.1016/j.cognition.2010.11.007

Oakes, L. M., Ross-Sheehy, S., & Luck, S. J. (2006, September). Rapid development of feature binding in visual short-term memory. *Psychological Science*, *17*(9), 781–787. doi: 10.1111/j.1467-9280.2006.01782.x

Padilla, R., Passos, W. L., Dias, T. L. B., Netto, S. L., & da Silva, E. A. B. (2021, January). A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit. *Electronics*, *10*(3), 279. doi: 10.3390/electronics10030279

Pandith, P., John, S., Bellon-Harn, M. L., & Manchaiah, V. (2021, February). Parental Perspectives on Storybook Reading in Indian Home Contexts. *Early Childhood Education Journal*. doi: 10.1007/s10643-020-01147-0

Papageorgiou, K. A., Smith, T. J., Wu, R., Johnson, M. H., Kirkham, N. Z., & Ronald, A. (2014, July). Individual Differences in Infant Fixation Duration Relate to Attention and Behavioral Control in Childhood. *Psychological Science*, *25*(7), 1371–1379. doi: 10.1177/0956797614531295

Patel, V., Rodrigues, M., & deSouza, N. (2002). Gender, poverty, and postnatal depression: a study of mothers in Goa, India. *The American journal of psychiatry*, *159 1*, 43–7. doi: https://doi.org/10.1176/appi.ajp.159.1.43

Perone, S., & Spencer, J. P. (2013a). Autonomous visual exploration creates developmental change in familiarity and novelty seeking behaviors. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00648

Perone, S., & Spencer, J. P. (2013b, January). Autonomy in Action: Linking the Act of Looking to Memory Formation in Infancy via Dynamic Neural Fields. *Cognitive Science*, *37*(1), 1–60. doi: 10.1111/cogs.12010

Pliatsikas, C., Veríssimo, J., Babcock, L., Pullman, M. Y., Glei, D. A., Weinstein, M., ... Ullman, M. T. (2019, June). Working memory in older adults declines with age, but is modulated by sex and education. *Quarterly Journal of Experimental Psychology*, *72*(6), 1308–1327. doi: 10.1177/1747021818791994

Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., ... Iyengar, S. S. (2019, September). A Survey on Deep Learning: Algorithms, Techniques, and Applications. *ACM Computing Surveys*, *51*(5), 1–36. doi: 10.1145/3234150

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016, June). You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 779–788). Las Vegas, NV, USA: IEEE. doi: 10.1109/CVPR.2016.91

Reuben, A., Moffitt, T. E., Caspi, A., Belsky, D. W., Harrington, H., Schroeder, F., ... Danese, A. (2016, October). Lest we forget: comparing retrospective and prospective assessments of adverse childhood experiences in the prediction of adult health. *Journal of Child Psychology and Psychiatry*, *57*(10), 1103–1112. doi: 10.1111/jcpp.12621

Richards, H. J., Benson, V., Donnelly, N., & Hadwin, J. A. (2014, February). Exploring the function of selective attention and hypervigilance for threat in anxiety. *Clinical Psychology Review*, *34*(1), 1–13. doi: 10.1016/j.cpr.2013.10.006

Riggs, K. J., McTaggart, J., Simpson, A., & Freeman, R. P. (2006, September). Changes in the capacity of visual working memory in 5- to 10-year-olds. *Journal of Experimental Child Psychology*, *95*(1), 18–26. doi: 10.1016/j.jecp.2006.03.009

Riggs, K. J., Simpson, A., & Potts, T. (2011, April). The development of visual short-term memory for multifeature items during middle childhood. *Journal of Experimental Child Psychology*, *108*(4), 802–809. doi: 10.1016/j.jecp.2010.11.006

Rogoff, B., Dahl, A., & Callanan, M. (2018, December). The importance of understanding children's lived experience. *Developmental Review*, *50*, 5–15. doi: 10.1016/j.dr.2018.05.006

Rogoff, B., Mistry, J., Göncü, A., Mosier, C., Chavajay, P., Heath, S. B., & Goncu, A. (1993). Guided Participation in Cultural Activity by Toddlers and Caregivers. *Monographs of the Society for Research in Child Development*, *58*(8), i. doi: 10.2307/1166109

Roopnarine, J. L., Ahmeduzzaman, M., Hossain, Z., & Riegraf, N. B. (1992, October). Parent–Infant Rough Play: Its Cultural Specificity. *Early Education & Development*, *3*(4), 298–311. doi: 10.1207/s15566935eed0304_3

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2001). Attention and recognition memory in the 1st year of life: A longitudinal study of preterm and full-term infants. *Developmental Psychology*, *37*(1), 135–151. doi: 10.1037/0012-1649.37.1.135

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2002). Processing speed in the 1st year of life: A longitudinal study of preterm and full-term infants. *Developmental Psychology*, *38*(6), 895–902. doi: 10.1037/0012-1649.38.6.895

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2005, May). The structure of infant cognition at 1 year. *Intelligence*, *33*(3), 231–250. doi: 10.1016/j.intell.2004.11.002

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2011, September). Modeling a cascade of effects: the role of speed and executive functioning in preterm/full-term differences in academic achievement. *Developmental*

Science, *14*(5), 1161–1175. doi: DOI:10.1111/j.1467-7687.2011.01068.x

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2012, November). Implications of Infant Cognition for Executive Functions at Age 11. *Psychological Science*, *23*(11), 1345–1355. doi: 10.1177/0956797612444902

Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology*, *18*(5), 704–713. doi: 10.1037/0012-1649.18.5.704

Ross-sheehy, S., Oakes, L. M., & Luck, S. J. (2003, November). The Development of Visual Short-Term Memory Capacity in Infants. *Child Development*, *74*(6), 1807–1822. doi: 10.1046/j.1467-8624.2003.00639.x

Saleem, S. M. (2020). Modified Kuppuswamy socioeconomic scale updated for the year 2020. *Indian Journal of Forensic and Community Medicine*, 4.

Salomo, D., & Liszkowski, U. (2013, July). Sociocultural Settings Influence the Emergence of Prelinguistic Deictic Gestures. *Child Development*, *84*(4), 1296–1307. doi: 10.1111/cdev.12026

Savarimuthu, R., Ezhilarasu, P., Charles, H., Antonisamy, B., Kurian, S., & Jacob, K. (2010, January). Post-Partum Depression in the Community: a Qualitative Study From Rural South India. *International Journal of Social Psychiatry*, *56*(1), 94–102. doi: https://doi.org/10.1177/0020764008097

Scaife, M., & Bruner, J. S. (1975, January). The capacity for joint visual attention in the infant. *Nature*, *253*(5489), 265–266. Retrieved from https://doi.org/10.1038/253265a0 doi: 10.1038/253265a0

Scheidecker, G., Oppong, S., Chaudhary, N., & Keller, H. (2021, September). How overstated scientific claims undermine ethical principles in parenting interventions. *BMJ Global Health*, *6*(9), e007323. doi:

10.1136/bmjgh-2021-007323

Schneiberg, S., Sveistrup, H., McFadyen, B., McKinley, P., & Levin, M. F. (2002, September). The development of coordination for reach-to-grasp movements in children. *Experimental Brain Research*, *146*(2), 142–154. doi: 10.1007/s00221-002-1156-z

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012, July). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. doi: 10.1038/nmeth.2089

Sethuraman, K., Lansdown, R., & Sullivan, K. (2006). Women's empowerment and domestic violence: the role of sociocultural determinants in maternal and child undernutrition in tribal and rural communities in south india. *Food and nutrition bulletin*, *27*(2), 128–143.

Shivalli, S., & Gururaj, N. (2015, April). Postnatal Depression among Rural Women in South India: Do Socio-Demographic, Obstetric and Pregnancy Outcome Have a Role to Play? *PLOS ONE*, *10*(4), e0122079. doi: 10.1371/journal.pone.0122079

Shroff, M., Griffiths, P., Adair, L., Suchindran, C., & Bentley, M. (2009, January). Maternal autonomy is inversely related to child stunting in Andhra Pradesh, India. *Maternal & Child Nutrition*, *5*(1), 64–74. doi: 10.1111/j.1740-8709.2008.00161.x

Simmering, V. R. (2012, April). The development of visual working memory capacity during early childhood. *Journal of Experimental Child Psychology*, *111*(4), 695–707. doi: 10.1016/j.jecp.2011.10.007

Simmering, V. R. (2016a). Iii. empirical tests of predictions comparing capacity estimates across tasks and development. *Monographs of the Society for Research in Child Development*, *81*(3), 56–81.

Simmering, V. R. (2016b). I. working memory capacity in context:

Modeling dynamic processes of behavior, memory, and development. *Monographs of the Society for Research in Child Development*, *81*(3), 7-24. Retrieved from https://srcd.onlinelibrary.wiley.com/doi/abs/10.1111/mono.12249 doi: https://doi.org/10.1111/mono.12249

Siposova, B., & Carpenter, M. (2019, August). A new look at joint attention and common knowledge. *Cognition*, *189*, 260–274. doi: 10.1016/j.cognition.2019.03.019

Skotheim, S., Braarud, H. C., Høie, K., Markhus, M. W., Malde, M. K., Graff, I. E., ... Stormark, K. M. (2013, June). Subclinical levels of maternal depression and infant sensitivity to social contingency. *Infant Behavior and Development*, *36*(3), 419–426. doi: 10.1016/j.infbeh.2013.03.009

Smith, K. E., Landry, S. H., & Swank, P. R. (2006, May). The Role of Early Maternal Responsiveness in Supporting School-Aged Cognitive Development for Children Who Vary in Birth Status. *Pediatrics*, *117*(5), 1608–1617. doi: 10.1542/peds.2005-1284

Smith, L. B., Yu, C., & Pereira, A. F. (2011, January). Not your mother's view: the dynamics of toddler visual experience: Dynamics of toddler visual experience. *Developmental Science*, *14*(1), 9–17. doi: 10.1111/j.1467-7687.2009.00947.x

Smith, L. B., Yu, C., Yoshida, H., & Fausey, C. M. (2015, May). Contributions of Head-Mounted Cameras to Studying the Visual Environments of Infants and Young Children. *Journal of Cognition and Development*, *16*(3), 407–419. doi: 10.1080/15248372.2014.933430

Soska, K. C., & Adolph, K. E. (2014, March). Postural Position Constrains Multimodal Object Exploration in Infants. *Infancy*, *19*(2), 138–161. doi: 10.1111/infa.12039

Spencer, J. P. (2020, December). The Development of Working Memory. *Cur-*

rent Directions in Psychological Science, *29*(6), 545–553. doi: 10.1177/0963721420959835

Spencer, J. P., Forbes, S. H., Naylor, S., Singh, V. P., Jackson, K., Deoni, S., ... Kumar, A. (2023, April). Poor air quality is associated with impaired visual cognition in the first two years of life: A longitudinal investigation. *eLife*, *12*, e83876. Retrieved 2023-05-17, from `https://elifesciences.org/articles/83876` doi: 10.7554/eLife.83876

Spencer, J. P., Perone, S., & Johnson, J. S. (2009, June). Dynamic Field Theory and Embodied Cognitive Dynamics. In J. Spencer (Ed.), *Toward a Unified Theory of Development Connectionism and Dynamic System Theory Re-Consider* (1st ed., pp. 86–118). Oxford University PressNew York. doi: 10.1093/acprof:oso/9780195300598.003.0005

Spencer, J. P., & Schoner, G. (2003, September). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science*, *6*(4), 392–412. doi: 10.1111/1467-7687.00295

Spiro, A. M. (2005, April). *Najar* or *Bhut* —Evil eye or ghost affliction: Gujarati views about illness causation. *Anthropology & Medicine*, *12*(1), 61–73. doi: 10.1080/13648470500049867

St Clair-Thompson, H. L., & Gathercole, S. E. (2006, April). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology*, *59*(4), 745–759. doi: 10.1080/17470210500162854

Subramanian, S., Nandy, S., Irving, M., Gordon, D., Lambert, H., & Davey Smith, G. (2006, May). The Mortality Divide in India: The Differential Contributions of Gender, Caste, and Standard of Living Across the Life Course. *American Journal of Public Health*, *96*(5), 818–825. Retrieved 2023-05-18, from `https://ajph.aphapublications.org/doi/`

full/10.2105/AJPH.2004.060103 doi: 10.2105/AJPH.2004.060103

Szucs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2013, November). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex*, *49*(10), 2674–2688. doi: 10.1016/j.cortex.2013.06.007

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001, May). Maternal Responsiveness and Children's Achievement of Language Milestones. *Child Development*, *72*(3), 748–767. doi: 10.1111/1467-8624.00313

Tamis-LeMonda, C. S., Shannon, J. D., Cabrera, N. J., & Lamb, M. E. (2004, December). Fathers and Mothers at Play With Their 2- and 3-Year-Olds: Contributions to Language and Cognitive Development. *Child Development*, *75*(6), 1806–1820. doi: 10.1111/j.1467-8624.2004.00818.x

Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2019, November). Routine Language: Speech Directed to Infants During Home Activities. *Child Development*, *90*(6), 2135–2152. doi: 10.1111/cdev.13089

Team, D. (2014). *Datavyu: A Video Coding Tool.* Databrary Project, New York University. Retrieved from http://datavyu.org

Team, R. C. (2017). R: A language and environment for statistical computing. *R Found. Stat. Comput. Vienna, Austria.*. Retrieved from /www.R-project.org/

Tinker, A., Parker, R., Lord, D., & Grear, K. (2010, January). Advancing newborn health: The Saving Newborn Lives initiative. *Global Public Health*, *5*(1), 28–47. Retrieved 2023-05-18, from https://www.tandfonline.com/doi/full/10.1080/17441690903286572 doi: 10.1080/17441690903286572

Tzutalin. (2015). *Labelimg.* Free Software: MIT License. Retrieved from https://github.com/tzutalin/labelImg

United Nations. (2015). *The 17 Sustainable Development Goals.* Retrieved from https://sdgs.un.org/goals

Vieites, V., & Reeb-Sutherland, B. C. (2017, May). Individual differences in non-clinical maternal depression impact infant affect and behavior during the still-face paradigm across the first year. *Infant Behavior and Development*, *47*, 13–21. doi: 10.1016/j.infbeh.2017.02.005

Viola, P., & Jones, M. J. (2004, May). Robust Real-Time Face Detection. *International Journal of Computer Vision*, *57*(2), 137–154. Retrieved 2023-05-08, from http://link.springer.com/10.1023/B:VISI.0000013087.49260.fb doi: 10.1023/B:VISI.0000013087.49260.fb

Vlach, H. A., & Johnson, S. P. (2013, June). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375–382. doi: 10.1016/j.cognition.2013.02.015

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 92–114. doi: 10.1037/0096-1523.27.1.92

Wagenmakers, E.-J., & Farrell, S. (2004, February). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. doi: 10.3758/BF03206482

Wass, S. V. (2014, August). Comparing methods for measuring peak look duration: Are individual differences observed on screen-based tasks also found in more ecologically valid contexts? *Infant Behavior and Development*, *37*(3), 315–325. doi: 10.1016/j.infbeh.2014.04.007

Wass, S. V., Clackson, K., Georgieva, S. D., Brightman, L., Nutbrown, R., & Leong, V. (2018, November). Infants' visual sustained attention is higher during joint play than solo play: is this due to increased endogenous attention control or exogenous stimulus capture? *Developmental Science*, *21*(6), e12667. doi: 10.1111/desc.12667

Weber, A. M., Diop, Y., Gillespie, D., Ratsifandrihamanana, L., & Darmstadt, G. L. (2021, July). Africa is not a museum: the ethics of encouraging new parenting practices in rural communities in low-income and middle-income countries. *BMJ Global Health*, *6*(7), e006218. doi: 10.1136/bmjgh-2021-006218

Werrett, J., & Clifford, C. (2006, February). Validation of the Punjabi version of the Edinburgh postnatal depression scale (EPDS). *International Journal of Nursing Studies*, *43*(2), 227–236. doi: 10.1016/j.ijnurstu.2004.12.007

West, K. L., & Iverson, J. M. (2017, July). Language learning is hands-on: Exploring links between infants' object manipulation and verbal input. *Cognitive Development*, *43*, 190–200. doi: 10.1016/j.cogdev.2017.05.004

White, H., Heck, A., Jubran, R., Chroust, A., & Bhatt, R. S. (2022, May). Average fixation duration in infancy: Stability and predictive utility. *Infancy*, infa.12483. doi: 10.1111/infa.12483

Wijeakumar, S., Kumar, A., Delgado Reyes, L. M., Tiwari, M., & Spencer, J. P. (2019, September). Early adversity in rural India impacts the brain networks underlying visual working memory. *Developmental Science*, *22*(5). doi: 10.1111/desc.12822

Willis, J. R., Kumar, V., Mohanty, S., Kumar, A., Singh, J. V., Ahuja, R. C., ... for the Saksham Study Group (2011, August). Utilization and perceptions of neonatal healthcare providers in rural Uttar Pradesh, India. *International Journal for Quality in Health Care*, *23*(4), 487–494. Retrieved 2023-05-18, from https://academic.oup.com/intqhc/article-lookup/doi/10.1093/intqhc/mzr030 doi: 10.1093/intqhc/mzr030

Xu, T., Chen, Y., & Smith, L. (2011). It's the child's body: The role of toddler and parent in selecting toddler's visual experience. In *2011 ieee international conference on development and learning (icdl)* (Vol. 2, pp. 1–6).

Yoshida, H., & Smith, L. B. (2008, May). What's in View for Toddlers? Using a Head Camera to Study Visual Experience. *Infancy*, *13*(3), 229–248. doi: 10.1080/15250000802004437

Yovsi, R. D., & Keller, H. (2007). The Architecture of Cosleeping among Wage-Earning and Subsistence Farming Cameroonian Nso Families. *Ethos*, *35*(1), 65–84. (Publisher: [American Anthropological Association, Wiley])

Yu, C., & Smith, L. (2016, May). The Social Origins of Sustained Attention in One-Year-Old Human Infants. *Current Biology*, *26*(9), 1235–1240. doi: 10.1016/j.cub.2016.03.026

Yu, C., & Smith, L. B. (2012, November). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262. doi: 10.1016/j.cognition.2012.06.016

Yu, C., & Smith, L. B. (2013, November). Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLoS ONE*, *8*(11), e79659. doi: 10.1371/journal.pone.0079659

Yu, C., Suanda, S. H., & Smith, L. B. (2019, January). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, *22*(1), e12735. doi: 10.1111/

desc.12735

Yu, C., Yurovsky, D., & Xu, T. (2012, January). Visual Data Mining: An Exploratory Approach to Analyzing Temporal Patterns of Eye Movements: VISUAL DATA MINING OF GAZE DATA. *Infancy*, *17*(1), 33–60. doi: 10.1111/j.1532-7078.2011.00095.x

Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018, October). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, *2*(10), 719–731. doi: 10.1038/s41551-018-0305-z

Yuan, L., Xu, T. L., Yu, C., & Smith, L. B. (2019, March). Sustained visual attention is more than seeing. *Journal of Experimental Child Psychology*, *179*, 324–336. doi: 10.1016/j.jecp.2018.11.020

Yurovsky, D., Smith, L. B., & Yu, C. (2013, March). Statistical word learning at scale: the baby's view is better. *Developmental Science*, n/a–n/a. doi: 10.1111/desc.12036

Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C.-L., & Grundmann, M. (2020). MediaPipe Hands: On-device Real-time Hand Tracking.
(Publisher: arXiv Version Number: 1) doi: 10.48550/ARXIV.2006.10214

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016, October). Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters*, *23*(10), 1499–1503. doi: 10.1109/LSP.2016.2603342

Zhang, L., Ssewanyana, D., Martin, M.-C., Lye, S., Moran, G., Abubakar, A., ... Malti, T. (2021, July). Supporting Child Development Through Parenting Interventions in Low- to Middle-Income Countries: An Updated Systematic Review. *Frontiers in Public Health*, *9*, 671988. Retrieved 2023-05-26, from https://www.frontiersin.org/articles/10.3389/fpubh.2021.671988/full doi: 10.3389/fpubh.2021.671988

Zhang, N., Luo, J., & Gao, W. (2020, September). Research on Face Detection Technology Based on MTCNN. In *2020 International Conference on Computer Network, Electronic and Automation (ICCNEA)* (pp. 154–158). Xi'an, China: IEEE. doi: 10.1109/ICCNEA50255.2020.00040

Zuiderveld, K. (1994). Contrast Limited Adaptive Histogram Equalization. In *Graphics Gems* (pp. 474–485). Elsevier. doi: 10.1016/B978-0-12-336156-1.50061-6

Appendix A

Appendix

SES_Demographic Form

question_id	question_english	type	option_values
rec_id	Record ID		
IndiaCode	IndiaCode		
GroupNo	GroupNo		
FormName	FormName		
SESTestDate	SES Test Date	Date	
DataCollectorID	DataCollectorID	Select	
VillageID	VillageID		
GPPRICode	GPPRICode		
HamletID	HamletID		
BabyBirthDate	Baby's birth date		
CalendarAge	Calendar Age		
BirthPlace	Delivery Place	Select	govt_health_facility, private_health_facility, home, others, dont_Know
BirthPlaceOth	Anyother delivery place	Text	
BabyStatus	What is baby's status?	Select	alive, dead
BabyDeathDate	Baby's Death Date	Date	
BabyGender	Baby's gender	Select	boy, girl
MotherEducation	What is mother's formal Highest education?	Select	illiterate, below_primary, primary_pass, eight_pass, highschool_pass, inter_pass, polytechnic_or_diploma, graduate, postgraduate, Ph.D
FatherEducation	What is father's formal Highest education?	Select	illiterate, below_primary, primary_pass, eight_pass, highschool_pass, inter_pass, polytechnic_or_diploma, graduate, postgraduate, Ph.D
SESEduStatus	SEE status as per mother and father education		
MotherStatus	What is status of mother?	Select	alive, dead
MotherKnowDOB	Does mother know her DOB?	Select	yes, no
MotherDOB	What is Date of birth of mother?	Date	
MotherAge	What is mother's age (in completed years)?	Int	

Page 1

Figure A.1: Demographic questionnaire to calculate socioeconomic status in India (page1).

SES_Demographic Form

MotherOccupation	What is the mother's current occupation?	Select	housewife, self_employed, government_employee, private_employee, farmer_(own_land), farmer_(leased_land), skilled_labour, unskilled_labour, traditional_worker, others
MotherOccupationOther	If other, then specify	Text	
PrimaryCaregiver	Who is the primary caregiver?	Select	grandparents, father, siblings/cousin, other_family_member, orphanage/aanganwadi, neighbours, others
PrimaryCaregiverOther	If other, then specify	Text	
FatherAge	What is father's age?	Int	
FatherOccupation	What is the father's current occupation?	Select	unemployed, selfemployed, government_employee_, non_government_employee, farmer_(own_land), farmer_(leased_land), skilled_labour_, unskilled_labour, _traditional_worker, others
FatherOccupationOther	If any other, then specify	Text	
FatherStay		Select	yes, no
Religion	What is your religion?	Select	hindu, muslim, others
FatherReligionOther	If any other religion, then specify	Text	
Caste	To what tribe/caste do you belong?	Select	general, OBC, sc/st
FamilyType	What type of family is yours?	Select	nuclear_family, joint_family
FamilyMembers	How many members are there in your family?	Int	
ChildrenUnder5	How many children under 5 are there in your family?	Int	
SleepingRooms	How many rooms in your house are used for sleeping?	Int	
MobileUsers	How many members of the family have a mobile phone?	Int	

Page 2

SES_Demographic Form

SmartPhoneUsers	How many members of the family have a smart phone?	Int	
MotherHasSmartPhone	Does mother has smartphone?	Select	yes, no
RoofMtl	What is the main material that the roof of your household is made of?	Select	natural_(thatch_grass), rudimentary(bamboo_wood), metal_asbestos, finished_(cement_tiles), others
RoofMtlOther	If any other material is used, then specify	Text	
FloorMtl	What is the main material that the floor of your household is made of?	Select	Natural_(mud_cowdung), Finished_(cement_tiles_brick_polished_wood), Others
FloorMtlOther	If other, then specify	Text	
WallMtl	What is the main material that the walls of your household are made of?	Select	natural_(thatch_grass), rudimentary_(bamboo_wood, finished_(cement_tiles), others
WallMtlOther	If other, then specify	Text	
WaterSource	What is the main source of drinking water for your family members?	Select	underground_water, piped_water, surface_water, tanker_truck, bottled_water, well, others
WaterSourceOther	If any other source, then specify	Text	
ToiletType	What type of toilet does your household use?	Select	in_open, flush_toilet, pit_toilet, common_toilet, others
ToiletTypeOther	If other, then specify	Text	
HouseholdOwnership	Do you own the house you live in?	Select	own_house, rented
Electricity	Does your household have electricity connection?	Select	yes, no
ElectricalEquipment	Select what your household have among the following	MSelect	TV, fridge, cooler, washing_machine, AC, inverter_or_generator, fan, none_of_above, others
ElectricityEquipOther	If other, then specify	Text	
AgriculturalLandOwnership	Does any member of your family own agricultural land?	Select	yes, no

Page 3

Figure A.2: Demographic questionnaire to calculate socioeconomic status in India (page 2 and 3).

SES_Demographic Form

AgricultureLandType	What do you have in cultivable land?	Select	beegha, biswa, BeeghaAndBiswa
Bigha	Area of cultivable land (Bigha)	Int	
Biswa	Area of cultivable land (Biswa)	Int	
MotorAssets	Do you have any of the following in your possession?	MSelect	two_wheeler, tractor, four_wheeler, pump_set, cycle, none_of_above
EconomicAnimals	Is there any economic animal in your house?	Select	yes, no
CowBuffalo	CowBuffaloOx	Int	
Goat	Goat	Int	
Horse	Horse	Int	
Pig	Pig	Int	
sheep	Sheep	Int	
HenDuck	HenDuck	Int	
others	Other Animals	Int	
EconomicAnimalsother	EconomicAnimalsother	Text	
CookingFuel	What type of fuel does your household use for cooking?	Select	LPG, wood, cowdung_cake, husk, charcoal, biogas, kerosene, solar_cooker, others
CookingFuelOther	If other, then specify	Text	
BankAccount	Does any member of your household have a bank account or post office account?	Select	yes, no
Income	What is your family's income?	Int	
Remarks		Text	
IncomeScore	Income Score calculated for Kuppuswamy Scale	Int	
MotherEduScore	Mother Education Score calculated for Kuppuswamy Scale	Int	
HusbandEduScore	Father Education Score calculated for Kuppuswamy Scale	Int	
EducationScoreAvgRound	Average education of mother and father Score calculated for Kuppuswamy Scale	Int	

Page 4

SES_Demographic Form

OccupationScore	Father occupation Score calculated for Kuppuswamy Scale	Int
SES_Score	Sum of income, education avg, occupation Score calculated for Kuppuswamy Scale	Int
Kuppuswamy_Scale	Kuppuswamy Scale based on SES	Text

Page 5

Figure A.3: Demographic questionnaire to calculate socioeconomic status in India (page 4 and 5).

Appendix B

Perinatal Services BC
An agency of the Provincial Health Services Authority

Edinburgh Perinatal/Postnatal Depression Scale (EPDS)

For use between **28–32 weeks** in **all** pregnancies and **6–8 weeks** postpartum

Name: ______________________ Date: ______________ Gestation in Weeks: ______

As you are having a baby, we would like to know how you are feeling. Please mark "X" in the box next to the answer which comes closest to how you have felt in the **past 7 days**–not just how you feel today.

In the past 7 days:

1. I have been able to laugh and see the funny side of things
   - 0 ☐ As much as I always could
   - 1 ☐ Not quite so much now
   - 2 ☐ Definitely not so much now
   - 3 ☐ Not at all
2. I have looked forward with enjoyment to things
   - 0 ☐ As much as I ever did
   - 1 ☐ Rather less than I used to
   - 2 ☐ Definitely less than I used to
   - 3 ☐ Hardly at all
3. I have blamed myself unnecessarily when things went wrong
   - 3 ☐ Yes, most of the time
   - 2 ☐ Yes, some of the time
   - 1 ☐ Not very often
   - 0 ☐ No, never
4. I have been anxious or worried for no good reason
   - 0 ☐ No, not at all
   - 1 ☐ Hardly ever
   - 2 ☐ Yes, sometimes
   - 3 ☐ Yes, very often
5. I have felt scared or panicky for no very good reason
   - 3 ☐ Yes, quite a lot
   - 2 ☐ Yes, sometimes
   - 1 ☐ No, not much
   - 0 ☐ No, not at all
6. Things have been getting on top of me
   - 3 ☐ Yes, most of the time I haven't been able to cope
   - 2 ☐ Yes, sometimes I haven't been coping as well as usual
   - 1 ☐ No, most of the time I have coped quite well
   - 0 ☐ No, I have been coping as well as ever
7. I have been so unhappy that I have had difficulty sleeping
   - 3 ☐ Yes, most of the time
   - 2 ☐ Yes, sometimes
   - 1 ☐ Not very often
   - 0 ☐ No, not at all
8. I have felt sad or miserable
   - 3 ☐ Yes, most of the time
   - 2 ☐ Yes, quite often
   - 1 ☐ Not very often
   - 0 ☐ No, not at all
9. I have been so unhappy that I have been crying
   - 3 ☐ Yes, most of the time
   - 2 ☐ Yes, quite often
   - 1 ☐ Only occasionally
   - 0 ☐ No, never
10. The thought of harming myself has occurred to me
    - 3 ☐ Yes, quite often
    - 2 ☐ Sometimes
    - 1 ☐ Hardly ever
    - 0 ☐ Never

Total Score ☐

Talk about your answers to the above questions with your health care provider.

Translations for care-provider use available on PSBC website: perinatalservicesbc.ca.

The Royal College of Psychiatrists 1987. From Cox, JL, Holden, JM, Sagovsky, R (1987). Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. British Journal of Psychiatry. 150, 782–786. Reprinted with permission.

Figure B.1: Edingburgh postnatal depression scale in English.

Baby IND Code: _______________ Date of administration: ___/___/_______

एडिन्बरा मानसिक विषाद प्रश्नावली

पिछले ७ दिनों में आपने जैसा अनुभव किया, उसके आधार पर हमें बताएँ कि इन दिनों आप कैसा महसूस कर रही हैं।

1. मैं रोज़मर्रा की ज़िंदगी में हँसी-मज़ाक़ का आनंद ले पा रही हूँ
   - ☐ उतना ही जितना कि हमेशा से आनंद ले पाती थी (पहले की अपेक्षा कोई बदलाव नहीं)
   - ☐ अब उतना ज़्यादा आनंद नहीं ले पाती हूँ जैसा कि पहले था
   - ☐ अब काफ़ी कम आनंद ले पाती हूँ
   - ☐ अब बिलकुल नहीं
2. मैं प्रसन्नता और सुखद अपेक्षा से आने वाले समय के बारे में सोचती हूँ
   - ☐ उतना ही जितना हमेशा से सोचती थी (पहले की अपेक्षा कोई बदलाव नहीं)
   - ☐ अब उतना नहीं जैसा कि पहले था
   - ☐ अब पहले से काफ़ी कम
   - ☐ नहीं के बराबर
3. जब कुछ भी बिगड़ता है तो मैं अनावश्यक रूप से अपने आप को दोषी मानती हूँ
   - ☐ हाँ, ज़्यादातर
   - ☐ हाँ, कभी कभी
   - ☐ हाँ, पर ऐसा कम बार होता है
   - ☐ नहीं, कभी नहीं
4. मुझे बिना किसी वजह/कारण के घबराहट या चिंता होती है
   - ☐ नहीं, कभी नहीं
   - ☐ ऐसा हुआ है, पर बहुत कम बार
   - ☐ हाँ, कभी कभी
   - ☐ हाँ, काफ़ी बार
5. मुझे बिना किसी वजह/कारण के डर या आतंक महसूस होता है
   - ☐ हाँ, काफ़ी बार
   - ☐ हाँ, कभी कभी
   - ☐ हाँ, पर ऐसा कम बार होता है
   - ☐ नहीं, कभी नहीं
6. ऐसा लगता है कि ज़िम्मेदारियों और काम के बोझ से मैं दबती चली जा रही हूँ, और अपने हालात का मैं ठीक प्रकार से सामना नहीं कर पा रही हूँ - अपने आप को असमर्थ और बेबस/ असहाय/ कमज़ोर महसूस कर रही हूँ
   - ☐ हाँ, ज़्यादातर समय
   - ☐ हाँ, कभी कभी ऐसा होता है
   - ☐ हाँ, पर ऐसा कम बार होता है
   - ☐ नहीं, कभी नहीं

Figure B.2: Edinburgh postnatal depression scale in Hindi.

Appendix C

VWM of Mother

अब कुछ समय के लिए, हम आपको टीवी पर दाएं और बाएं किनारे पर चित्र वाले एक वीडियो दिखाएंगे। हम चाहते हैं कि आप इन चित्रों को ध्यान से देखें। वीडियो पूरा होने के बाद हम आपसे चित्र के बारे में कुछ प्रश्न पूछेंगे। जब आप वीडियो देखेंगी तो हम इस कैमरे से देखेंगे कि आप कहां देख रही हैं। उसे देखने के लिए हम आपके माथे पर स्टिकर लगायंगे।
क्या आपका कोई सवाल है?

विडियो देखने के बाद यह प्रश्न पूछे:

प्रश्न 1: आपको यह जो चित्र (विडियो) दिखाया गया वह कैसा लगा?
प्रश्न 2: क्या आपने दोनों तरफ के चित्रों के बीच में कोई अंतर देखा?
प्रश्न 3: जब चित्र आते हैं तो कभी में 2 रंग आते हैं कभी 4 रंग आते हैं कभी 6 रंग आते हैं क्या उनमे आपको कुछ अंतर लगा?
प्रश्न 4: क्या आप इस विडियो को दुबारा देखना चाहेंगी? यदि हाँ तो क्यों? यदि नहीं तो क्यों नहीं?
प्रश्न 5: क्या इस विडियो के बारे में कुछ और भी बताना चाहेंगी?
प्रश्न 6: क्या विडियो देखने के दौरान आप कुछ और भी सोच रही थी?

Baby ID:________________ Date of Test:_____________________

Ans-1: ____
____
____

Ans-2: ____
____
____

Ans-3: ____
____
____

Figure C.1: Instructions for the caregivers VWM task in Hindi.

Appendix D

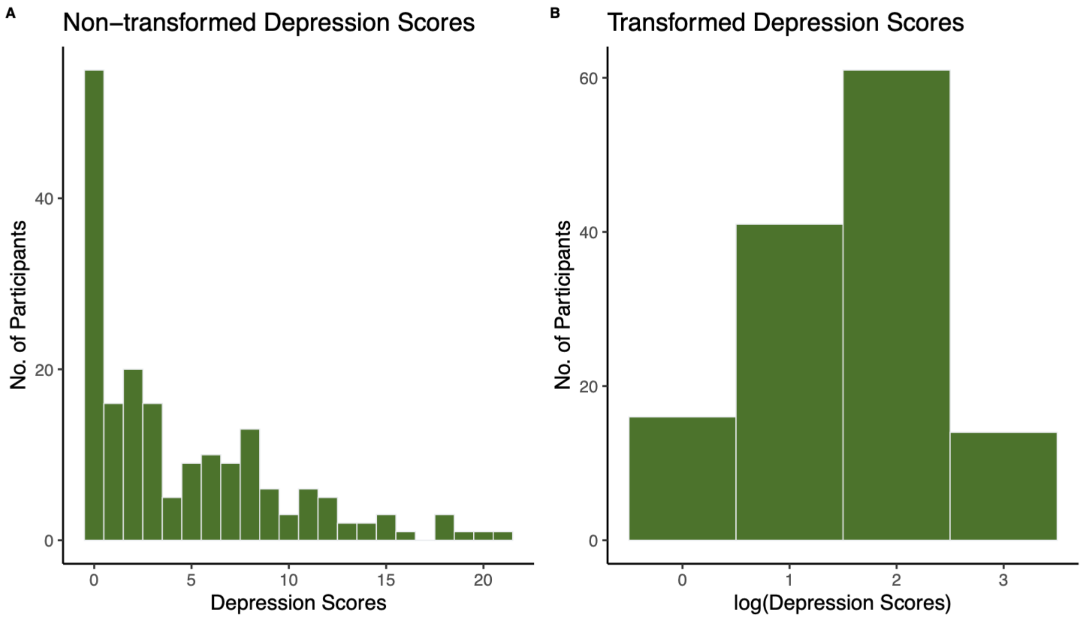


Figure D.1: Frequency histogram of Depression scores before (A) and after (B) logarithmic transformation of Depression scores.

Appendix E

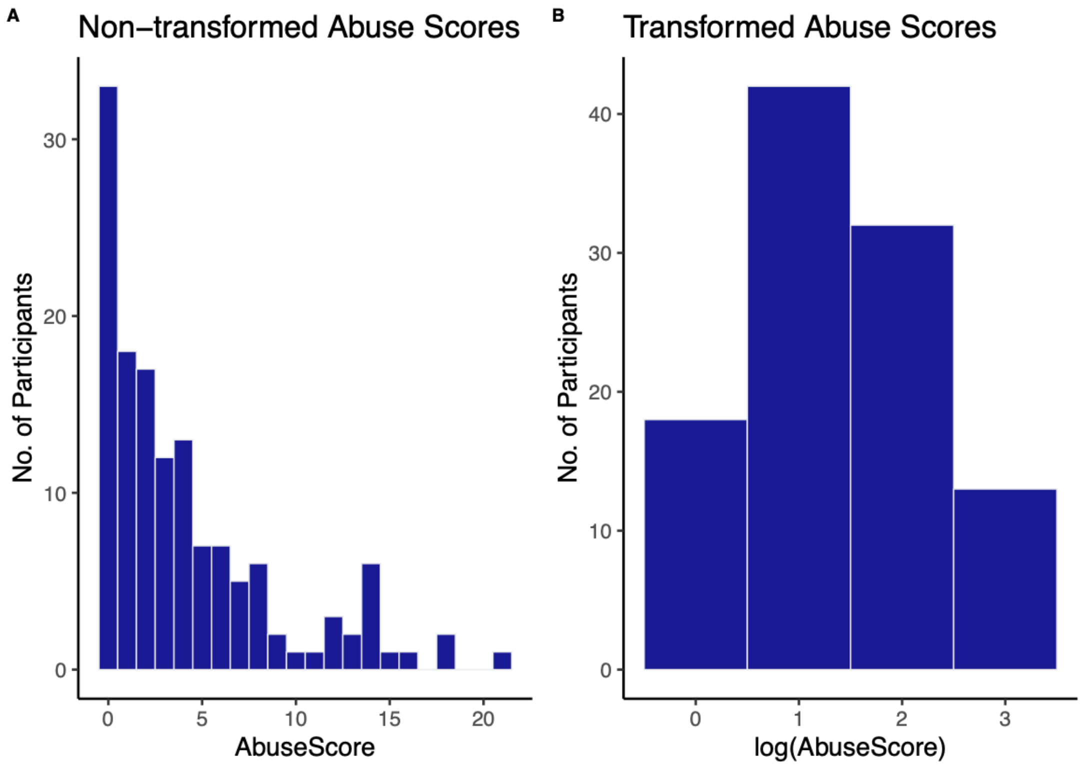


Figure E.1: Frequency histogram of Abuse scores before (A) and after (B) logarithmic transformation of Abuse scores.

Appendix F

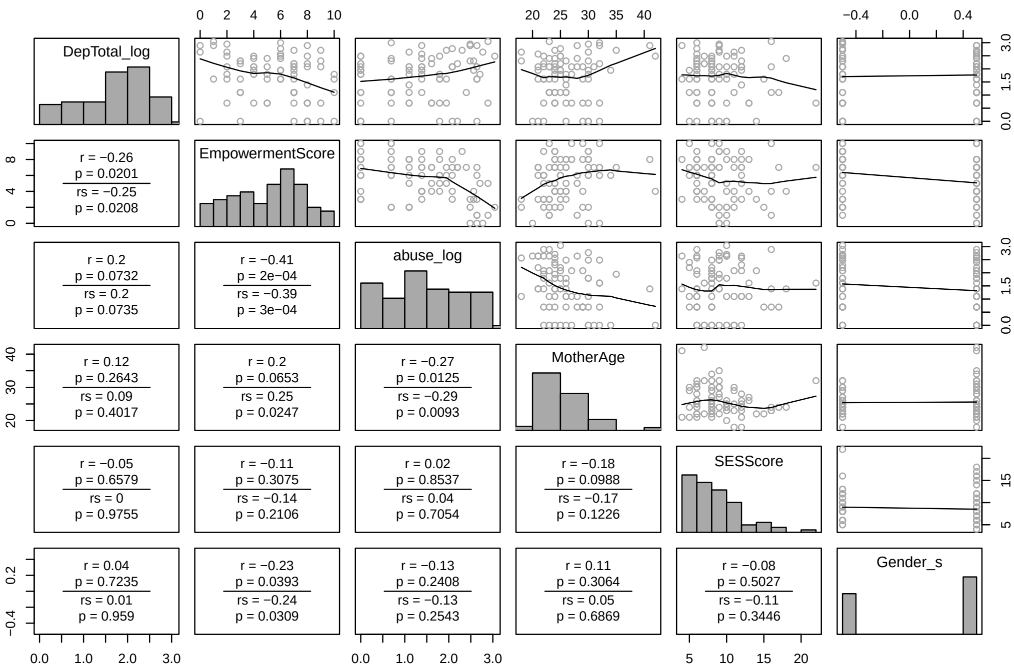


Figure F.1: Correlation matrix for outcome variables, i.e., EPDS, Empowerment, and IPV and each of their predictors, i.e., Mothers' Age, SES Score, and Infants' Gender. Plots include histograms and scatterplots with Pearson (r) and Spearman (rs) correlations in the lower triangle with p-values.

Appendix G

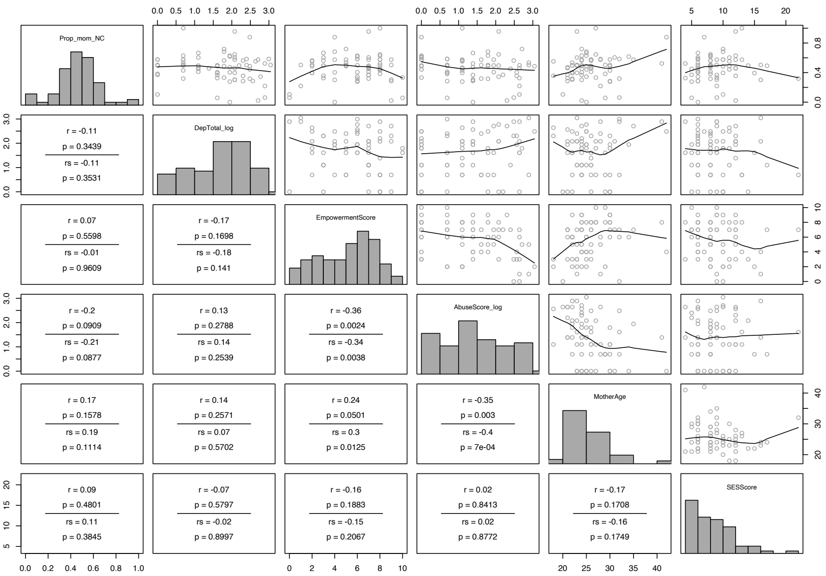


Figure G.1: Correlation matrix for the outcome variable, i.e., Proportion of looking to change side at **Low Load** condition and each of their predictors, i.e., EPDS, Empowerment Score, Abuse Score, Mothers' Age, and SES Score. Plots include histograms and scatterplots with Pearson (r) and Spearman (rs) correlations in the lower triangle with p-values.

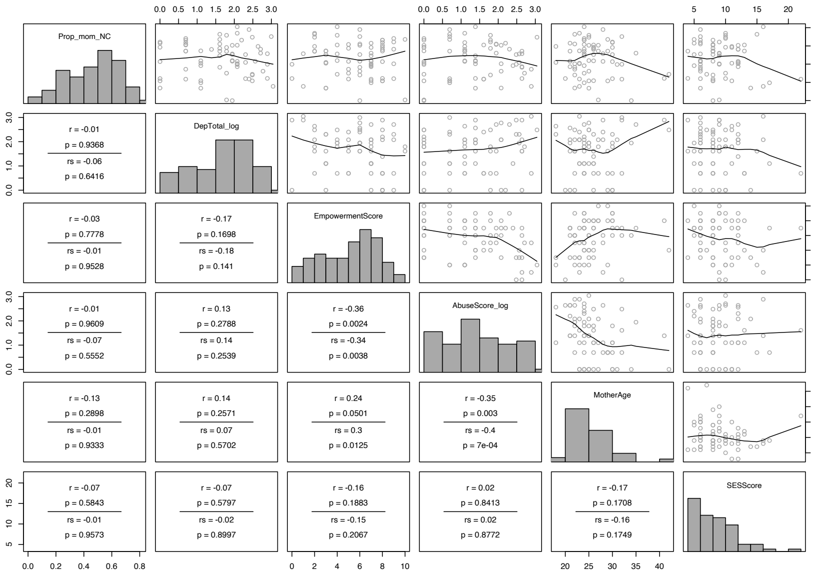


Figure G.2: Correlation matrix for the outcome variable, i.e., Proportion of looking to change side at **Medium Load** condition and each of their predictors, i.e., EPDS, Empowerment Score, Abuse Score, Mothers' Age, and SES Score. Plots include histograms and scatterplots with Pearson (r) and Spearman (rs) correlations in the lower triangle with p-values.

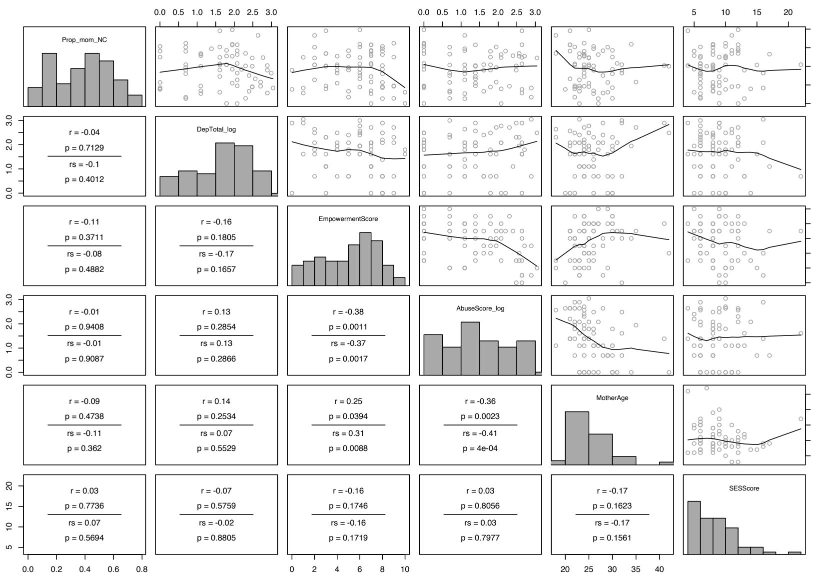


Figure G.3: Correlation matrix for the outcome variable, i.e., Proportion of looking to change side at **High Load** condition and each of their predictors, i.e., EPDS, Empowerment Score, Abuse Score, Mothers' Age, and SES Score. Plots include histograms and scatterplots with Pearson (r) and Spearman (rs) correlations in the lower triangle with p-values.

Appendix H

Table H.1: Results for Model 1: Maternal VWM-Scores with the main effect of and 3-way interaction between Load, SES Score, and Mothers' Age as predictors

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	1377.39	1	<.001 ***
Load	14.16	2	<.001 ***
SESScore_s	0.05	1	.83
MotherAge_s	0.03	1	.87
Load:SESScore_s	2.08	2	.35
Load:MotherAge_s	8.01	2	<.05 *
SESScore_s:MotherAge_s	0.01	1	.93
Load:SESScore_s:MotherAge_s	0.76	2	.68

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table H.2: Results for Model 2: Maternal VWM-Scores with the main effect of Load, SES Score, and Mothers' Age as predictors

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	1389.21	1	<.001 ***
Load	13.93	2	<.001 ***
SESScore_s	0.04	1	.84
MotherAge_s	0.03	1	.88

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table H.3: Results for model comparison

Model	AIC	BIC	logLik	deviance	Chisq	Df	p-value
Model2	-170.04	-142.67	92.02	-184.04			
Model3	-175.37	-140.17	96.63	-193.37	9.32	2	<.01 **
Model1	-168.20	-113.45	98.10	-196.20	2.84	5	.73

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Appendix I

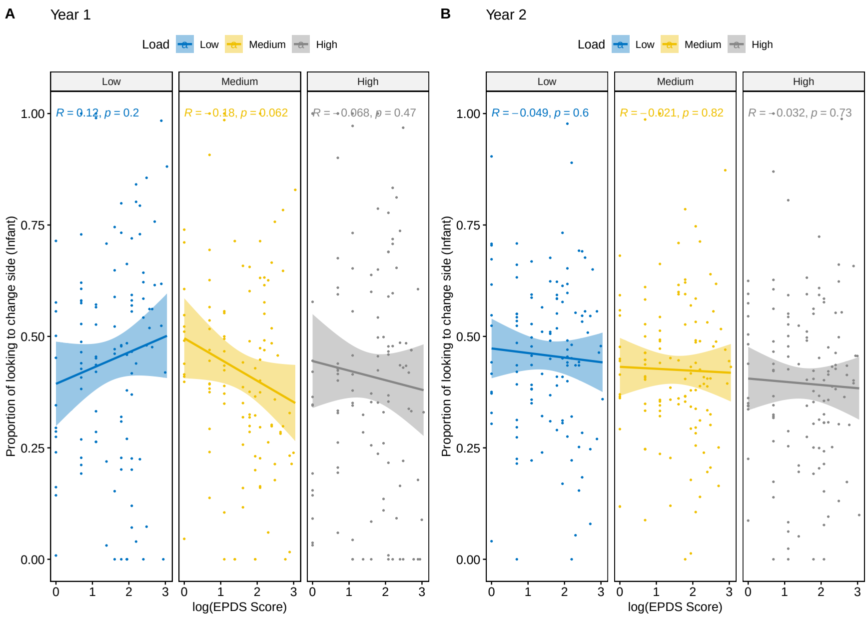


Figure I.1: Scatterplot for the outcome variable, i.e., Infants' proportion of looking to change side for year 1 (A) and year 2 (B) by Load and their predictor i.e., Maternal EPDS scores. Plots include pairwise correlation coefficients.

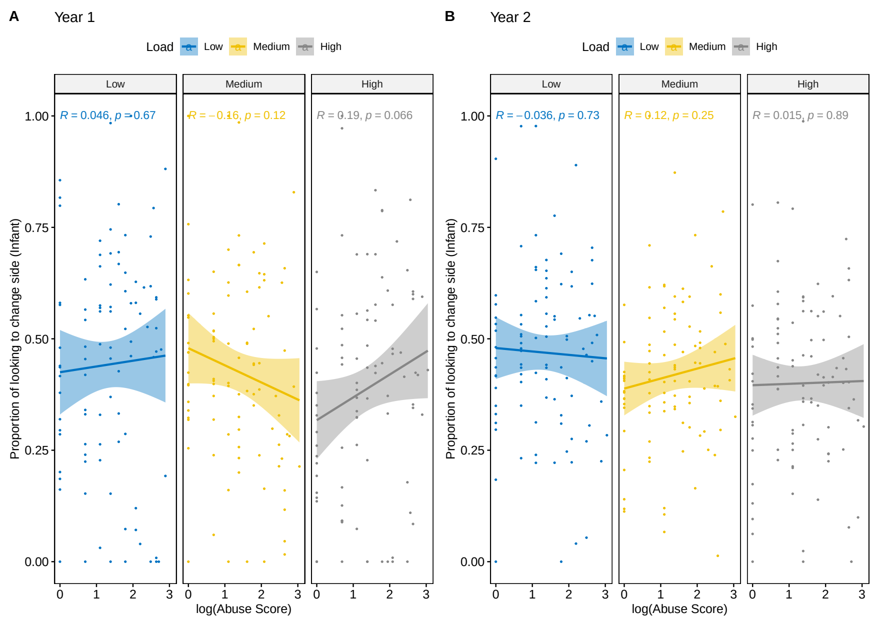


Figure I.2: Scatterplot for the outcome variable, i.e., Infants' proportion of looking to change side for year 1 (A) and year 2 (B) by Load and their predictor i.e., Maternal IPV scores. Plots include pairwise correlation coefficients.

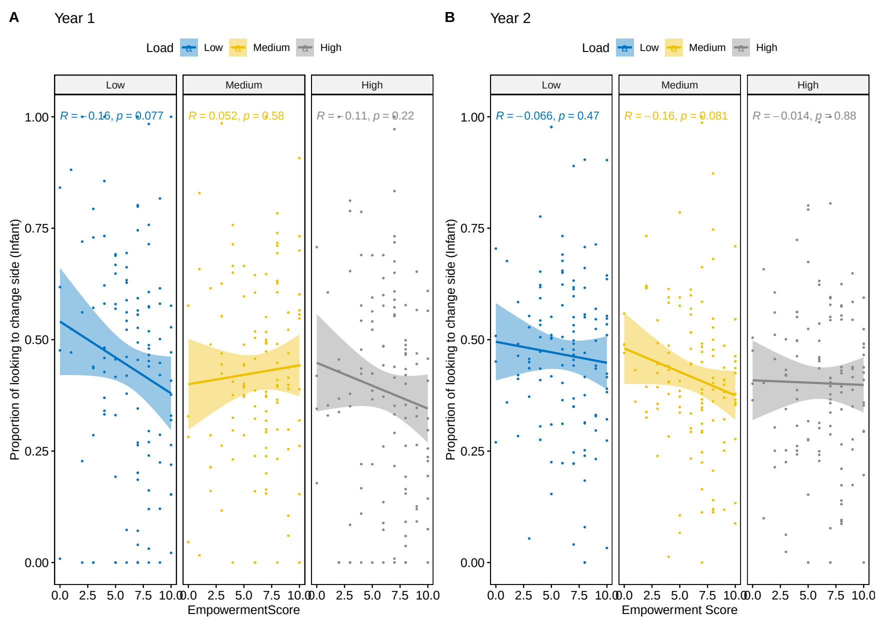


Figure I.3: Scatterplot for the outcome variable, i.e., Infants' proportion of looking to change side for year 1 (A) and year 2 (B) by Load and their predictor i.e., Maternal Sense of Empowerment scores. Plots include pairwise correlation coefficients.

Appendix J

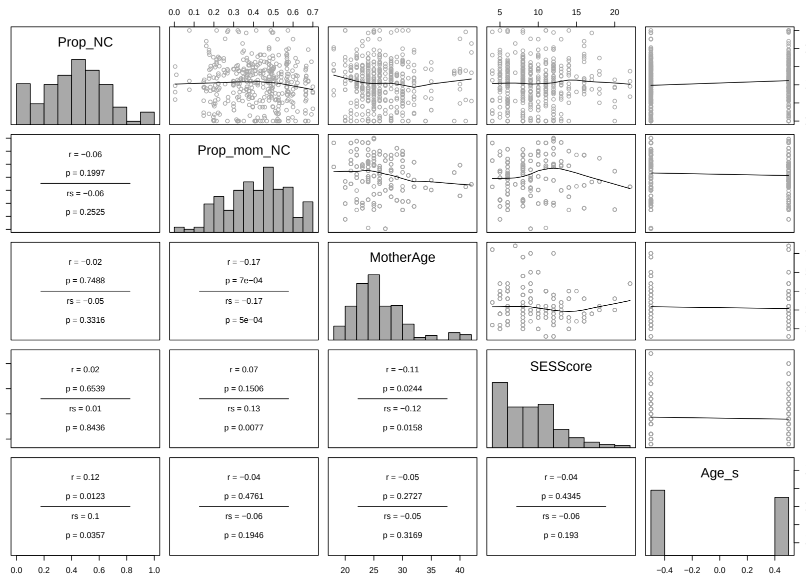


Figure J.1: Correlation matrix for the outcome variable, i.e., Proportion of looking to change side at **Low Load condition in Year 1** and each of their predictors, i.e., Mothers' VWM-PL Scores (aggregated), Mothers' Age, SES Score, and Infant age (6 months old corresponds to - 0.5). Plots include histograms and scatterplots with Pearson (r) and Spearman (rs) correlations in the lower triangle with p-values.

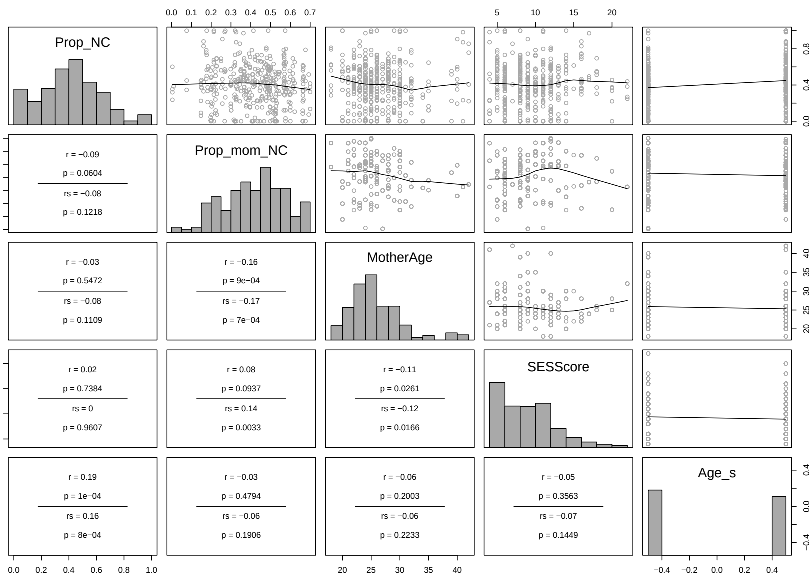


Figure J.2: Correlation matrix for the outcome variable, i.e., Proportion of looking to change side at **Medium Load condition in Year 1** and each of their predictors, i.e., Mothers' VWM-PL Scores (aggregated), Mothers' Age, SES Score, and Infant age (6 months old corresponds to - 0.5). Plots include histograms and scatterplots with Pearson (r) and Spearman (rs) correlations in the lower triangle with p-values.

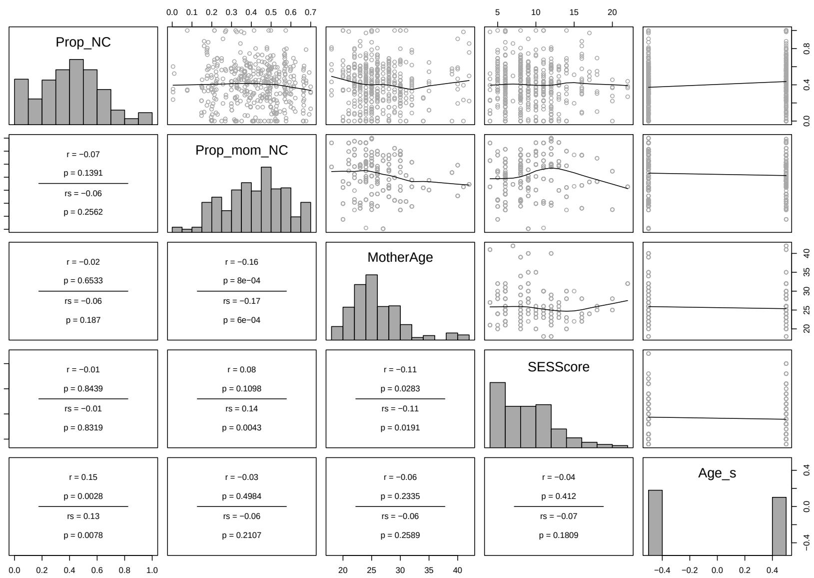


Figure J.3: Correlation matrix for the outcome variable, i.e., Proportion of looking to change side at **High Load condition in Year 1** and each of their predictors, i.e., Mothers' VWM-PL Scores (aggregated), Mothers' Age, SES Score, and Infant age (6 months old corresponds to - 0.5). Plots include histograms and scatterplots with Pearson (r) and Spearman (rs) correlations in the lower triangle with p-values.

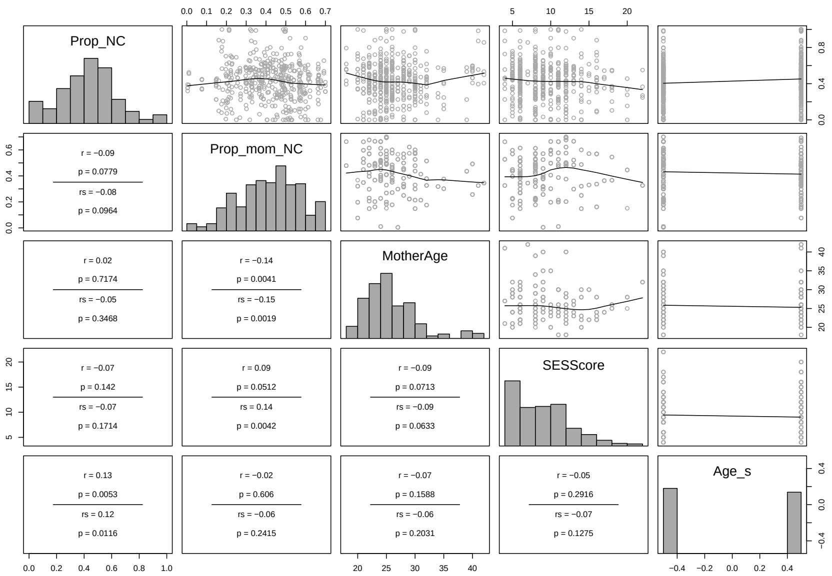


Figure J.4: Correlation matrix for the outcome variable, i.e., Proportion of looking to change side at **Low Load condition in Year 2** and each of their predictors, i.e., Mothers' VWM-PL Scores (aggregated), Mothers' Age, SES Score, and Infant age (6 months old corresponds to - 0.5). Plots include histograms and scatterplots with Pearson (r) and Spearman (rs) correlations in the lower triangle with p-values.

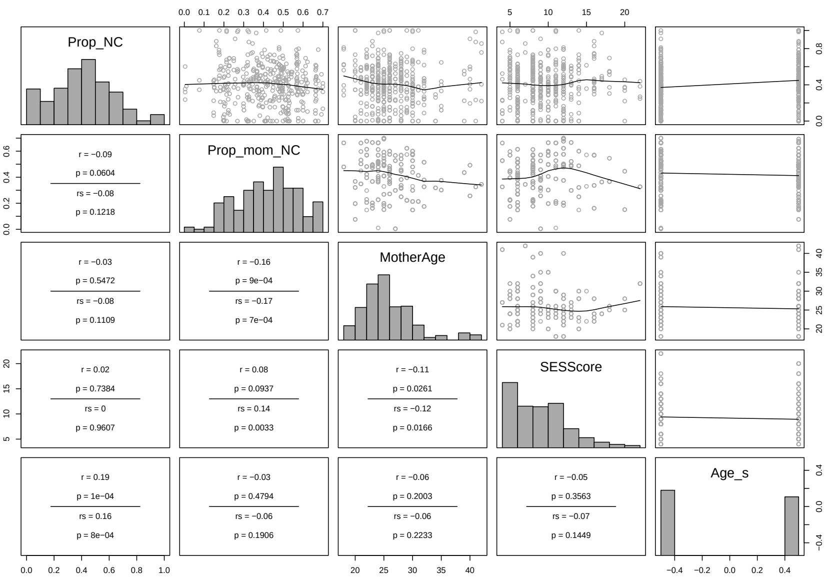


Figure J.5: Correlation matrix for the outcome variable, i.e., Proportion of looking to change side at **Medium Load condition in Year 2** and each of their predictors, i.e., Mothers' VWM-PL Scores (aggregated), Mothers' Age, SES Score, and Infant age (6 months old corresponds to - 0.5). Plots include histograms and scatterplots with Pearson (r) and Spearman (rs) correlations in the lower triangle with p-values.

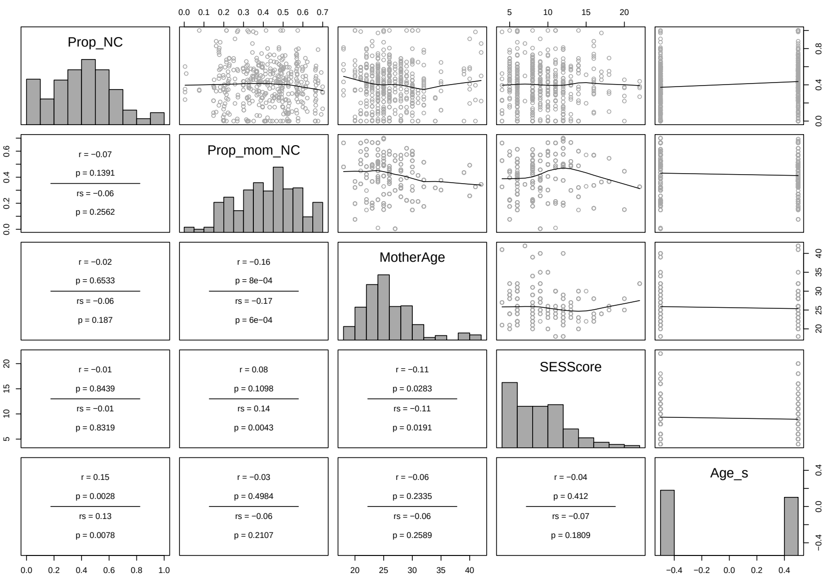


Figure J.6: Correlation matrix for the outcome variable, i.e., Proportion of looking to change side at **High Load condition in Year 2** and each of their predictors, i.e., Mothers' VWM-PL Scores (aggregated), Mothers' Age, SES Score, and Infant age (6 months old corresponds to - 0.5). Plots include histograms and scatterplots with Pearson (r) and Spearman (rs) correlations in the lower triangle with p-values.

Appendix K

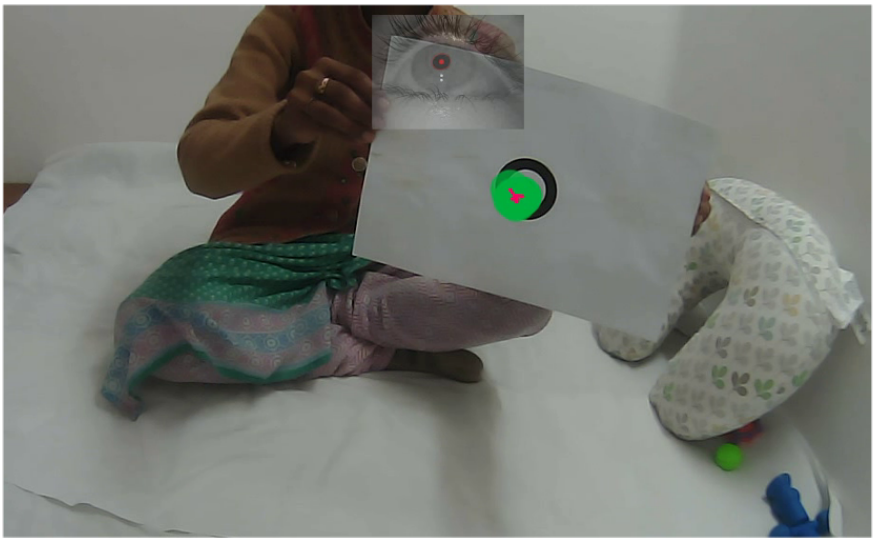


Figure K.1: An example of mapping gaze point (green circle) on the calibration marker during the post-hoc calibration process

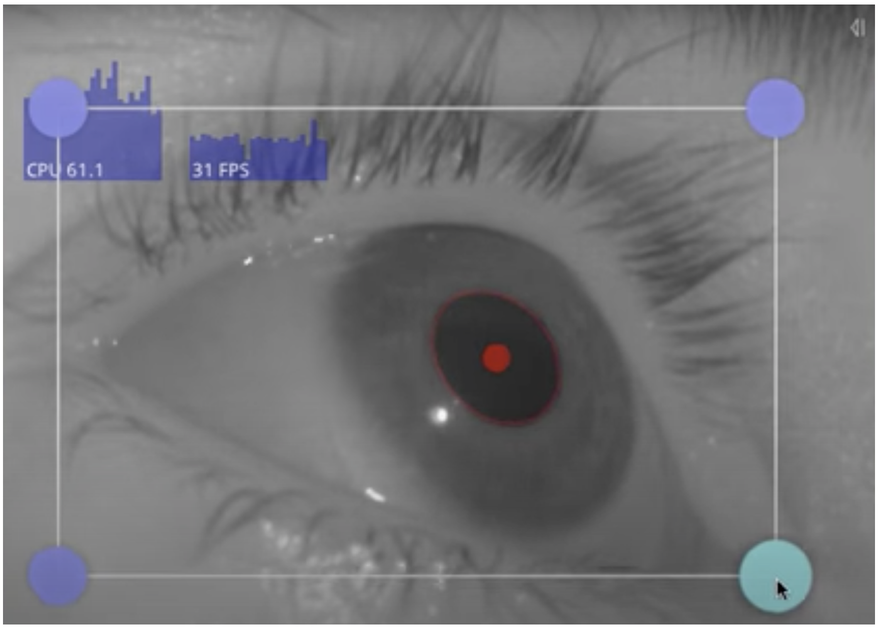


Figure K.2: An example of trimming and selecting the region of interest using the ROI parameter in the Pupil player software. The grey box with four circles could be moved to adjust the ROI in a way that trimmed any long, and dark eyelashes.

Appendix L

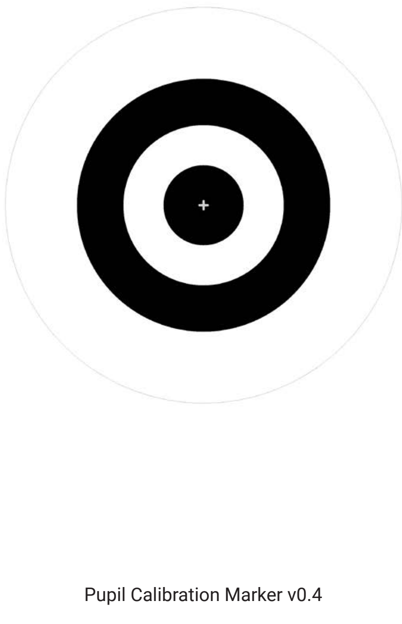


Figure L.1: Pupil Labs Calibration Marker

Appendix M

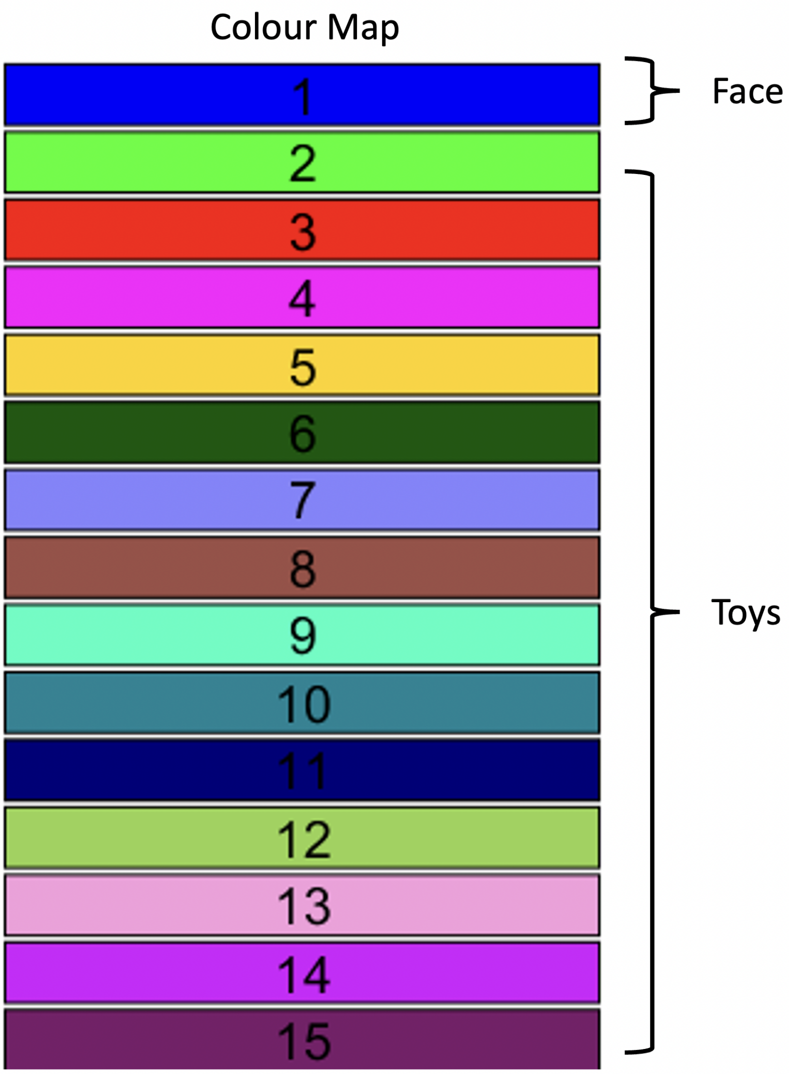


Figure M.1: Colour map for timevp visualisation

Appendix N

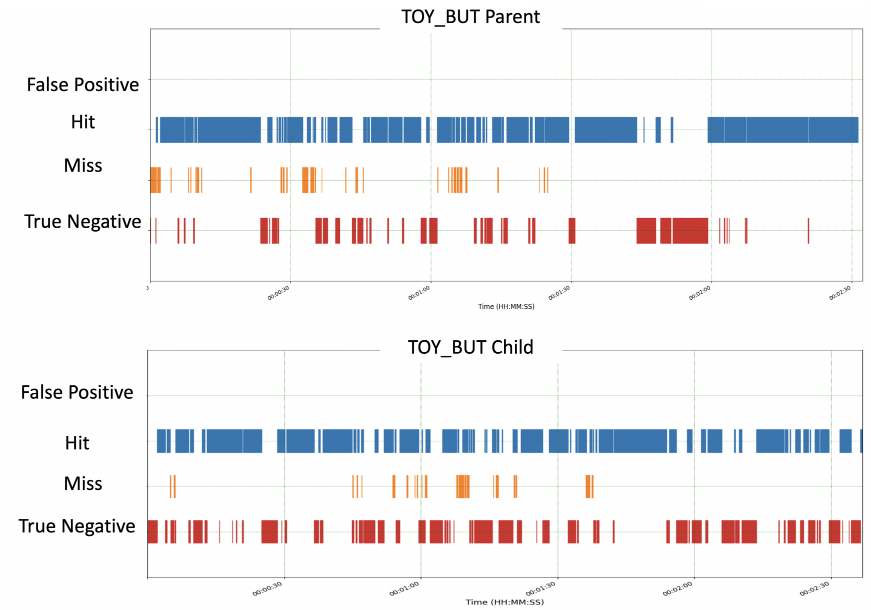


Figure N.1: YOLO results from manually coded parameters in BORIS. The top plot shows observation from a parent's headcamera and the bottom plot shows observation from an infant's headcamera for the toy butterfly (UK).

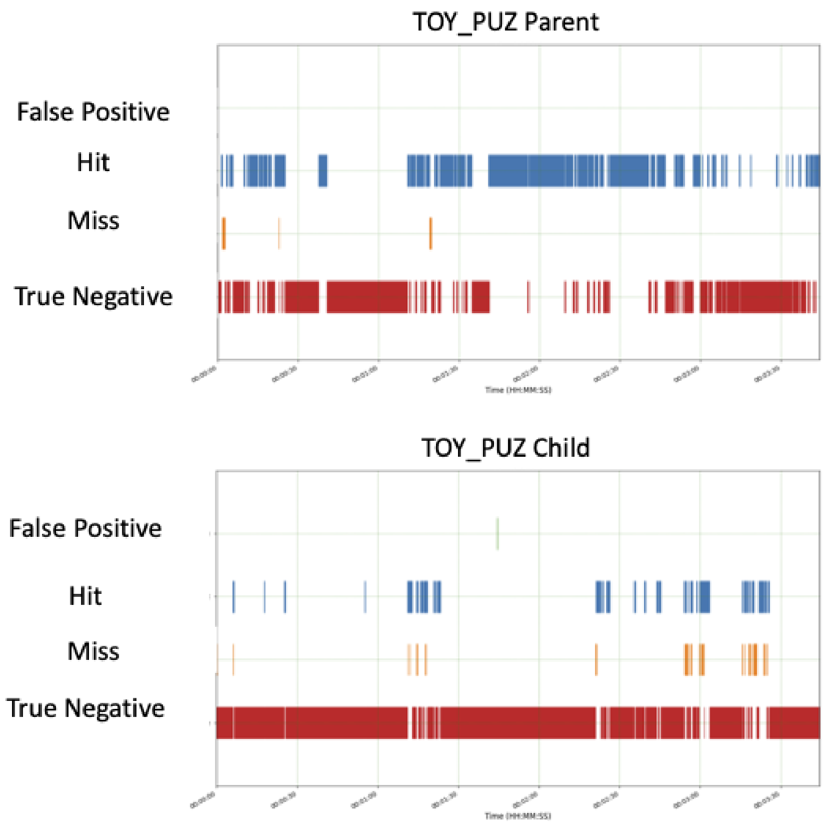


Figure N.2: YOLO results from manually coded parameters in BORIS. The top plot shows observation from a parent's headcamera and the bottom plot shows observation from an infant's headcamera for the toy puzzle (India).

Appendix O

Table O.1: Median for Infants' MLD across targets (toys, face) and conditions (alone, infant-led JA and caregiver-led JA).

Infant			
Toys			
	Alone	Infant-led	Caregiver-led
Cohort	Median	Median	Median
6UK	0.17	0.97	0.2
6IND	0.16	0.7	0.17
9IND	0.2	1.06	0.24
Faces			
	Alone	Infant-led	Caregiver-led
Cohort	Median	Median	Median
6UK	0.13	1.17	0.26
6IND	0.16	2.23	0.16
9IND	0.13	1.13	0.3

Note. MLD means Mean Look Duration. JA means Joint Attention. 6UK represents 6-month-old infants from the UK. 6IND represents 6-month-old infants from India. 9IND represents 9-month-old infants from India.

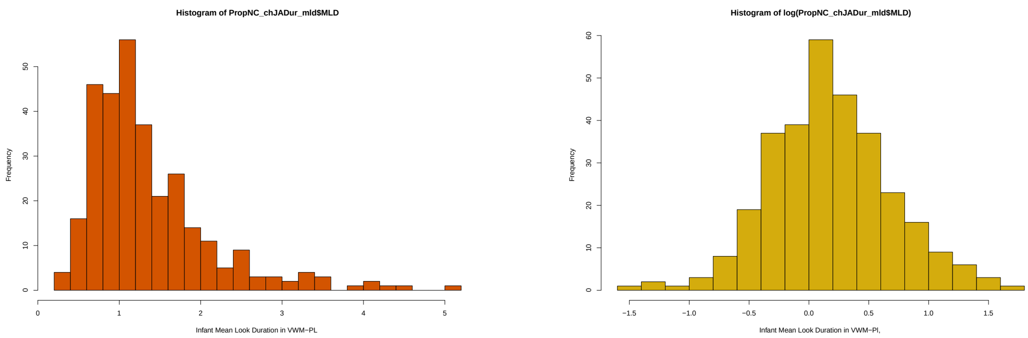
Table O.2: Median for Caregivers' MLD across targets (toys, face) and conditions (alone, infant-led JA and caregiver-led JA).

Caregiver			
Toys			
Cohort	Alone Median	Infant-led Median	Caregiver-led Median
6UK	0.13	0.13	0.3
6IND	0.17	0.2	0.73
9IND	0.17	0.23	0.63

Faces			
Cohort	Alone Median	Infant-led Median	Caregiver-led Median
6UK	0.14	0.14	0.52
6IND	0.16	0.23	0.76
9IND	0.14	0.26	0.8

Note. MLD means Mean Look Duration. JA means Joint Attention. 6UK represents 6-month-old infants from the UK. 6IND represents 6-month-old infants from India. 9IND represents 9-month-old infants from India.

Appendix P



(a) Non-transformed MLD scores from VWMP-PL task for infants

(b) Logarithmic transformed MLD scores from VWMP-PL task for infants

Figure P.1: Histogram of MLD scores from VWMP-PL task for infants.

Appendix Q

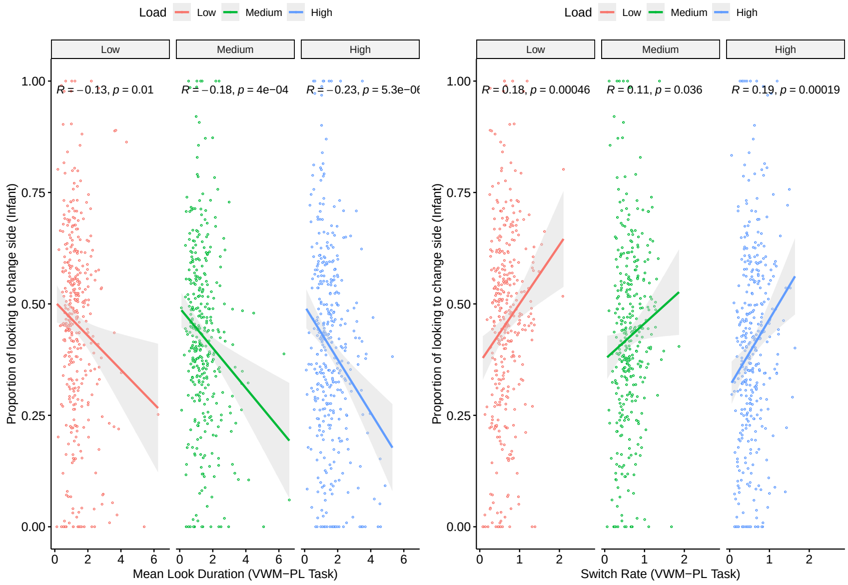


Figure Q.1: Scatterplot showing correlations between infants' VWM-PL score (i.e., the proportion of looking to change side) and (A) Mean Look duration; (B) Switch Rate obtain from VWM-PL task.

Appendix R

Table R.1: Results for infant MLD in VMW-PL task with infants MLD (alone) in dyadic interaction as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	4.84	1	<.05 *
Load	10.57	2	<.01 **
Year_s	20.59	1	<.001 ***
SESScore_c	1.96	1	.16
Age_s	0.71	1	.40
ChAloneDur_mean_c	2.50	1	.11
Load:ChAloneDur_mean_c	1.44	2	.49

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table R.2: Results for infant MLD in VMW-PL task with infants MLD in infant-led joint attention episode in dyadic interaction as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	37.29	1	<.001 ***
Load	10.35	2	<.01 **
Year_s	23.57	1	<.001 ***
SESScore_c	2.81	1	.09 .
Age_s	2.19	1	.14
ChJADur_mean_c	1.52	1	.22
Load:ChJADur_mean_c	4.32	2	.12

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table R.3: Results for infant MLD in VMW-PL task with infants MLD in caregiver-led joint attention episode in dyadic interaction as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	36.29	1	<.001 ***
Load	10.68	2	<.01 **
Year_s	21.17	1	<.001 ***
SESScore_c	2.29	1	.13
Age_s	1.62	1	.20
ChJADur_mean_c	0.61	1	.44
Load:ChJADur_mean_c	5.44	2	.07 .

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Appendix S

Table S.1: Results for infant Switch Rate in VMW-PL task with infants Switch Rate in dyadic interaction as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	8.18	1	<.01 **
Load	1.19	2	.55
Year_s	0.15	1	.70
SESScore_c	0.94	1	.33
Age_s	0.53	1	.47
individual_switches	0.06	1	.81
Load:individual_switches	0.30	2	.86

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Appendix T

Table T.1: Results for infant VWM scores with their MLD in infant-led joint-attention Episode as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	327.04	1	<.001***
Load	10.27	2	<.01 **
Year_s	0.01	1	.93
SESScore_c	0.19	1	.66
Age_s	3.28	1	.07 .
ChJADur_mean_c	0.63	1	.43
Load:ChJADur_mean_c	1.56	2	.46

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table T.2: Results for infant VWM scores with their MLD in caregiver-led joint-attention Episode as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	345.21	1	<.001 ***
Load	10.02	2	<.01 **
Year_s	0.0007	1	.98
SESScore_c	0.07	1	.79
Age_s	2.76	1	.09 .
ChJADur_mean_c	0.69	1	.41
Load:ChJADur_mean_c	2.99	2	.22

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Appendix U

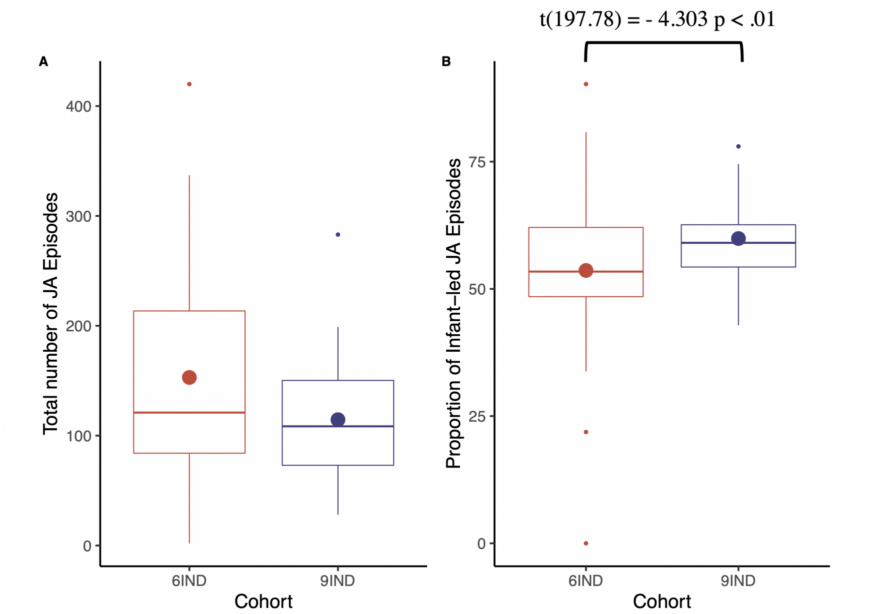
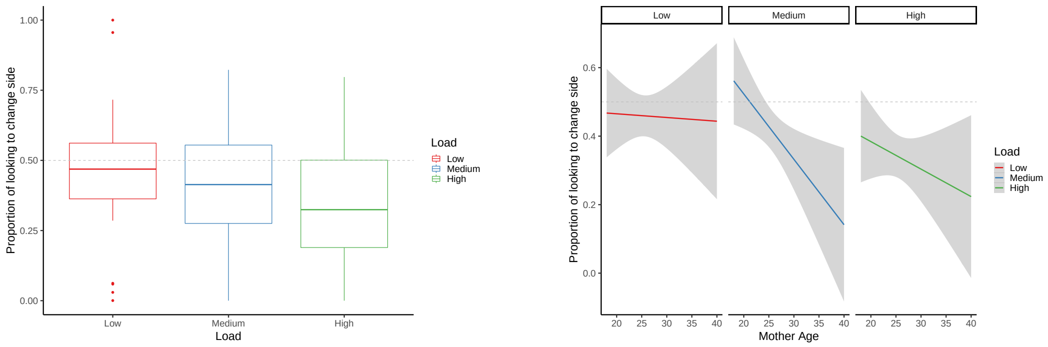


Figure U.1: Image A indicates the total number of joint attention episodes for 6- and 9-month infant dyads. Image B indicates the proportion of infant-led joint attention bouts in percentage.

Appendix V



(a) Main effect of Load in Caregivers' VWM Baseline Model (N = 46)

(b) Load and caregivers' age trend in the VWM Baseline Model (N = 46)

Figure V.1: Histogram of MLD scores from VWM-PL task for infants.

Appendix W

Table W.1: Results for caregivers' VWM performance with caregivers MLD (alone) in dyadic interaction as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

Variable	χ^2	Df	p-value
(Intercept)	237.06	1	<.001 ***
Load	10.97	2	<.01 **
SESScore_s	0.18	1	.67
MotherAge_s	0.02	1	.90
ParAloneDur_mean_c	0.002	1	.96
Load:MotherAge_s	3.70	2	.16
MotherAge_s:ParAloneDur_mean_c	0.04	1	.84

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table W.2: Results for caregivers' VWM performance with caregivers MLD in infant-led joint attention episode as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

variable	χ^2	Df	p-value
(Intercept)	238.39	1	<.001 ***
Load	10.97	2	<.01 **
SESScore_s	0.24	1	.63
MotherAge_s	0.01	1	.92
ParJADur_mean_c	0.53	1	.47
Load:MotherAge_s	3.70	2	.16

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*

Table W.3: Results for caregivers' VWM performance with their MLD in caregiver-led joint attention episode as a predictor.

Analysis of Deviance Table (Type III Wald chi-square tests)

variable	χ^2	Df	p-value
(Intercept)	244.07	1	<.001 ***
Load	10.97	2	<.01 **
SESScore_s	0.41	1	.52
MotherAge_s	0.02	1	.89
ParJADur_mean_c	3.26	1	.07 .
Load:MotherAge_s	3.70	2	.16

*Note. Blank indicates p >.05, * indicates p <.05, ** indicates p <.01, *** indicates p <.001*