Neural oscillatory insight into the endogenous cognitive processes and inter-personal contingencies that drive infant attention and support joint action during early infant-caregiver interaction

Emily Angela Mitchell Phillips

A thesis submitted in partial fulfilment of the requirement of the University of East London for the degree of Doctor of Philosophy

January 2023

ABSTRACT

Infants' ability to engage in joint attention with others towards the end of the first year is fundamental to language acquisition and shared cognition. Despite this, our understanding of the endogenous cognitive mechanisms that drive infant attention during shared caregiver-infant interaction, and support dynamic inter-personal co-ordination is, at the moment, limited.

Traditional approaches to joint attention development centred on understanding how caregivers didactically structure infant learning through ostensive communication. More recent perspectives, however, drawing on dynamic systems views of early cognition, have emphasised the role of fast-acting, multi-level, sensorimotor processes that operate across the dyad to support joint action and social learning. Newly developed micro-analysis approaches to studying early interaction have shown that infants use sensory cues to rapidly coordinate their attention with an adult partner, and statistical regularities in these cues are thought to extend infant attention and support word learning.

To fully understand the contribution of joint interactions to early cognitive development however, we need to examine the mechanisms, endogenous to the infant, that support infant engagement and interpersonal contingency on a moment-by-moment basis. One way we can examine this is through the application of neurocognitive methods, such as electroencephalogram (EEG) recordings, to studying the dynamics of naturalistic, free-flowing interactions.

Analysing time-locked and continuous associations between infants' neural activity, infant attention, and inter-dyadic behavioural coordination, this thesis assesses the sub-second cognitive processes that influence how infants allocate their attention during triadic caregiver-infant play.

First, neural evidence is presented to show that, whilst infants do not play a proactive role in creating episodes of mutual attention, they are sensitive to when their gaze is followed by an adult partner. Second, extended infant attention episodes are shown to be influenced, jointly, by attentional processes endogenous to the infant and reactive modulations in caregiver behaviour in response to changes in infant attention and cognitive engagement. Finally, the applicability of continuous methods to assessing speech-brain tracking by infants to their caregivers' speech signal during naturalistic interactions is examined, and the role of behaviour-brain entrainment in creating and maintaining episodes of joint attention considered.

Discussion focusses on the contribution of the findings to our understanding of active learning processes that operate across the dyad during early interaction, and that support the development of shared cognition. Models of early language learning in the context of the findings are considered, and directions for future work put forward.

DECLARATION

This thesis focusses on data collected in one large study, a Research Project Grant from the Leverhulme Trust RPG-2018-281 (PI: Sam Wass). The ethics approval letter and ethics application forms for the project are located in Appendix D (section 1.1 and 1.2).

Data collection was primarily performed by myself and two other PhD students working on the project, Megan Whitehorn and Ira Marriott Haresign. All three PhD students were responsible for writing the ethics application for this study, the design of the paradigms used for data collection, as well as data management and participant recruitment. I was responsible for the supervision and training of all the behavioural coding included in this thesis.

In Chapters 3, 4 and 5 the pre-processing of the EEG data was conducted by one of the other PhD students working on the project, Ira Marriott-Haresign.

Otherwise all of the data analysis, as well as the formulation and testing of the hypotheses and the writing of the text, was entirely my own work.

First-author publications arising from this thesis

Chapter 3 represents the work of the following publication:

Phillips, E.A.M, Goupil, L., Whitehorn, M., Bruce-Gardyne, E., Csolsim, F.A., Marriott-Haresign, I. & Wass, S.V. (in press). *Proactive or reactive? Neural oscillatory insight* *into the leader-follower dynamics of early infant-caregiver interaction*. Proceedings of the National Academy of Science.

Chapter 4 represents the work of an article in preparation for publication:

Phillips, E.A.M, Goupil, L., Whitehorn, M., Bruce-Gardyne, E., Csolsim, F.A., Navsheen, k., Marriott-Haresign, I. & Wass, S.V. (pre-print). *Endogenous oscillatory rhythms and interactive contingencies jointly influence infant attention during early infant-caregiver interaction.*

During my PhD I also contributed to the following publications

Marriott Haresign, I., **Phillips, E. A. M.**, Whitehorn, M., Lamagna, F., Eliano, M., Goupil, L., Jones, E. J. H., & Wass, S. V. (2022). Gaze onsets during naturalistic infant-caregiver interaction associate with 'sender' but not 'receiver' neural responses, and do not lead to changes in inter-brain synchrony. *BioRxiv*, 2022.05.27.493545. https://doi.org/10.1101/2022.05.27.493545

Marriott Haresign, I., **Phillips, E.**, Whitehorn, M., Noreika, V., Jones, E. J. H., Leong, V., & Wass, S. V. (2021). Automatic classification of ICA components from infant EEG using MARA. *Developmental Cognitive Neuroscience*, *52*, 101024. https://doi.org/10.1016/j.dcn.2021.101024

Wass, S. V., Whitehorn, M., Marriott Haresign, I., **Phillips, E**., & Leong, V. (2020). Interpersonal Neural Entrainment during Early Social Interaction. *Trends in Cognitive Sciences*, *24*(4), 329–342. https://doi.org/10.1016/j.tics.2020.01.006

Marriott Haresign, I., **Phillips, E. A. M**., Whitehorn, M., Goupil, L., Noreika, V., Leong, V., & Wass, S. V. (2022). Measuring the temporal dynamics of inter-personal neural entrainment in continuous child-adult EEG hyperscanning data. *Developmental cognitive neuroscience*, *54*, 101093.

Wass, S., **Phillips, E**., Smith, C., Fatimehin, E. O., & Goupil, L. (2022). Vocal communication is tied to interpersonal arousal coupling in caregiver-infant dyads. *Elife*, *11*, e77399.

TABLE OF CONTENTS

CHAPTER 5 - Using continuous methods of analysis to examine speech-brain tracking during naturalistic caregiver-infant object play..*145*

TABLE OF FIGURES

TABLE OF TABLES

ACKNOWLEDGEMENTS

There are a number of people I would like to thank for their role in supporting this piece of work. First, I would like to thank Sam Wass for the opportunity to be involved with this project, with whom it has been a fantastic experience to work. Sam's energy, interest and commitment to the research, and to the people involved in it, has made conducting this work, and working at the BabyLab more generally, not only academically rigorous and challenging, but fun, meaningful and exciting. I am very grateful for all the opportunities I have had to develop professionally, analytically and, most of all, technically, and for all the support and expertise Sam has provided in this.

Second to that, of course, thank you to Megan Whitehorn and Ira Marriott-Haresign who have been such brilliant colleagues and friends to work with over the past four years. We have definitely faced a good few challenges on the way, including a global pandemic, and multiple lockdowns, but our complementary approaches and skills, and commitment to the research and each other have driven all our projects' achievements. A big thank you also to Louise Goupil who has played a fundamental part in the development of the theoretical grounding of much of my work, and whose comments on my work are always so insightful and meticulous. More widely, thank you to all members of the UEL BabyLab, which is now so big and buzzy! What a great experience it has been to work with such a talented, driven and accepting group of people from whom I have learnt so much.

An important mention also to John Duncan and Lynne Murray with whom I worked before I joined the PhD. The opportunity to work with two such experienced and influential academics was brilliant, and a time I feel will always influence my work.

On a personal note, I’d like to thank all wider family and friends for their support, in Ullapool (all those with fur and feathers included), Sidmouth, Shrewsbury, Machynlleth and London, particularly, of course, Grandma. A mention also to Peter Roderick; no doubt about your influence on my career in academia. Special thanks also to my three oldest and closest friends, Alice Stevenson, Ellie Reid and Lizzie Griffin, who are always there for words of positivity, encouragement and advice. Finally, thank you to my Mum and Dad, to whom this thesis is dedicated: thank you for all your brilliant support, unerring encouragement, your insightfulness, and your values, and everything else.

This thesis is dedicated to my Mum and Dad.

CHAPTER 1 – Thesis overview

This thesis aims to examine the neural and cognitive mechanisms that drive infant attention and support inter-personal coordination during early joint interactions. In doing so, the thesis elucidates the sub-second interactive processes that govern the allocation of infant attention on a moment-by-moment basis, which might be particularly important to understanding the learning mechanisms involved in the development of intentionally mediated forms of communication, and language acquisition. The thesis incorporates 3 separate empirical analyses conducted on data collected in the same interactive context, where infant EEG activity was recorded during naturalistic, free-flowing interactions with their caregiver. Beginning with a focus on the cognitive mechanisms that drive and maintain shared episodes of caregiver and infant attention (Chapter 3), Chapters 4 and 5 go on to examine the micro-processes that structure infant attention and drive fluctuations in infants' endogenous control over their attention.

In Chapter 2 the thesis starts by outlining traditional socio-pragmatic accounts of the development of joint attention and communication that emphasise the role of intentionally-mediated behaviours in guiding infant attention in shared interactions. This is contrasted with dynamic-systems perspectives that emphasise instead the importance of dynamic sensorimotor contingencies in structuring infant behaviour and in driving shared perceptions. Differential theories on early language acquisition derived from these models is outlined. The chapter ends with a discussion on the potential for neurocognitive methods to support new insight into the intra-individual and bidirectional processes that drive the allocation of infant attention in shared interactions.

Chapter 3 examines the leader-follower dynamics that drive the development of co-ordinated attention in shared interaction, combining neural and behavioural analyses to test whether infants, at the end of the first year, show signs of proactive attention-sharing. In particular, this chapter examines fluctuations in infant neural oscillatory activity and the probability of their behaviourally ostensive cues increasing around moments of infant- vs. adult-led episodes of shared attention. Findings are discussed relative to embodied accounts of the learning mechanisms involved in the development of infants' active control over their own behaviour, and the attention of their partner.

Chapters 4 and 5 build on these findings in examining the continuous and event-locked intra- and inter-personal dynamics that structure infant attention and drive fluctuations in infants' endogenous cognitive processing. Specifically, in Chapter 3, forwards-predicative and reactive changes in caregiver behaviour. and infants' endogenous neural activity, are assessed relative the duration of infant attention episodes towards objects and their partner. In Chapter 5, the applicability of continuous methods of analysis to examining speech-brain tracking in early naturalistic interactions is assessed, and the implications of the development of these methods of analysis for assessing behaviour-brain tracking in naturalistic interactions is considered relative to dynamic systems perspectives on early cognitive development.

In the General Discussion of Chapter 6, the thesis considers the implications of the findings of Chapters 3, 4 and 5 for our understanding of the development of infants' active control over their own behaviour and attention in social interactions. The thesis ends with a consideration of the development of further analytical procedures and methods for examining the mechanisms that drive shared attention, perceptions and cognitions in shared interactions, and how we might use these methods to further our understanding of the inter-dyadic processes that support early language acquisition.

CHAPTER 2 - General introduction

2.1 Introduction

The influence of early joint attentional interactions with a caregiver on infants' language, attention and socio-cognitive development is well established. We know, for example, that consistent and contingent responding by a caregiver to their infants' communicative behaviours associates with superior language outcomes (Tamis-LeMonda et al., 2001), that more reciprocal interactions predict developments in aspects of cognition and attention (Murray, De Pascalis, Tomlinson, et al., 2016), and that, infants who more frequently engage in attention-sharing behaviours with their caregiver develop a better understanding of the pragmatic and referential aspects of shared communication (Goldin-Meadow et al., 2007). Despite this, we currently know little about the neural and cognitive mechanisms that drive infant attention and support inter-personal coordination on a moment-by-moment basis. Examining this is crucial to our understanding of how infants begin to integrate and learn from the multi-level, dynamic and hierarchically nested processes that organise joint action and behaviour.

In the sections that follow, different accounts of the cognitive mechanisms that influence how infants allocate their attention in shared interactions are outlined. First, socio-pragmatic accounts of early joint attention development and their implications for how we understand infant attention and learning in early interaction are considered. Second, dynamic systems perspectives of the influence of social interactions on early cognition are contrasted to the traditional perspective; considering, in particular, the multi-level, micro-behavioural processes associated with the development and maintenance of co-ordinated attention. Third, the application of neurocognitive methods to studying early social interactions and informing our understanding of the mechanisms that drive micro-second modulations in infant attention and behaviour when interacting with a social partner are discussed.

2.2 Joint attention and the development of shared intentionality

Traditional approaches to understanding early joint attention development have largely focussed on identifying at what stage in development infants begin to engage in intentionally mediated forms of shared communication (Csibra & Gergely, 2009; Tomasello, 2016; Tomasello et al., 2007; Tomasello & Carpenter, 2007). Infants are sensitive to ostensive signals such as direct gaze (Farroni et al., 2002) and infant-directed speech (IDS) from the first few days of life (Cooper & Aslin, 1994), and, over the first six months develop in their ability to respond to the social cues of a communicative partner (Senju et al., 2008). Towards the end of the first year, at around 9-12 months, infants are thought to begin to engage in episodes of attention sharing, where they understand others as intentional agents, whose communicative behaviours are the result of goal-directed thought, and that their own intentionally motivated behaviours can influence the attention, actions and mental states of their social partner (Siposova & Carpenter, 2019; Tomasello et al., 2007). This is often referred to as the ‘9-month revolution’, and is thought to result from the ontogenetic motivation of human infants to engage in collaborative forms of joint action (Tomasello et al., 2005).

In particular, the socio-pragmatic model emphasises the role of intentionally-mediated communication in driving early language development: asserting that only once infants can understand and create shared internal representations with others can they begin to learn from the referential and pragmatic aspects of joint attention (Lieven, 2016; Tomasello, 2000). This perspective has driven prolific research aimed at examining how infants begin to respond to a social partners' initiations for joint attention, at what stage infants begin to initiate episodes of joint attention during shared interactions, as well as the longitudinal associations between early joint attention and later language abilities (Behne et al., 2005; Brooks & Meltzoff, 2015; Carpenter et al., 1998; Donnellan et al., 2020; Liszkowski, Carpenter, et al., 2008; Liszkowski & Tomasello, 2011).

In the sections that follow, work tracking the development of infants' ability to respond to and initiate ostensively guided episodes of shared communication with a social partner is outlined and discussed relative theories of the mechanisms that drive infant attention in early interactions.

2.3 Development of the ability to respond to social signals in the first year

During joint interactions, participating individuals are thought to act alternately as senders and receivers of social information (Csibra & Gergely, 2009). Ostensive cues are used by both partners to signal communicative intent: sending information at moments they wish to convey particularly important information to the recipient (Csibra, 2010; Csibra & Gergely, 2009). A vast literature documents the development of infants' sensitivity to ostensive signals over the course of the first year (Brooks & Meltzoff, 2015; Farroni et al., 2002, 2003; Senju et al., 2008), and a particular focus of this work has been on how and when infant begin to understand the referential nature of these cues.

Infants are sensitive to eye gaze soon after birth, preferentially orienting towards faces with direct gaze from 3-5 days (Farroni et al., 2002), and by 4 months, show increased attention and greater neural responses to images of adults engaged in direct gaze (Farroni et al., 2002; Grossmann et al., 2013). For example, recording electroencephalography (EEG) from 4-month-olds whilst they viewed pictures of adults engaging in direct vs. indirect gaze, Farroni et al. (2002) showed that averaged N170 event-related potential (ERP) amplitudes, over occipital electrodes, were greater where infants viewed images of adults looking directly towards them. Similarly, infants prefer infant-directed speech (IDS), compared to adult-directed speech (ADS) from the first few days of life (Cooper & Aslin, 1994), and by 5 months show an enhanced Nc ERP component to words produced with IDS intonation (Parise & Csibra, 2013). Compared to ADS, IDS is less syntactically complex, more repetitive, involves more prosodic variability, has a slower tempo, and has recently been associated with a shift in timbre (Papoušek et al., 1990; Piazza et al., 2017; Soderstrom, 2007).

The mechanisms driving infants' responsivity to ostensively produced signals produced by a social partner develop over the course of the first year. Newborns and 3-month-old infants, for example, have been shown to orient towards targets cued by gaze shifts (Hood et al., 1998). It is thought that this early gaze-following reflects rudimentary mechanisms, with infants shifting their attention in the direction of lateral motion, rather than using gaze direction as a referential cue (Farroni et al., 2000, 2003). In the latter half of the first year, however, infants are thought to begin to understand the referential nature of eye gaze. For example, at 6 and 9 months, infants have been found to only follow an adults' gaze when their attention shift is preceded by an ostensive signal (Senju et al., 2006, 2008; Senju & Csibra, 2008). Senju et al., (2008) presented 6-month-old infants with a video of an experimenter shifting their gaze towards one of two objects located on either side of them. In one condition, the experimenter looked directly towards the infant before shifting their gaze to the target object, whilst, in the other, a cartoon covered the experimenters' head during the period of direct eye contact. Only infants in the direct eye contact condition were more likely to look towards the target object when the experimenter shifted their attention. Where infants in the no eye contact condition heard IDS before the attention-shift, however, they looked towards the target object reliably more often, compared to infants who heard ADS. This finding has been interpreted as evidence that, by 6 months, infants interpret a social partner's directed gaze to indicate their referential intent.

In the socio-pragmatic perspective, infants' ability to respond to and understand the meaning of ostensive signals is fundamental to early language acquisition, whereby infants' relational understanding between an interacting partners' cues and the objects to which their attention is directed supports early word-object mappings (Carpenter et al., 1998; Donnellan et al., 2020; Tomasello & Carpenter, 2007). Corresponding to this prediction, for example, Brooks and Meltzoff (2008) demonstrated that 10-12 month old infants' ability to accurately and contingently follow the gaze of an adult partner towards one of two objects on a table predicted infants' receptive and expressive vocabulary at two years of age.

2.4 Infants deliberately direct the attention of a social partner towards the end of the first year

Towards the end of the first year, as well responding to adults' signals for shared attention, infants are thought to begin to deliberately initiate episodes of shared attention, through intentionally-produced communicative cues such as gestures and vocalisations (Carpenter et al., 1998; Donnellan et al., 2020; Liszkowski & Tomasello, 2011; Tomasello et al., 2007). This later emerging ability is, in-part, thought to reflect the development of the executive attention system, which develops rapidly over the course of the first year, corresponding to the maturation of the pre-frontal cortex (Colombo, 2001; Gogtay et al., 2004; Mundy & Newell, 2007).

It is generally thought that infant pointing at the end of the first year serves a declarative function (Liszkowski, Carpenter, et al., 2008; Liszkowski & Tomasello, 2011; Tomasello et al., 2007): i.e. that infants proactively create episodes of joint attention with a social partner so that attention towards a stimulus is not only shared, but common ground between the two partners (Siposova & Carpenter, 2019). A series of studies conducted by Liszkowski and colleagues supports this perspective. In their paradigm, infants were seated next to an experimenter and on each trial an object appeared from behind a curtain. In one condition, the adult experimenter responded to infant points towards the object with joint attentional behaviours (looking back and forth between the infant and the object) whilst, in the other, the adult spoke excitedly about the object but did not alternate their gaze between the object and the infant (Liszkowski et al., 2004; Liszkowski et al., 2008). Recording the pointing behaviour of infants in each condition, they found that in the non-JA trials, infants pointed more frequently towards the object (Liszkowski et al., 2004), and increased the frequency of their vocalisations (Liszkowski et al., 2008). The authors interpreted findings to suggest that infants point to modulate the behaviours of their partner, and persist in gesturing where their gestures are not met with the predicted contingent response.

Developments in infants' knowledge that their own behavioural cues signal intention to share attention with an interactive partner are thought to be particularly important in driving the development of a communication system (Donnellan et al., 2020; Goldin-Meadow et al., 2007). Socio-pragmatic perspectives have emphasised that where infants initiate shared attention with a social partner, infants are able to map caregivers' speech inputs to the object of their joint focus, based on the understanding that they have influenced the focus of their caregivers' attention, and the caregiver is aware of their focus of attention, and the infants' intention to share attention (Tomasello et al., 2007). Supporting this perspective, the frequency with which infants engage in pointing gestures towards objects during interactions with their caregivers at the end of the first year positively associates with later language outcomes (Carpenter et al., 1998; Fenson et al., 1994; Goldin-Meadow et al., 2007), particularly where infants engage in joint attentional referencing (Carpenter et al., 1998; Donnellan et al., 2020).

Overall, the socio-pragmatic account has emphasised that, towards the end of the first year, infant attention during social interactions begins to be driven by their engagement in intentionally mediated forms of communication. Largely, this work emphasises the role of the caregiver in structuring infant attention and learning in joint attentional frames: infants follow, or later, lead an adult partners' attention towards a shared focus of attention, and the caregiver didactically scaffolds infant learning on the basis of shared internal representations (Csibra & Gergely, 2009; Tomasello & Carpenter, 2007). More recent work, however, has highlighted the active role infants might play in selecting the information they learn about and in regulating the behavioural inputs of their partner. This works is outlined in the next section.

2.5 Active learning from a socio-pragmatic perspective

Active information seeking strategies have been shown to influence where infants look and for how long they engage with stimuli in their environment (Goupil & Proust, 2023; Kidd et al., 2012; Poli et al., 2020). For example, as early as 7-8 months, infants allocate their attention towards stimuli that are neither too complex nor too predictable (Kidd et al., 2012; Kidd et al., 2014). Kidd et al. (2012) presented infants with a sequence of images designed so that each image, as it occurred in sequence, varied in terms of its predictability (i.e. its likelihood of occurring, based on the sequence of images leading up to that image being presented). The authors found that infants' likelihood of looking away from the image showed a U-shaped relationship with image complexity, such that infants were less likely to look away from an image with mid-range levels of predictability. Infants therefore appeared to direct their attention towards stimuli with most opportunity for information gain (i.e. stimuli that were not too complex and therefore cognitively demanding to process, nor stimuli that were overly simple; Kidd et al., 2012; Kidd & Hayden, 2015). Presenting 8-month-old infants with different shapes, appearing as targets and cues within each trial frame, Poli et al. (2020) have also recently shown that, as well as allocating their attention towards stimuli with mid-range levels of predictability, infants look longer to sequencies of stimuli with greater opportunity for learning progress. Rather than passive recipients of information, these findings suggest that intrinsic, motivational factors, such as interest and curiosity can guide how infants allocate their attention in the environment (Kidd & Hayden, 2015). The influence of these endogenous factors is likely to vary as a function of individual differences in infants' information processing abilities (Piccardi et al., 2020), as well as moment-by-moment fluctuations in their cognitive engagement with external changes in the environment (Piccardi & Gliga, 2022).

These findings have informed the view that, in shared interactions at the end of the first year, infants' ostensive communication with an adult partner might be guided by epistemic feelings of curiosity: actively directing their partners' attention towards specific stimuli in order to receive new information about their environment (Begus et al., 2014; Begus & Southgate, 2018). Building on previous theories of the intentionally-motivated behaviours of infants in social interactions (Tomasello et al., 2007), these accounts argue that infants are aware of how their ostensive behaviours influence the behaviours of a social partner, and use these cues to regulate the amount and the timing of the information they receive (Begus & Southgate, 2018).

Beyond declarative signalling, a series of findings have revealed that, by the end of the first year, infants use ostensive signals interrogatively, with the aim of directly eliciting information from a social partner about their environment (Begus et al., 2014; Goupil et al., 2016; Kovács et al., 2014). Engaging in shared attention episodes towards objects with an experimenter, 12-month-old infants have been found to increase the frequency of their pointing across experimental trails where the experimenter consistently provides new information about an object. In comparison, infants decrease their rate of pointing where experimenters provide information the infant already knows, or no information at all (Kovács et al., 2014). This replicated previous findings among infants aged 16 months (Begus et al., 2014), and is also in line with reports that, in naturalistic interactions with their caregiver, 11-12-month-old infants increase the rate of their pointing gestures and vocalisations, where caregivers are consistently and contingently responsive to these cues (Gros-Louis et al., 2014).

Responding to infants' ostensive signals has also been shown to have immediate effects on infant learning, possibly indicating a readiness by the infant to receive new information (Begus & Southgate, 2018). For example, Begus et al. (2014) have shown that, among infants aged 16 months, infant learning about the function of objects is better where this information is presented contingent to their interests. Engaging in a structured paradigm with an experimenter, infants were presented with 4 objects. Once infants pointed towards one of the four objects, the experimenter either demonstrated the function of that object, or an object to which the infant had not gestured towards. Recording infant actions on the demonstrated objects in the time after the experimental phase, infants were found to more accurately replicate the actions of the objects that they had pointed towards. As early as 12 months, infants have also been found to learn object labels better where this information is presented contingent on their object-directed vocalisations (Goldstein et al., 2010). These findings of the direct association between infant learning and the contingent placement of object labels relative to their cues is in line with many previous findings documenting associations between caregiver responsivity to infant behaviour, and later language development (Murray et al., 1993; Tamis-LeMonda et al., 2001).

Similar to the traditional socio-pragmatic perspective, then, active learning accounts emphasise that, by the end of the first year, infants are proactive in creating episodes of shared attention with their adult partner, and, beyond this, learn better where their interests are responded to. However, much of this work on joint attention development and active ostensive signalling has been conducted using structured, experimental paradigms, where a researcher engages in clear, repetitive behaviours aimed at eliciting either a response to their initiations for shared attention in the infant, or an initiation for shared attention *by* the infant. On each experimental trial, therefore, the adult's behaviour is spatially precise and temporally stable (Yu & Smith, 2013): far from the fast-changing multi-layered complexity of naturalistic, free-flowing interactions (Knoblich & Sebanz, 2008; Wass et al., 2020). To understand the processes that support infants' engagement in coordinated episodes of attention, as well as the mechanisms that support language learning, recent dynamic system perspectives have emphasised that we need to study social interactions in naturalistic contexts and at the micro-behaviural level. The dynamic systems approach and its implications to our understanding of the mechanisms that drive infant attention in shared interactions is outlined in the next section.

2.6 Embodied approaches to understanding the influence of early interactions on infant cognition

Dynamic systems approaches to understanding social interaction emphasise the role of sensorimotor process in driving attention and social learning. Rather than the consolidation of abstract representations, based on experience and cultural knowledge (Tomasello & Carpenter, 2007), embodied perspectives emphasise that cognition emerges through the interaction of an individual with their immediate environment, where mental constructs are generated in-the-moment through the integration of disparate and nonstationary sensorimotor activity (L. B. Smith, 2005; L. Smith & Gasser, 2005, p. 200). Applying this perspective to social interaction, dynamic system theorists propose that the fast-acting sensory-motor coordination of both interacting partners drives and maintains shared perceptions and cognitions, such that social interactions are hierarchically nested and characterised by a dynamic, inter-dyadic flow of information from cognition to action to sensory experience (Knoblich & Sebanz, 2008; Yu & Smith, 2012, 2017).

In very early face-to-face interactions, consistent and timely sensorimotor inputs by the caregiver as a function of infant behaviours are thought catalytic to the development of higher-order cognitive processing (L. B. Smith & Breazeal, 2007). In particular, repeated and consistent feedback to the production of specific infant behaviours assigns meaning to infant behaviour, driving the development of their understanding about the intentions of others, and how their own intentionally-motivated behaviours affect those of their partner (Oudeyer & Smith, 2016; L. B. Smith & Breazeal, 2007; L. Smith & Gasser, 2005). Indeed, even in the first few months of life, modulations in infant behaviours, and communicative cues elicit temporally contingent and functionally relevant change in caregiver behaviours (Cohn & Tronick, 1987; Murray, De Pascalis, Bozicevic, et al., 2016; Murray & Trevarthen, 1986, p. 198). In very early face-to-face interactions, caregiver facial expressions frequently mirror and dynamically adapt to micro-second fluctuations in infant affect and vocalisations (Feldman, 2007; Murray, De Pascalis, Bozicevic, et al., 2016; Tronick et al., 1977), responding differentially to varied affective behaviours. Recording caregiver-infant face-to-face interactions at 2 months, for example, Murray and colleagues showed that social expressions produced by the infant were more likely to be mirrored by the caregiver or marked with positive affect, compared to negative expressions, and biological movements (Murray, De Pascalis, Bozicevic, et al., 2016).

Similarly, throughout the first year, caregivers respond to infant vocalisations with temporally consistent contingency (Bornstein & Azuma, 1992; Gratier et al., 2015; Jaffe et al., 2001; Van Egeren et al., 2001), and modulate their vocal feedback, depending on the type of sound produced by the infant (Albert et al., 2018; Goldstein et al., 2003; Yoo et al., 2018). For example, analysing day-long home recordings of caregivers and 3-month-old infants Yoo et al. (2018) showed that a pause-response sequence largely characterised the behaviour of caregivers in the time after speech-like vocalisations produced by the infant. In comparison, infant cry vocalisations, defined as distresses sounds produced with negative affect, elicited vocal responses from the caregiver that over-lapped with the end of the vocalisation. Interacting in lab-based settings with infants aged 9 months, caregivers have also been shown to respond more often and with grater semantic complexity to more mature protophones produced by infants (Albert et al., 2018).

Not just passively receptive to these responses, infants are sensitive to the contingencies of an adult partner from early in the first year. For example, at 2-3 months, where caregivers pause interacting with their infant after a period of free play, infants increase their frequency of smiles, as well as the number and duration of their vocalisations (Bigelow & Power, 2016; Bourvis et al., 2018), and increase the length of their vocalisations once caregivers begin interacting again (Bourvis et al., 2018). Repeated experience of contingency at 3 months has also been shown to already influence how infants interact in novel situations (Bigelow & Rochat, 2006). Comparing 2-3-month-old infants' interactions with their caregivers to where they interacted with a stranger, Bigelow et al. (2006) showed that infants were less responsive to the smiles and vocalisations of the stranger when the frequency of their behavioural responsivity was either greater or less frequent compared to the responsivity of their caregiver.

2.7 Sensorimotor contingencies and joint attention at the end of the first year

Recently, there has been a move towards understanding how fast-acting sensorimotor contingencies operate across the dyad during inter-dyadic interactions in later infancy, to support triadic and coordinated visual attention. Rather than focussing on the developmental onset of the production of ostensive communication and referential understanding

(Tomasello, et al., 2007), dynamic systems approaches have begun to examine the moment-by-moment fluctuations in each individuals' behaviour, that create and maintain episodes of joint attention towards a common referent (e.g. an object; Yu & Smith, 2012). This work is opening new avenues for understanding the inter-dyadic sensorimotor dynamics that support the development of intentionally-mediated communication, with important implications for our understanding of early language acquisition (Yu et al., 2021).

In a series of recent studies conducted by Yu and Smith, micro-behavioural analysis of caregiver-infant behaviours during triadic play with objects at the end of the first year has suggested that, rather than controlled, intentionally-mediated ostensive signalling (Begus et al., 2014; Tomasello et al., 2007), fast-acting, sensorimotor contingencies largely drive the allocation and coordination of infant attention. For example, analysing the intra- and inter-dyadic associations between caregiver and infant touch to objects and their visual attention, Yu and Smith (2013) showed that episodes of joint attention were largely achieved and perceived by infants through attending towards their own hand actions, and that of their adult partners'. Caregivers and infants engaged in mutual gaze for less than 10% of the interaction, and, whereas caregivers increased looking towards the infants' face 2000ms before following the infants' attention, infants only increased looking 200ms before jointly attending to the same object as their partner. Instead, infant attention was strongly associated with the hand actions of both partners, and coupling between their attention, and the adult's hand actions markedly increased where infants followed their partners' attention towards an abject.

In a later longitudinal analysis of infants aged 12 and 18 months, Yu and Smith (2017) also showed that, at both ages, caregivers and infants used multiple coordination pathways to achieve episodes of joint attention, varying in the sequence of each individuals' hand and eye movements towards an object in the time before and just after the onset of joint attention. Again, at both 12 and 18 months, infants rarely followed the gaze of their partner towards the object, though infant-led pathways to joint attention became more common at 18 months. The finding that 12-18-month-old infants rarely use the gaze of their partner to co-ordinate their attention during triadic interactions has now been replicated across different labs, employing eye-tracking techniques during table-top play (Custode & Tamis-LeMonda, 2020; Franchak et al., 2011), as well as floor-based, free-moving interactions with their caregivers (Franchak et al., 2018).

Challenging the view that, towards the end of the first year, infant attention is driven by intentionally mediated forms of communication (Tomasello et al., 2007), then, infants do not appear to routinely engage in active attention-sharing behaviours in naturalistic, free-flowing interactions. Instead, similar to observations of early face-to-face interactions (Bigelow & Power, 2016; Murray, De Pascalis, Bozicevic, et al., 2016), their attention is largely *reactive* to the sensorimotor coordination dynamics of the interaction: following and leading their partners' attention towards objects via fast-acting, salient and visually precise object manipulations (Yu & Smith, 2017).

Despite that infants do not look to their caregiver to identify where they are attending, jointly looking towards an object at the same time as the adult has been found to lead to immediate increases in infants' attention towards an object. In a sample of infants aged 11 – 13 months, Yu and Smith (2016) tested differences in the duration of episodes of sustained attention (defined as unbroken infant object looks lasting 3s or longer) by the infant where their caregiver jointly attended towards the object, compared to where they looked towards the object on their own. They found that sustained attention episodes accompanied by their adult partners' gaze lasted significantly longer; a difference that could not be explained by the time it took caregivers to join their infants' object look. This finding has since been replicated among infants aged 12 months (Abney et al., 2020; Suarez-Rivera et al., 2019), and, comparing joint table-top play to solo infant play, Wass et al. (2018a) have shown that infant attention durations during joint play are, overall, longer, compared to where infants play with the same objects on their own (see also McQuillan et al., 2020 for similar findings).

Taken together, these micro-behavioural findings suggest that, towards the end of the first year, rather than the organisation of shared attentional episodes through ostensive communication and the establishment of shared internal representations, infant attention in social interaction is largely responsive to and possibly scaffolded by low-level sensory inputs that operate across the infant-caregiver dyad (Yu & Smith, 2013, 2017). In the next section, the implications of these findings to understanding early language acquisition, in the context of the possible mechanisms driving infant attention is considered.

2.8 Implications of embodied cognitive approaches to understanding the mechanisms that drive infant attention and support early language acquisition

That infant attention is most often focussed on objects during shared interactions is difficult to reconcile with the socio-pragmatic perspective on early language development that emphasises the role of shared intentions in creating clean referential mappings between the caregivers' speech inputs and the objects to which they are attending (Lieven, 2016; Tomasello et al., 2007). Recent work examining micro-second relationships between infant gaze and caregiver speech, during online, free-flowing interactions has highlighted, instead, the quality of infant attention at the time of object naming, as well as the timing of object labels produced by the caregiver, relative to infant attention (Yu et al., 2019; Yu & Smith, 2012). For example, recording infant gaze behaviour with head-mounted eye trackers, Yu and Smith (2012) have shown that infants learnt object labels better where the caregiver provided an object label at a time that one object was most dominant in the infants' field of view, and that caregivers most often produced labels at these moments. Associating the duration and timing of caregiver and infant looking behaviour with later language outcomes, the duration of infant attention towards objects at 12 months has been found to predict language development at 15 months, whilst the extent to which caregivers' monitor their infants' attention by alternating their gaze between the object and the infant at 9 months, predicts infant vocabulary aged 12 months (Abney et al., 2020; Yu et al., 2019).

Overall, these studies suggest that infant learning is associated with child-centred factors (i.e. the quality of their attention to objects) and adult centred factors (the attunedness of the caregiver to the infants' behaviour and their timely placement of speech inputs). But, what we have little understanding of, and, at the moment, little theoretical discussion of (although see Yu et al., 2021; Yu & Smith, 2012), is the intra- and inter-individual hierarchical processes that drive the child- and adult-centred factors that associate with moment-by-moment learning. For example, the visual dominance of an object in the child's field of view could be the result of the sensorimotor behaviours of the infant, higher level cognitive factors such as interest, or adult-led factors, such as modulations in the saliency of their behaviour (Yu & Smith, 2012). Nor do we understand the inter-dyadic factors that increase the duration of infant attention in shared interactions, and how this might support infants' early control over their own attention and behaviour. In the section below, different models to understanding the endogenous and interactive mechanisms that drive and maintain infant attention in shared interactions in the context of the embodiment perspective are discussed.

2.9 Models of the endogenous and interactive contingencies driving and maintaining infant attention in shared interactions

One possibility is that the dynamics of the interaction, at different levels of hierarchical integration, are predominantly adult-led: i.e. the behaviours of the adult largely regulate the allocation of infant attention, and, possibly fluctuations in infants' endogenous attention control, as well as the timing of the caregiver's speech and behavioural inputs. As mentioned previously, for example, caregiver's hand actions on objects could create visual saliencies in the environment to which the infant is responsive (Yu & Smith, 2017).

As well as overt attention capture, other covert, fast-changing processes in the caregivers' speech signal could also influence infant attention on a moment-by-moment basis. For example, prosodic features of IDS are particularly salient, indexed by the fundamental frequency (F0) of the voice, which is acoustically perceived as pitch (Spinelli et al., 2017). The mean F0 in IDS is higher compared to ADS speech, and the width of F0 variation within and between utterances is greater (Soderstrom, 2007). In controlled looking-time paradigms variation in the fundamental frequency of speech has been shown to influence infant attention more than other features of IDS, as well as affecting infants' own vocal production (Cooper & Aslin, 1994). In a recent study, Nencheva et al. (2019) have also shown that infant arousal synchrony to a speech stimulus (indexed by inter-participant pupil synchrony) was higher when they listened to rising and bell shaped F0 contours, compared to where they heard flat, or falling contours, and greater arousal synchrony predicted better learning.

Whilst little work has examined associations between infant attention and fluctuations in caregiver F0 on a moment-by-moment basis, early work conducted by Stern showed that, during interaction with infants aged 2-6 months, mothers used rising contours to re-engage infant attention toward their face, and bell-shaped contours to maintain infant positive affect (Stern et al., 1983). It is possible that, during triadic interactions, caregivers modulate the saliency of their voice to capture and maintain infant attention, and to highlight and regulate moments of information transfer (Wass et al., 2018a). Similarly, beyond the environmental saliencies created through their actions on objects (Yu & Smith, 2013, 2017), the temporal patterning of caregiver body and hand movement could also operate as an exogenous mechanism of attention capture (Meyer et al., 2022). For example, in an experimental paradigm, Meyer et al. (2022) recently showed that variability in the hand movements of an experimenter associated with increased cognitive engagement among infants aged 15 months (as indexed by infant electroencephalography (EEG) activity, occurring at theta frequencies (4-5Hz)); and that variability, rather than the amplitude of the adults' hand and arm movements, predicted greater infant engagement.

Within this framework, it is possible that, rather than repeated and reactive contingent responsivity to isolated behaviours, temporal dependency between infant and caregiver attention is driven by infant behaviours becoming periodically coupled to the behavioural modulations of their partner (Wass et al., 2022). Similar to inter-dyadic patterns of vocalisations in adults and marmoset monkeys (Takahashi et al., 2013; Wilson & Wilson, 2005), in early infant-caregiver interactions, vocal pauses in one partner's vocalisations can be predicted from those of the other (Gratier et al., 2015; Jaffe et al., 2001), and, during face-to-face interactions at the end of the first year, caregiver-infant facial affects become consistently temporally aligned (Feldman et al., 1999). Oscillatory entrainment by the infant to their caregivers' behaviour could, in particular, support the organisation and stability of their own attention and sensorimotor behaviours, as well as driving the development of expectancies and predictions related to the inter-dyadic contingencies of the interaction (K. J. Friston, 2019; S. Wass, 2022).

Alternatively, rather than leading the infants' attention, it could be that the adults' role is in monitoring and contingently responding to re-orientations in their infant's gaze. In following the focus of their infant's attention at moments that they reorient towards a new object, the caregiver 'catches' and extends infant attention with reactive and dynamic change in their salient ostensive behaviours, which infants are responsive to (Yu & Smith, 2012, 2016). Indeed, we know that at the macro-level, caregiver responsivity to their infant's behavioural cues and vocalisations associates with better infant learning and language outcomes (Bornstein & Azuma, 2020; Tamis-LeMonda et al., 2001). In more recent work, the behavioural contingency of the caregiver to micro-dynamic shifts in infant attention has been found to increase the duration and quality of infant attention to objects in shared interaction (Miller & Gros-Louis, 2013; Yu & Smith, 2012). Examining the relationship between the micro-second leader-follower dynamics and infant learning, Yu and Smith (2012) also showed that moment-to-moment object name learning was better where caregivers' presented object labels contingent on infant-led looks towards objects.

In this alternative infant-led perspective, we need to consider the mechanisms that might drive the timing of shifts in infant attention, to which adults respond. One possibility is that endogenous, motivational or epistemic processes, such as interest, cause re-orientations in infant attention (Kidd et al., 2012). Though infants at the end of the first year may not routinely engage in active attention-sharing behaviours (Begus et al., 2014), it is nevertheless possible that they attend towards the things that interest them, and learn better from their caregivers' inputs where their attention is intrinsically guided (Kidd et al., 2012; 2014). An alternative possibility, however, is that, rather than motivational factors, *endogenous regulatory* mechanisms drive temporally consistent shifts in infant attention (Nuthmann & Henderson, 2010; Stallworthy et al., 2020), and the contingent responsivity of the adult partner serves, in places, to extend the durations of infant attention through allostatic inter-dyadic process (McQuillan et al., 2020; Wass et al., 2018a; Yu & Smith, 2016). In particular, the fast-changing behaviours of the caregiver, relative to the slow-moving behaviours of their infants, could serve as an up-regulatory mechanism: catching infant attention and increasing their engagement through amplificatory mechanisms that operate across the dyad (Abney et al., 2020; Gratier et al., 2015; Wass, 2021a). Supporting this perspective to some extent, Suarez-Rivera et al. (2019) showed that longer sustained attention episodes by infants during early interactions were associated with more frequent multi-modal behavioural inputs by the caregiver (Suarez-Rivera et al., 2019). Caregiver speech inputs to infants where they are successful in catching and scaffolding infant attention could be particularly important to driving early language development.

Though micro-behavioural analyses of caregiver-infant interaction have attempted to delineate between exogenous and endogenous influences on infant attention, as well as their interaction, in maintaining infant attention (Suarez-Rivera et al., 2019; Wass et al., 2018a), using behavioural methods alone, it impossible to examine this fully (Wass et al., 2020). The next section outlines the potential of neurocognitive methods to furthering our understanding of the mechanisms that drive infant attention in early infant-caregiver interactions.

2.10 Using neurocognitive methods to understand the mechanisms that drive infant attention during naturalistic free-flowing interactions

Understanding the mechanisms that drive, and are elicited by, moment-to-moment inter-dyadic contingencies in naturalistic social interactions is difficult using behavioural methods alone, given that similar behaviours (e.g. looks towards objects and partners) can occur across different levels of attentional and intentional engagement (Siposova & Carpenter, 2019; Wass et al., 2018a), and, fluctuations in the relationship between endogenous processing and attention could be driven by both exogenous influence (i.e. factors external to the infant) and intrinsically-motivated process (those factors internal to the infant (e.g. epistemic reasoning (Kidd et al., 2012), or interest)). Using neurocognitive methods to examine associations between brain and behaviour during online interactions provides a way to more fully understand the hierarchically organised processes that operate across the dyad during early interactions, and how these processes dynamically structure the allocation of infant attention and influence learning at the sub-second scale (Hamilton, 2021; Redcay et al., 2012; Sebanz & Knoblich, 2009; Wass et al., 2020).

In recent years, the application of neurocognitive methods to examining the dynamics of social interactions has become increasingly popular among adult populations (Hamilton, 2021; Hasson et al., 2012), and more recently, in caregiver-infant dyads (Leong, Byrne, Clackson, Georgieva, et al., 2017). Similar to much of the behavioural research conducted on early joint attention development, work examining the neurocognitive mechanisms supporting early social interaction and communication has largely been conducted in controlled experimental paradigms, where the neural activity of participants is recorded in response to pre-recorded and live experimenters engaging in socially ostensive signalling

(Parise & Csibra, 2013; Senju et al., 2006). This work has been insightful in demonstrating neural responsivity by infants to ostensive cues from as early as 2-3 months, as well as giving insight into socially guided information encoding processes that support early learning (Begus et al., 2016). Yet, it is clear from the micro-behavioural work outlined above that, to understand the mechanisms that control infant attention, and dynamically support learning, we need to record infant brain activity during reciprocal free-flowing interactions (Yu & Smith, 2017).

Electroencephalography (EEG) measures electrical brain activity at the sub-second scale, and can be decomposed into different oscillatory frequency bands: delta (1-3Hz), theta (3-6Hz), alpha (6-9Hz), beta (9-20Hz) and gamma (> 20Hz), which are generally slower for infants, compared to the functionally equivalent bands examined in adults (Georgieva et al., 2020; Orekhova et al., 2006). Activity at different frequencies has been associated with a wide range of cognitive processes in early infancy, including attention (Xie et al., 2018), the processing of ostensive social signals (Senju et al., 2006), action imitation and production (C. D. Monroy et al., 2019), and object recognition (Begus et al., 2015). Infants' oscillatory activity occurring at theta and alpha frequencies has been extensively investigated: associating with attentional and encoding processes particularly relevant to understanding the intra- and inter-individual dynamics that drive and maintain infant attention, and inter-dyadic entrainment in shared interactions.

In infancy theta activity (occurring around 3-6Hz) has been associated with endogenously controlled episodes of attention, as well as information encoding processes (Begus & Bonawitz, 2020). For example, Orekhova et al. (1999) compared infant EEG activity during periods of anticipatory attention, where infants anticipated the re-appearance of an experimenter from behind a screen during a peak-a-boo game, to a condition where infant attention was externally elicited by an object appearing abruptly in the infant's visual field. Theta activity during episodes of anticipatory attention was significantly higher, compared to episodes of externally-driven attention, and infants' ability to maintain their attention during these periods of anticipatory attention correlated with infant theta activity occurring over fronto-central electrodes. The association between endogenously-driven episodes of sustained attention and increases in infant theta activity has been further supported in both-screen-based, and naturalistic object play (Orekhova, 1999; Wass et al., 2018; Xie et al., 2018). Recording infant EEG activity whilst 6-12-month-old infants watched cartoon videos, Xie et al. (2018) showed that infant theta activity fluctuated over the duration of an attention episode, such that activity over fronto-central electrodes increased where infants entered heart-rate defined periods of sustained attention. This effect reached significance at 10 months. Corresponding to this finding, during solitary naturalistic play with objects, infant theta activity over central electrodes in the time before and just after infants look towards an object has been found to predict the duration of infant attention towards that object (Orekhova et al., 2006; Wass et al., 2018b).

As well as associating with endogenously controlled patterns of attention in early infancy, increases in theta activity also correlate with information encoding processes. For example, averaging infant EEG activity over each object-directed attention episode during free-moving object exploration, Begus et al. (2015) have shown that infants' later recognition of the same objects was better for those that elicited greater increases in fronto-central theta activity. In 12-month-old infants, theta activity has also been associated with an expectation to receive information from an individual that they perceive as being informative (Begus et al., 2016). The extent to which these findings implicate theta in leaning processes in infancy is unclear (cf. Begus & Bonawitz, 2020), but they nevertheless add further support to the association between theta activity and endogenously controlled attention. Given this association, theta power is a particularly informative oscillatory rhythm to study in naturalistic interactions to understand if, and where, infant attention is endogenously guided, and how fluctuations in endogenous neural activity associate with the interactive dynamics of the interaction (Wass et al., 2018b).

Alpha activity occurring at 8-12Hz in adults, and around 6-9Hz in infancy has also been extensively studied relative to attentional processes, and linked, in particular, to patterns of online predictive processing (C. D. Monroy et al., 2019; Rayson et al., 2019). In adults, alpha desychronisation (i.e. a reduction in alpha activity) is thought to represent release from inhibition during sensory information processing (Klimesch et al., 2007). Reduced alpha activity has been identified at the onset of a predicted stimulus (Thut, 2006) and, in social paradigms, predicting the outcome of another persons' action is associated with alpha desychronisation over pre-central motor cortices (Hari et al., 2015; Kilner et al., 2004; Muthukumaraswamy et al., 2004). In infants similar patterns of alpha suppression (6-9Hz) have been shown over motor areas when observing the predicted outcome of another individuals' manual behaviour (C. D. Monroy et al., 2019, 2019; Southgate et al., 2010), and has even been associated with infants' learning of predictable regularities when viewing repeated behavioural sequencies (C. D. Monroy et al., 2019). Examining change in oscillatory activity at alpha frequencies around specific behavioural events in an interaction could therefore be particularly informative to our understanding of where infants are predicting and encoding the behavioural contingencies of their partner. In fact, one recent study, conducted in a controlled experimental setting, has shown that when a pre-recorded adult-experimenter follows an infant's gaze towards a cued object presented on a screen, infants aged 9 months showed greater reduction in EEG activity at alpha frequencies, compared to a condition where their gaze was not followed (Rayson et al., 2019). Whether similar patterns of neural activity associate with the online contingencies of naturalistic, free-flowing interactions is yet to be investigated.

To examine associations between infant neural oscillatory activity and their attention or sensitivity to the behaviours of their partner, event-locked and continuous methods of analysis can be employed (Begus et al., 2016; Wass et al., 2018b). In studies where the neurocognitive mechanisms of infant attention are studied in isolation, it is common to event-lock neural activity to a specific behavioural event, such as the direct eye gaze of a pre-recorded experimenter, and analyse changes in neural oscillatory activity around these events, using time-frequency or averaging methods (Parise & Csibra, 2013). In shared interactions the same approach can be taken to compare differences in neural activity occurring in the time just before and after behaviourally defined inter-dyadic moments (e.g. mutual vs. non-mutual gaze; adult- vs. infant-led episodes of shared attention), or over the duration of an attention episode.

In continuous methods of analysis, the relationship between infants’ EEG activity and the behavioural time series of one or both interacting partners is assessed (Piazza et al., 2018; Wass et al., 2018b). This method of analysis involves computing the association between infant EEG activity and a continuous behavioural signal over a range of time lags in order to assess forwards- and backward-predictive relationships between the signals. In this way intra-individual associations between fluctuations in infants’ own brain activity and modulations in their attention can be assessed: are, for example, longer infant look durations forward-predicted by increased in their own endogenous neural activity. Critically, inter-individual dynamics can also be examined. For example, recording fNIRS from 9-month-old infants whilst they engaged in shared interactions with their adult partner, Piazza et al. (2018) showed that fluctuations in infants' pre-frontal cortex activity dynamically forward-predicted modulations in the fundamental frequency of their caregiver's speech.

Examining oscillatory structures in caregiver behavioural signals, it is also possible to assess oscillatory entrainment (see previous section) between infants' neural activity and the behavioural time-series of their partner. In particular, slow varying amplitude modulations in the amplitude envelope of adult- and infant-directed speech oscillate at delta, theta and gamma-rate frequencies, corresponding to the stress, syllabic and phonetic patterning of the phonological information in speech (Goswami & Leong, 2013; Leong & Goswami, 2015). Neural entrainment to the amplitude envelope of the speech signal has been extensively investigated in adults, where concurrent entrainment to theta-rate frequencies is thought to be particularly important to speech processing and intelligibility (Luo & Poeppel, 2007). Recent work with infants has also shown neural entrainment at delta and theta frequencies to pre-recorded sung nursery rhymes (Attaheri et al., 2022), as well as IDS (Jessen et al., 2019; Kalashnikova et al., 2018). Examining speech-brain entrainment during naturalistic free-flowing interactions is particularly important to informing our understanding of early language acquisition and inter-dyadic attention, in opening avenues to test how early phonological processing drives and is driven by intra- and inter-personal attentional process, as well as its association with the timing and complexity of the caregivers' semantic inputs (Nencheva & Lew-Williams, 2022; Yu & Smith, 2012).

2.11 Summary

In summary, micro-behavioural work conducted in naturalistic settings has shown that infant attention is largely dependent on the sensorimotor contingencies of early interaction, and attending with an adult partner leads to immediate increases in infant attentiveness and learning (Yu & Smith, 2013, 2016, 2017). To understand the mechanisms guiding infant attention and supporting inter-dyadic behavioural coordination, however, we need to record infant neural activity at the sub-second scale and examine associations between their endogenous oscillatory activity and inter-dyadic patterns of attention and behaviour.

CHAPTER 3 - Proactive or reactive? Neural oscillatory insight into the leader-follower dynamics of early infant-caregiver interaction

This chapter is a publication of the original article examining infants' neural and behavioural activity in the time before and after infants either lead or follow their partners' attention during an interaction, to test whether infants play a proactive role in creating and maintaining episodes of mutual attention during naturalistic interactions at the end of the first year (Phillips et al., in press). Subheadings, figure placement, figure and table style, and citation style have been adapted to conform to the thesis format. The supplementary materials for this chapter are available in Appendix A.

Abstract

We know that infants' ability to coordinate attention with others towards the end of the first year is fundamental to language acquisition and social cognition. Yet, we understand little about the neural and cognitive mechanisms driving infant attention in shared interaction: do infants play a proactive role in creating episodes of joint attention? Recording EEG from 12-month-old infants whilst they engaged in table-top play with their caregiver, we examined the communicative behaviours and neural activity preceding and following infant- vs. adult-led joint attention. Infant-led episodes of joint attention appeared largely reactive: they were not associated with increased theta power, a neural marker of endogenously driven attention, and infants did not increase their ostensive signals before the initiation. Infants were, however, sensitive to whether their initiations were responded to. When caregivers joined their

attentional focus, infants showed increased alpha suppression, a pattern of neural activity associated with predictive processing. Our results suggest that at 10-12 months, infants are not routinely proactive in creating joint attention episodes yet. They do, however, anticipate behavioural contingency, a potentially foundational mechanism for the emergence of intentional communication.

3.1 Introduction

Temporal and spatial coordination of one's gaze with another's, or joint attention, is fundamental to successful social interaction and shared cognition (Tomasello et al., 2005). Shared perception, afforded by joint attention, is thought to form the basis of shared intentions and human-specific forms of collective actions (Frith & Frith, 2007; Sebanz & Knoblich, 2009). The ability to engage in reciprocally mediated joint attention, where both partners lead and follow each others' attention, develops towards the end of the first year, and is a key milestone in developmental trajectories of language learning and social cognition (Carpenter et al., 1998; Donnellan et al., 2020; Iverson & Goldin-Meadow, 2005). A distinction is made between 'mutual' and 'shared' joint attention. The former involves two individuals mutually attending to the same environmental stimulus together, at the same time; to be considered *shared* attention, however, mutual attention must be intentional – i.e. the partner who leads the other's attention towards a stimulus checks that the other partner has perceived it (e.g. by checking the partner's gaze), and the follower communicates that attention is shared (Siposova & Carpenter, 2019).

The onset of intentional, proactive communication is debated (Donnellan et al., 2020), but a popular view has been that, already by the end of the first year, infants achieve episodes of joint attention through the establishment of shared intentionality; using ostensive signals deliberately, to direct and share the attention of a communicative partner (Donnellan et al., 2020; Tomasello et al., 2005, 2007). For example, 9-12-month-old infants are thought to use declarative gestures and vocalisations to direct the attention of an experimenter (Carpenter et al., 1998), and modify their behaviour depending on their success (Liszkowski, Albrecht, et al., 2008; Liszkowski et al., 2004). It is argued that infants' ability to proactively initiate and engage in triadic forms of shared attention towards the end of the first year is catalytic to early language acquisition and socio-cognitive learning, in creating a joint attentional frame where the focus and meaning of the adult partner's communication is not only shared between the adult and the infant, but also common ground between them (i.e., both partners are attending to the same thing, and they both know that the other partner is attending to the same thing as them (Siposova & Carpenter, 2019; Tomasello et al., 2007)). More recently, it has been suggested that infants initiate joint attention, not only to share attention, but to directly elicit information from a social partner about their environment: communicating intentionally and actively to regulate when and how they learn (Begus & Southgate, 2018). For example, infants aged 12-months point in an interrogative manner (Begus et al., 2014; Kovács et al., 2014), and, by 20-months, they look towards their caregiver to ask for help when uncertain (Goupil et al., 2016).

However, much previous work on joint attention development has been conducted using structured, experimental paradigms, where a researcher engages in clear, repetitive behaviours aimed at eliciting either a response to their initiations for shared attention in the infant, or an initiation for shared attention *by* the infant. On each experimental trial, therefore, the adult's behaviour is spatially precise and temporally stable (Yu & Smith, 2013): far from the fast-changing multi-layered complexity of naturalistic, free-flowing interactions (Sebanz

& Knoblich, 2009; Wass et al., 2018a; Wass et al., 2020). This saliency and predictability may help infants deploy specific "communicative" behaviours at an age where they would not necessarily do so spontaneously during naturalistic interactions with their caregivers, and without necessarily grasping the communicative nature of these behaviours yet. That is, over the course of the trial-by-trial repetitions, infants might learn that a change in their behaviour affects change in the experimenter, giving rise to a behavioural response otherwise absent in sporadic, isolated, naturalistic contexts.

Consistent with this idea, recent micro-behavioural analysis of caregiver-infant table-top play has shown that, in fact, during naturalistic interactions at the end of the first year, infants rarely engage in active attention-sharing behaviours. For instance, they have been found to look to their caregivers infrequently (Wass et al., 2018a; Yu & Smith, 2013, 2017), and check the focus of their partners' gaze before following their attention less than 10% of the time (Yu & Smith, 2013). Instead of routinely deploying communicative behaviours, infants at this age most often look directly towards objects, and join their caregiver's gaze through attending towards the adults' hands, as they manipulate the attended object (Yu & Smith, 2013, 2017). Infrequent looks to the caregivers' face by 13-14-month-old infants has also been observed during free-moving, naturalistic interactions, both in laboratory settings, and home-based recordings (Custode & Tamis-LeMonda, 2020; Franchak et al., 2011).

These findings from naturalistic interactions challenge the view that, towards the end of the first year, joint attention is already frequently achieved through proactive communication from the infant (i.e. using ones' own gaze to signal communicative intention, and using partner gaze to infer intention), and suggest that, at this point, shifts in infant attention might instead be mostly *reactive* to the behaviours of their partner. That infant attention is predominantly reactive during online social interaction has important implications for our current understanding of the learning mechanisms involved in the development of joint attention, and how these mechanisms support language acquisition, as well as early socio-cognitive skills (Yu & Smith, 2012).

However, understanding how joint attention is established in caregiver-infant dyads, and addressing the mechanisms driving infant attention in shared interaction, is difficult using behavioural methods alone. This is because similar behaviours (e.g., looks towards objects or partners) can occur across different levels of attentional and intentional engagement (Siposova & Carpenter, 2019; Wass et al., 2018a). Electroencephalography (EEG) provides a method to explore sub-second changes in neural activity at different oscillatory frequencies, which have previously been associated with broad mechanisms of cognitive engagement, in infancy and adulthood (Wass et al., 2020). Comparing EEG activity before, after and during specific inter-dyadic moments in a free-flowing interaction thus allows insight into the fast-changing cognitive processes that govern how each partner's attention is allocated.

Theta activity (3-6Hz) is an oscillatory rhythm associated with endogenously driven attention and information-encoding processes in early infancy (Begus & Bonawitz, 2020). In particular, EEG activity in the theta range has been found to increase over fronto-central electrodes during episodes of endogenously controlled attention. For example, theta activity increases where infants anticipate the next action of an experimenter (Orekhova et al., 1999), and whilst 10-12 month-old infants view cartoon videos, theta activity increases over frontal electrodes during heart-rate defined periods of sustained attention (Xie et al., 2018). Both anticipatory looking and engaging in bouts of sustained attention rely on infants' skill in endogenously controlling how their attention is allocated in the environment. Corresponding to this, fronto-central theta activity increases during self-guided object exploration, and theta activity occurring in the time before infants look towards an object has been found to predict the length of time infants pay attention to that object during solitary play (Orekhova et al., 2006; Wass et al., 2018b).

We hypothesised that, if controlled top-down processes drive infant attention when they lead their partner's attention towards an object, theta activity would increase in the time-window preceding infant-initiated looks to mutual attention, compared to adult-led looks. To explore whether communicative signalling necessary for *shared* joint attention (Siposova & Carpenter, 2019) also preceded moments of infant-led mutual attention, we compared the probability of infants looking to their partner or vocalising in the time before look onset. Based on findings from experimental paradigms, an increase in ostensive signalling before infant-led attention was expected (Liszkowski, Albrecht, et al., 2008; Liszkowski et al., 2004). As a secondary research question, we also examined whether proactive engagement with their partner in the time before an infant-led look affected whether the look was followed by the caregiver. It was hypothesised that infant theta activity and their use of ostensive signals would increase in the time before infant-led looks to mutual attention, compared to nonmutual attention.

A key process involved in the deliberate and intentional re-orientation of a social-partners' attention in shared interaction, is the anticipation of the partner's response in the time after the initiation (Hamilton, 2021; Hasson & Frith, 2016; Konvalinka & Roepstorff, 2012; Siposova & Carpenter, 2019). We therefore also compared infant neural oscillatory activity and ostensive signalling occurring immediately after the onset of infant- and adult-led looks to mutual attention. Naturalistic, observational studies have shown that infants are sensitive to the contingency of an adult partner. For example, responding contingently to an infant's gestures immediately improves the quality and quantity of the attention that they pay to objects (Mason, 2019; Mason et al., 2019); and when caregivers behave redirectively (i.e. non-contingently), infants' visual attention durations immediately decrease (Miller et al., 2018; Miller & Gros-Louis, 2013). To our knowledge, however, no previous work has investigated whether infants proactively anticipate, or predict, a response by the partner to their behaviour, i.e. do they check whether their partner has perceived their new attentional focus, and communicate about it, once attention is shared (Siposova & Carpenter, 2019).

As well as examining infants' behavioural cues signalling the anticipation of joint attention *after* leading their partner's attention, here, we also investigate whether we can identify neural markers of predictive processing in the time following gaze onset. In adults, alpha desychronisation is thought to represent release from inhibition during sensory information processing (Klimesch et al., 2007). Reduced alpha activity has been identified at the onset of a predicted stimulus (Thut, 2006) and, in social paradigms, predicting the outcome of another persons' action is associated with alpha desychronisation over pre-central motor cortices (Hari et al., 1998; Kilner et al., 2004; Muthukumaraswamy et al., 2004). In infancy, similar patterns of alpha suppression (6-9Hz) have been shown over motor areas when observing the predicted outcome of another individuals' manual behaviour (M. Meyer et al., 2016; C. Monroy et al., 2021; Southgate et al., 2010), and one recent study also showed alpha desynchronisation over central-parietal areas when infants viewed the behavioural response of a video-recorded experimenter to their own behaviour, who followed the infant's gaze towards an object (Rayson et al., 2019) (see also ref. Hoehl et al., 2014)). If infants anticipate the behavioural response of their partner where they lead a look towards an object, alpha desychronisation would be expected to occur in the time after infant-led looks to mutual attention; with infants encoding the predicted outcome of their initiation towards an object, on their partner's behaviour (i.e. following their attention).

Based on the view that infants deliberately and proactively initiate shared attention with their partner during social interaction (Tomasello et al., 2005, 2007), we hypothesised that infant looks to their partner's face would increase in the time after infant-led looks to mutual attention (i.e. that they would check whether their partner had followed their attention towards a new object of interest). It was further hypothesised that infant vocalisations would show some increase in the time after infant- and adult-led looks to mutual attention, with the infant communicating, intentionally, to their partner about the shared focus of attention (Siposova & Carpenter, 2019; Tomasello et al., 2007). Consistent with previous neurophysiological findings (Hoehl et al., 2014; M. Meyer et al., 2016; C. Monroy et al., 2021; Rayson et al., 2019; Southgate et al., 2010), we hypothesised that, if infants anticipate the behavioural contingency of their adult partner where they lead attention towards an object, decreased oscillatory activity in the alpha range (6-9Hz) would occur in the time after infant-initiated looks to mutual attention, compared to adult-initiated looks, and infant-initiated looks to nonmutual attention.

3.2 Methods

3.2.1 Participants

Fifty-eight caregiver-infant dyads took part in the study; 37 participants contributed useable data (13 excluded due to recording error, 2 excluded due to infant fussiness, 6 excluded due to poor quality infant EEG (see artifact rejection and pre-processing section for more information on EEG exclusion criteria)). Power calculations estimated in the Leverhulme Trust grant (RPG-2018-281) suggested that a sample of this size should offer >96% power for two-tailed significance of 0.05. The final sample included 18 females and 19 males; mean age, 11.12 months (SD=1.33). All caregivers were female. Participants were recruited through baby groups and Children's' Centers in the Boroughs of Newham and Tower Hamlets, as well as through online platforms such as Facebook, Twitter and Instagram. Written informed consent was obtained from all participants before taking part in the study, and consent to publish was obtained for all identifiable images used. All experimental procedures were reviewed and approved by the University of East London Ethics Committee.

3.2.2 Experimental set-up

Caregivers and infants were seated facing each other on opposite sides of a 65cm wide table. Infants were seated in a high-chair, within easy reach of the toys (see Figure 3.1c). The shared toy play comprised two sections, with a different set of toys in each section, each lasting ~5 minutes each. Two different sets of three small, age-appropriate toys were used in each section; this number was chosen to encourage caregiver and infant attention to move between the objects, whilst leaving the table uncluttered enough for caregiver and infant gaze behaviour to be accurately recorded (cf. Yu & Smith, 2017).

At the beginning of the play session, a researcher placed the toys on the table, in the same order for each participant, and asked the caregiver to play with their infant just as they would at home. Both researchers stayed behind a screen out of view of caregiver and infant, except for the short break between play sessions. The mean length of joint toy play recorded, combining the first and second play sections was 9.92 minutes (SD=2.31).

3.2.3 Equipment

EEG signals were recorded using a 32-channel BioSemi gel-based ActiveTwo system with a sampling rate of 512Hz with no online filtering using Actiview Software. The interaction was filmed using three Canon LEGRIA HF R806 camcorders recording at 50 fps. Caregiver and infant vocalisations were also recorded throughout the play session, using a ZOOM H4n Pro Handy Recorder and Sennheiner EW 112P G4-R receiver.

Two cameras faced the infant: one placed on the left of the caregiver, and one on the right (see Figure 3.1c). Cameras were placed so that the infant's gaze and the three objects placed on the table were clearly visible, as well as a side-view of the caregiver's torso and head. One camera faced the caregiver, positioned just behind the left or right side of the infant's high-chair (counter-balanced across participants). One microphone was attached to the caregiver's clothing and the other to the infant's high-chair.

Caregiver and infant cameras were synchronised to the EEG via radio frequency (RF) receiver LED boxes attached to each camera. The RF boxes simultaneously received trigger signals from a single source (computer running MATLAB) at the beginning of each play section, and concurrently emitted light impulses, visible in each camera. Microphone data was synchronised with the infants' video stream via a xylophone ding recorded in the infant camera and both microphones, which was hand identified in the recordings by trained coders. All systems were extensively tested and found to be free of latency and drift between EEG, camera and microphone to an accuracy of +/- 20 ms.

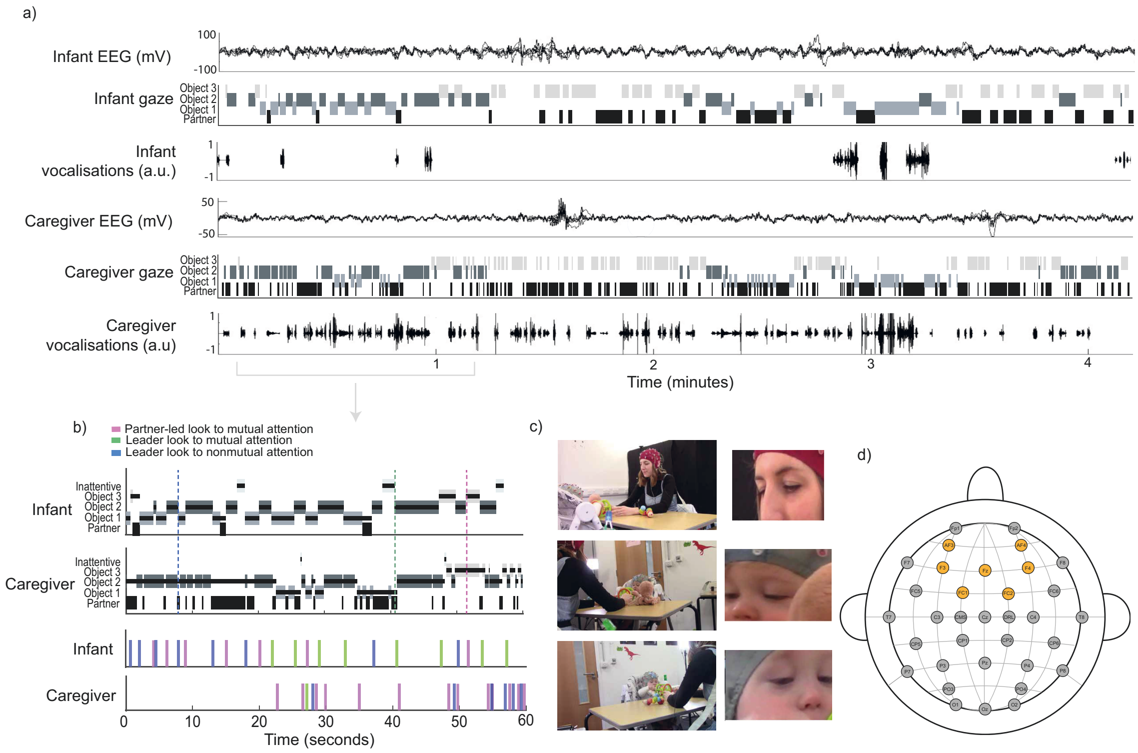

Figure 3. 1 Example data collected during one five minute interaction for one dyad, camera angles used for coding and EEG montage. a) Raw data sample, showing (from top) infant EEG over fronto-central electrodes, after pre-processing, infant gaze behaviour, infant vocalisations, adult EEG over fronto-central electrodes, adult gaze behaviour, adult vocalisations. b) Example of interpolated looks (thin black lines) superimposed on caregiver and infant looking behaviour before interpolation (thick grey lines). Coloured dashed lined indicate examples of different look types in the infant gaze time series (top). Spike trains for infant and caregiver looks, coloured according to look type (bottom). c) Example camera angles for caregiver and infant (right and left), as well as zoomed-in images of caregiver and infant faces, used for coding. d) Topographical map showing electrode locations on the bio-semi 32-cap; fronto-central electrodes included in the infant time-frequency analysis are highlighted in orange (AF3, AF4, FC1, FC2, F3, F4, Fz).

3.2.4 Video coding

The visual attention of caregiver and infant was manually coded using custom-built MATLAB scripts that provided a zoomed-in image of caregiver and infant faces (see Figure 3.1c). Coders indicated the start frame (i.e. to the closest 20ms, at 50fps) that caregiver or infant looked to one of the three objects, to their partner, or looked away from the objects or their partner (i.e. became inattentive). Partner looks included all looks to the partner's face; looks to any other parts of the body or the cap were coded as inattentive. Periods where the researcher was within camera frame were marked as uncodable, as well as instances where the caregiver or infant gaze was blocked or obscured by an object, or their eyes were outside the camera frame. Video coding was completed by two coders, who were trained by the first author. Inter-rater reliability analysis on 10% of coded interactions (conducted on either play section 1 or play section 2), dividing data into 20ms bins, indicated strong reliability between coders (kappa=0.9 for caregiver coding and kappa=0.8 for infant coding).

3.2.5 Vocalisation coding

The onset and offset times of caregiver and infant vocalisations were coded using custom-built MATLAB scripts that allowed coders to identify the onset and offset of a vocalisation based on the spectrogram, as well as auditory sound. A vocalisation was defined as a continuous sound produced by the caregiver or infant, with a pause less than 500ms. Due to the labour-intensive nature of the vocalisation coding, vocal coding was completed for a sub-sample of the caregiver-infant dyads (n=19). Inter-rater reliability on 10% of coded interactions (conducted on either play section 1 or play section 2), dividing data into 1ms bins, again indicated strong reliability between coders (kappa=0.8).

3.2.6 Behavioural look extraction and analysis

3.2.6.1 Data pre-processing

The aim of our analysis was to identify moments where the infant's attention transitioned from one play object to another, and to examine whether the infant or the caregiver initiated the transition. Before doing this, however, we first interpolated through infant and caregiver looks to their partner. This is because, as shown in Figure 3.1a, 3.1b, during periods of concurrent looking towards an object, caregivers, and, to a lesser extent, infants, alternated their attention frequently between the object and their partner. Without interpolation, each subsequent look back to the object would be classified as a separate follower look to the object. This procedure thus allowed us to accurately identify moments in the interaction where the infant was leading and following their partner's attention, while considering the dynamic nature of joint attention documented in previous studies (Yu & Smith, 2013).

Interpolation involved identifying moments where the caregiver or infant looked up to their partner and then interpolating through that look, so that the partner look became an extension of the preceding object look. No threshold was set for interpolation: a new look was considered to have started at the beginning of each new object look (see Figure 3.1b). After interpolation, the first and last frame of all attention episodes were extracted. Infant object looks were categorised into adult-led and infant-led looks. Infant-led looks were subdivided into two further categories: infant-led looks to mutual and nonmutual attention (see Table 1 for description of each look category). Looks that followed or preceded uncodable gaze behaviour were excluded from analysis, as well as leader looks where the partner's gaze in the time after look onset preceded an uncodable period.

Attention episode	Definition
Adult-led looks to mutual attention	The start of the attention episode was taken from the frame that the infant first shifted their gaze towards an object that the adult was already looking towards, at any point in the time that the adult was still attending towards the object.
Infant-led looks to mutual attention	The start of the attention episode was taken from the frame that the infant first shifted their gaze towards an object that the parent was not already looking at, and the adult subsequently joined the infant's gaze towards the object, at any point in the time that the infant was still attending towards the object.
Infant-led looks to nonmutual attention	The start of the attention episode was taken from the frame that the infant first shifted their gaze towards an object that the adult was not already looking at, and the adult did not follow the infant's look towards the object at any point in the time that the infant was still attending towards the object.

Table 1.1 Definition of infant attention episode categories.

3.2.6.2 Cluster-based permutation analysis

To test for significant differences in the likelihood of ostensive signals during the time-periods before and after infant-led and adult-led looks, a permutation-based temporal clustering analysis was conducted (Maris & Oostenveld, 2007). This approach controls for family-wise error rate using a non-parametric Monte Carlo method. For each comparison, a t-statistic was independently calculated at each frame, using a paired t-test, and significant effects were thresholded at an alpha level of 0.05 (two-tailed). Clusters constituted consecutive frames with significant effects. One thousand permutations were then conducted whereby one data stream was shuffled randomly in time, and the largest cluster of contiguous significant effects identified. The Monte Carlo estimate of the permutation p-value was calculated as the proportion of random clusters longer than the clusters found in the observed data. Contiguous clusters >95th centile were considered significant, corresponding to a critical alpha level of 0.05 (two-sided test).

3.2.7 Infant EEG analysis

3.2.7.1 Artifact rejection and pre-processing

A fully automatic artifact rejection procedure including ICA was adopted, following procedures from commonly used toolboxes for EEG pre-processing in adults (Bigdely-Shamlo et al., 2015; Mullen, 2012) and infants (Debnath et al., 2020; Gabard-Durnam et al., 2018), and optimised and tested for use with our naturalistic infant EEG data (Georgieva et al., 2020; Marriott Haresign et al., 2021). This was composed of the following steps: first, EEG data were high-pass filtered at 1Hz (FIR filter with a Hamming window applied: order 3381 and 0.25/ 25% transition slope, passband edge of 1Hz and a cut-off frequency at -6dB of 0.75Hz). Although there is debate over the appropriateness of high pass filters when measuring ERPs (see (Widmann & Schröger, 2012), previous work suggests that this approach obtains the best possible ICA decomposition with our data (Dimigen, 2020; Marriott Haresign et al., 2021). Second, line noise was eliminated using the EEGLAB (Bigdely-Shamlo et al., 2015) function *clean_line.m* (Mullen, 2012).

Third, the data were referenced to a robust average reference (Bigdely-Shamlo et al., 2015). The robust reference was obtained by rejecting channels using the EEGLAB *clean_channels.m* function with the default settings and averaging the remaining channels. Fourth, noisy channels were rejected, using the EEGLAB function *clean_channels.m.* The function input parameters 'correlation threshold' and 'noise threshold' (inputs one and two) were set at 0.7 and 3 respectively; all other input parameters were set at their default values. Fifth, the channels identified in the previous stage were interpolated back, using the EEGLAB function eeg_interp.m. Interpolation is commonly carried out either before or after ICA cleaning but, in general, has been shown to make little difference to the overall decomposition (Delorme & Makeig, 2004) . Infants with over 21% (7) electrodes interpolated were excluded from analysis. After exclusion, the mean number of electrodes interpolated for infants was 0.19 (SD=0.67) for play section 1, and 2.36 (SD=1.87) for play section 2.

Sixth, the data were low-pass filtered at 20Hz, again using an FIR filter with a Hamming window applied identically to the high-pass filter. Seventh, continuous data were automatically rejected in a sliding 1s epoch based on the percentage of channels (set here at 70% of channels) that exceed 5 standard deviations of the mean channel EEG power. For example, if more than 70% of channels in each 1-sec epoch exceed 5 times the standard deviation of the mean power for all channels then this epoch is marked for rejection. This step was applied very coarsely to remove only the very worst sections of data (where almost all channels were affected), which can arise during times when infants fuss or pull the caps. This step was applied at this point in the pipeline so that these sections of data were not inputted into the ICA. The mean percentage of data removed in play section 1 was 7.96 (SD=7.44), and 3.16 (SD=4.32) for play section 2.

Data collected from the entire course of the play session (including play section 1 and play section two, as well as two further five minute interactions) were then concatenated and ICAs were computed on the continuous data using the EEGLAB function runica.m. The mean percentage of ICA components rejected was 52.03% (SD=19.18). After ICA rejection, data from each play section were re-split. For representative examples of artifactual ICA components identified in our naturalistic EEG data, see Marriott-Haresign et al. (2021).

3.2.7.2 Time-frequency analysis

Each infant look onset was identified in the EEG signal, and activity occurring 2500ms before to 2500ms after look onset extracted, across all channels. An additional 200ms was also extracted immediately prior to this segment to serve as the pre-look baseline. Only look epochs with 25% or fewer data points excluded during artifact rejection were included in analysis, and missing data points were set to NaN.

Time-frequency decomposition was conducted on each look epoch via continuous Morlet wavelet convolution, whereby the EEG signal at each channel was convolved with Gaussian-windowed complex sine-waves, ranging from 1-16Hz, in linearly spaced intervals. This frequency range was selected as the frequency range least sensitive to movement artifacts inherent in naturalistic infant EEG, which affects both low (<2Hz) and high (>16Hz) frequency activity (Georgieva et al., 2020). The width of the Gaussian was set to 7 cycles. Before wavelet convolution, the epoched data was reshaped into continuous data, and afterwards transformed back to individual epochs. To remove distortion introduced by wavelet convolution, the first and last 500ms of each epoch was chopped off, so that the epochs were 4200ms in length. After convolution, power was extracted as the absolute value squared, resulting from the complex signal, before averaging power values at each time point over all looks. The condition-specific baseline period used was 2200-2000ms before look onset. Averaged power time-series occurring 2000ms before and after look onset were normalised by transforming the baseline-corrected signal to a decibel (dB) scale (Cohen, 2014).

3.2.7.3 Cluster-based permutation analysis – EEG data

Two approaches were used for analysing the EEG data. First, 2-dimensional (frequency x time) clusters were calculated based on data collapsed in topographical space, over fronto-central electrodes (Figure 3.1d). Second, 3-dimensional (frequency x time x electrode) clusters were calculated based on the entire data. For the first analysis, normalised power was averaged over fronto-central electrodes (AF3, AF4, FC1, FC2, F3, F4, Fz; see Figure 3.1d for locations), and compared between looks. This electrode cluster was chosen based on previous infant literature (Braithwaite et al., 2020). Only participants contributing 5 usable trials or more to both look categories, in one comparison, were included in each analysis (see Figure S1 for details on the number of epochs included before and after artifact rejection, for each type of attention episode).

For the first analysis, the normalised power time-series before and after look onset was compared for each look category, using the FieldTrip function ft_frqstatistics (Maris & Oostenveld, 2007). The cluster-based permutation approach controls for family-wise error rate using a non-parametric Monte Carlo method. Corresponding time*frequency points were compared between conditions using paired sample t-tests, and adjacent significant time*frequency points ($p<0.05$, two-tailed) were clustered together using the 'maxsum' cluster statistic, which sums together the t statistics of each significant time-frequency point in each cluster. The largest cluster was retained. This procedure was then repeated 1000 times, randomising and reshuffling participant data points between conditions on each permutation. The Monte Carlo estimate of the permutation p-value was calculated as the proportion summed test statistics larger than the observed summed t-statistic: here, clusters with a summed t-statistic >95th centile (corresponding to a p-value of 0.025 (two-sided test)) were accepted as significant.

Second, in order to examine how the distribution of results varied topographically over the brain, an additional 3-dimensional cluster-based permutation analysis was conducted to examine time-frequency-electrode space for clusters of significant data points. All 32 channels, at frequencies 2-16Hz were included in the permutation. The ‘minnbchan’ parameter was set to 0. Clusters with a p-value <0.025 (two-sided test) were considered significant.

3.3 Results

The results section is in three parts. Section 3.3.1 presents descriptive statistics on infant and adult gaze and vocal behaviour. Section 3.3.2 compares the attentional, behavioural and neural dynamics preceding a) infant-led vs adult-led looks to mutual attention, and b) infant-led looks to mutual vs nonmutual attention. Section 3.3.3 repeats this analysis in the time-period following look onset. As behavioural cues are slower-changing in comparison to EEG activity, a 5000ms time-window was used to compare infant behaviour in the time before and after a look onset (customary for this type of research (Yu & Smith, 2013)), whilst a 2000ms time-window examined infant EEG activity. See Table 3.1 (Methods section) for a detailed description of how adult-led looks to mutual attention and infant-led looks to mutual attention and nonmutual attention were defined.

3.3.1 Descriptive statistics

Prior to testing our main hypotheses, we conducted three descriptive analyses. Before interpolating through looks to partner, we investigated the proportion of time that caregivers and infants spent vocalising, looking to their partner, attending to objects, and inattentive during the interaction (Figure 3.2a, 3.2b). Second, after interpolating through looks to partner, we tested how many times per minute infants and adults engaged in episodes of mutual attention (infant or adult-led), and nonmutual attention. Finally, we examined the length of infant attention episodes, and of caregiver-infant mutual attention episodes (Figure 3.2c, 3.2d).

Infants spent the majority of the time looking towards objects, whereas caregivers divided their attention between their infant and the objects (consistent with Yu & Smith, 2013; Figure 3.2a). Infant vocalisations were infrequent, whereas adult vocalisations were more frequent (Figure 3.2b). Comparisons between caregivers and infants were significant using two-tailed independent t-tests: looks to objects [$t(72)=10.81$, $p<0.001$, $d=0.84$], looks to partner [$t(72)=-14.01$, $p<0.001$, $d=-1.02$] and vocalisations [$t(36)=-5.61$, p<0.001, $d=-0.60$]. The proportion of time spent in states of inattention did not differ [$t\ (72)=1.32$, $p=0.198$, $d=0.08$].

Infant object look durations were positively skewed before log transform (consistent with Yu & Smith, 2016), as were episodes of caregiver-infant mutual attention (Figure 3.1c). The number of times each type of attention episode occurred per minute was similar for caregivers and infants, with leader looks to mutual attention the most infrequently occurring category (Figure 3.1d). Two-tailed independent t-tests showed that infants followed their partner's attention significantly more often per minute compared to their caregivers [$t(72)=2.94$, p=0.004, $d=0.77$]; all other comparisons were not significant (leader to nonmutual looks [$t(72)=1.49$, $p=0.139$, $d=0.35$]; leader to mutual looks [$t(72)=-0.73$, $p=0.467$, $d=-0.14$]).

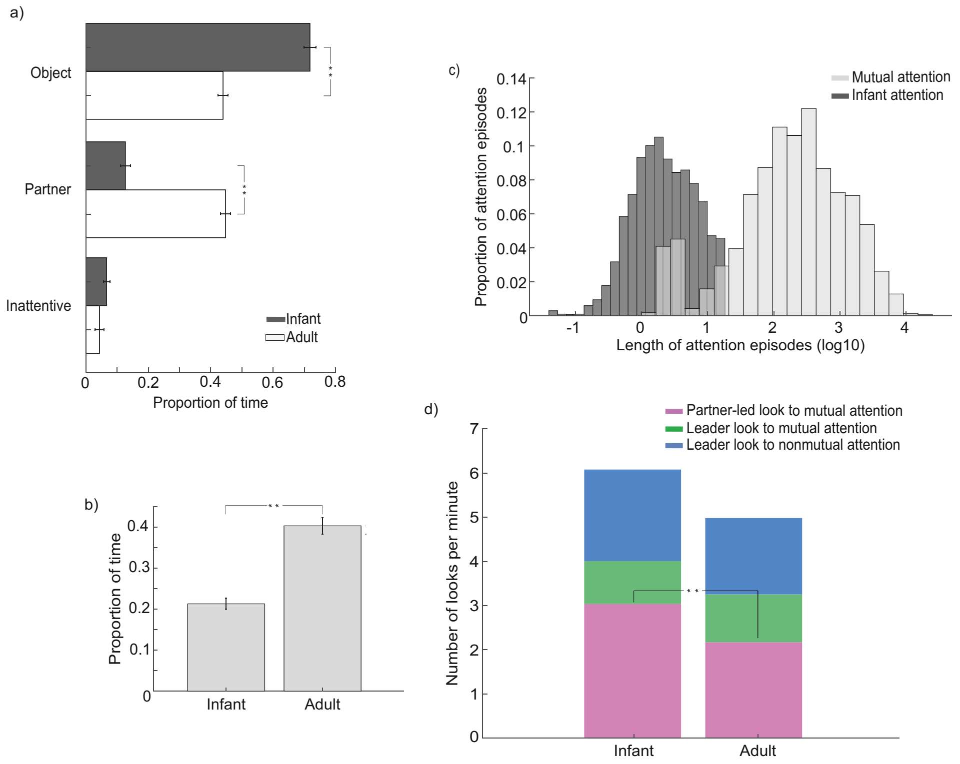

Figure 3. 2 Caregiver and infant attention, and vocal behaviour in shared play. a) Bar plots show the proportion of time caregivers and infants spent looking to their partner, towards objects, and inattentive, during the interaction. Two-tailed independent t-tests (n=37) compared proportions between caregivers and infants for each look category (*p<0.05, **p<0.01; error bars show the SEM). b) Bar plot shows the mean proportion of time caregivers and infants spent vocalising during the interaction. A two-tailed independent t-test (n=37) compared proportions between caregivers and infants (*p<0.05, **p<0.01; error bars show the SEM). c) Histograms show log-transformed infant object look durations, and length of mutual attention episodes, across all types of looks, after interpolation. d) Bar plots show the number of times infants and adults engaged in one of three possible attentional states per minute: partner-led looks to mutual attention, leader looks to mutual attention, leader looks to nonmutual attention. Two-tailed independent t-tests compared the number of

attention episodes per minute between caregivers and infants, for each look category (n=37; *p<0.05, **p<0.01).

3.3.2 Before look onset: are infants proactively initiating joint attention episodes?

This section is in two parts. Section 3.2.2.1 compares infants' use of ostensive signals and their neural oscillatory activity occurring before infant-led looks to mutual attention, and adult-led looks to mutual attention, in order to test for differences between infant- and adult-initiated mutual attention episodes. Section 3.3.2.2 subsequently compares infant-led object looks resulting in mutual and nonmutual attention, in order to test for differences between infant-led looks that were followed, or not followed, by their adult partner.

3.3.2.1 Infant-led vs adult-led mutual attention

Ostensive signals and infant attention

First, we tested whether infants were more likely to use ostensive signals before an infant-led mutual attention episode, compared to where they followed their caregiver's look into mutual attention. If true, this would support the hypothesis that infants proactively lead their caregiver's attention to objects during naturalistic table-top play. To investigate this, we conducted a probability analysis examining ostensive signals in the time-window +/- 5000ms relative to each look type.

For each look type (infant-led to mutual and adult-led to mutual), the frame at which an object look onset occurred was identified in the vocalisation and partner look time-series separately, and the 5000ms preceding look onset extracted. The probability of the behaviour occurring at each 20ms frame was then calculated as the proportion of looks where each ostensive signal (looks to the partner's face and vocalisations) was present in that frame. Results of the probability analysis are presented in Figure 3.3a-d. Cluster-based permutation analysis (see Methods) indicated that infants were significantly more likely to look towards their caregiver in the time-period immediately preceding an episode of infant-led mutual attention, compared to an adult-led mutual attention episode (Figure 3.3a). There were no other significant differences between look types (Figure 3.3a, 3.3b).

Any significant difference between adult-led and infant-led attention could, however, be driven either by an increase relative to baseline in looking prior to infant-led attention episodes, or by a decrease relative to baseline prior to adult-led attention episodes. To differentiate between these hypotheses, we generated a random probability time-series of partner looks and vocalisations in the time immediately before and after each type of infant object look by inserting a random event into each ostensive cue time series, and extracting the 5000ms preceding the event. A cluster-based permutation analysis was again conducted using paired t-tests to investigate where the behavioural time-series differed from chance (Figure 3.3a-d). Results indicated that the probability of infants looking to their partner was below levels expected by chance in the 1s time-period before the onset of adult-led attention. Overall, then, these results suggest that infants are less likely to look to their partner during the time-window preceding adult-led looks to mutual attention.

In addition, to investigate whether infant attentiveness differed in the time before look onset between each look type, using the uninterpolated gaze time series, we also examined how frequently infant attention changed in the 5000ms time-period leading up to each attention episode, and the length of infant gaze towards the previous object for each type of look (Figure 3.3e, 3.3f). Two-tailed paired t-tests showed no difference in the number of object looks occurring 5s before the onset of an infant-led look to mutual attention [mean=1.85, SEM=0.08], compared to an adult-led look to mutual attention [mean=1.17, SEM=0.06; $t(36)=-0.27$, $p=0.792$, $d=-0.24$]; nor was there a difference in the length of infant attention to the previous object [$t(36)=-0.61$, $p=0.544$, $d=-0.10$].

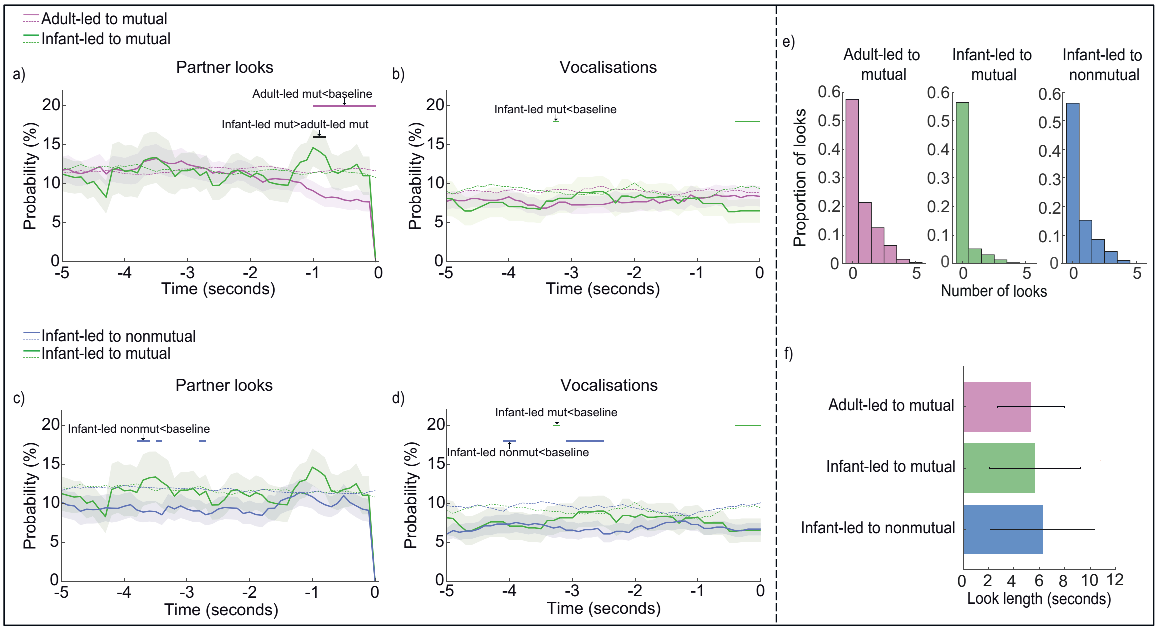

Figure 3. 3 Probability of ostensive signals (partner looks and vocalisations) and infant attentiveness in the time-period before infant look onset. Probability time-course for: a) Partner looks before infant-led mutual attention vs adult-led mutual attention; b) Vocalisations before infant-led mutual attention vs adult-led mutual attention; c) Partner looks before infant-led mutual attention vs infant-led nonmutual attention; d) Vocalisations before infant-led mutual attention vs infant-led nonmutual attention. In each case shaded areas show the SEM (n=37 for a) and c), n=19 for b) and d)), and horizontal black lines show

the areas of significant difference, between attention episodes, identified by the cluster-based permutation analysis (Monte-Carlo p value<0.05). Dotted lines show the baseline time-series, plotted for each attention episode. Horizontal coloured lines show the areas of significant difference between each attention episode and baseline, identified by the cluster-based permutation analysis (Monte-Carlo p value<0.05). e) Histograms show the distribution of the number of object looks in the 5s time-period before look onset for each attention episode. f) Bar plots show the mean length of infant attention towards the object immediately preceding look onset for all three attentional states. Error bars indicate the SEM (n=37). Two-tailed paired t-tests compared the length of infant attention towards the previous object, between each attention episode, which indicated no significant differences.

Neural oscillatory activity

We compared how the neural oscillatory activity differed in the time before infant-led and adult-led mutual attention episodes. Results of the time-frequency analysis are presented in Figure 3.4. Two-dimensional cluster-based permutation analysis revealed no significant clusters of time*frequency points, comparing between infant-led looks to mutual attention, and adult-led looks to mutual attention. Three-dimensional cluster-based permutation analysis, including all electrode by time by frequency points, also revealed no significant clusters. Contrary to what would be expected if infants were deliberately orienting their partners towards objects when shifting their gaze to an unattended object, this primary analysis suggests that there were no significant differences in infants' neural activity in the time-windows before they led their partner's attention towards an object.

In our naturalistic data, some of the epochs included in each look category will also have contained additional object and partner looks during the 2000ms before the onset of the look to which the data were event-locked. Even though eye movement-related artifacts were removed through ICA decomposition during pre-processing (see Methods), we also conducted an additional analysis to examine the possibility that this may have contributed to the null result. The results suggested that it did not: the average proportion of looks with object and partner looks occurring in the time before look onset did not differ between attention episodes (Figure S2, S3, S4). Conducting analyses including looks with no shifts in infant attention before look onset *only* was not possible due to low trial numbers (Figure S2). We therefore conducted a secondary analysis excluding neural activity within each look epoch, in the time before look onset, where an infant was not continuously focused on one object/ the partner (see Figure S5). Two-dimensional cluster-based permutation analysis again revealed no significant clusters of time*frequency points (Figure S5).

For comparison with the behavioral analysis presented at the beginning of this section, Figure S6 shows EEG activity over the same time-windows (-5000ms). Two-dimensional cluster-based permutation analysis again revealed no significant clusters of time*frequency points, comparing between infant-led looks and adult-led looks to mutual attention.

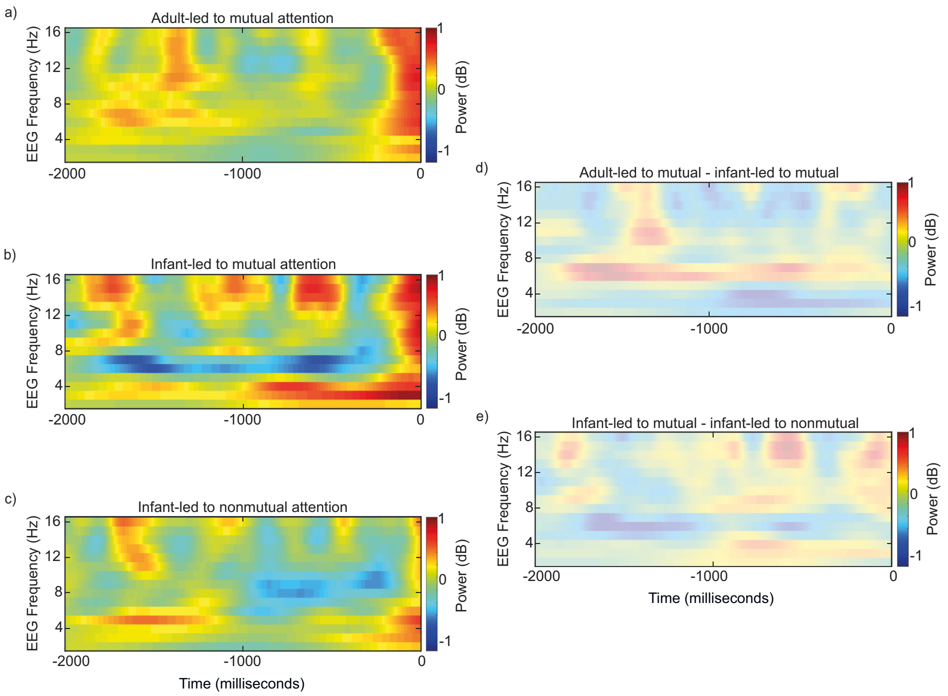

Figure 3. 4 Comparison of infant EEG activity in the 2000ms preceding infant- and adult-led attention episodes. Time-frequency plots show infant EEG activity (2-16Hz) occurring 2000ms before look onset, for (a) adult-led looks to mutual attention, (b) infant-led looks to mutual attention and (c) infant-led looks to nonmutual attention, over fronto-central electrodes (AF3, AF4, FC1, FC2, F3, F4, Fz). Time 0 indicates infant gaze onset. d) Difference in EEG activity between infant- and adult-led looks to mutual attention (adult-led – infant-led). Cluster-based permutation analyses showed no significant clusters of time*frequency points of the difference between attention episodes, and so no significant clusters have been highlighted. e) Difference in EEG activity between infant-led looks to mutual and nonmutual attention (infant-led to mutual - infant-led to nonmutual). Cluster-based permutation analyses showed no significant clusters of time*frequency points of the difference between attention episodes, and so no significant clusters have been highlighted.

3.3.2.2 Followed vs not followed infant-led looks

Ostensive signals and infant attention

Here, we also tested whether infants were more likely to use ostensive signals before an infant-led mutual attention episode, compared to an infant-led nonmutual attention episode, in order to examine differences between infant-led looks that were followed, or not, by their adult partner. No significant differences were observed either in the likelihood of the infant looking to their caregiver in the time-period preceding a look (Figure 3.3c), or in the likelihood of the infant vocalising (Figure 3.3d). In addition, no significant differences were observed in the duration [$t(36)=1.01$, $p=0.321$, $d=0.12$; Figure 3.3f], or number [$t(36)=1.45$, $p=0.157$, $d=0.24$] of infant objects looks in the time-period preceding infant-led looks to mutual attention [mean=1.85, SEM=0.08], and nonmutual attention [mean=1.06, SEM=0.07; Figure 3.3e].

Neural oscillatory activity

We also compared how neural oscillatory activity differed in the time before infant-led mutual attention episodes and infant-led nonmutual attention episodes (Figure 3.4). No significant differences were observed using either the 2-dimensional (Figure 3.4) or the 3-dimensional cluster-based permutation analyses. Again, the number of looks including object and partner looks before each attention episode did not differ (Figure S2, S3, S4). EEG activity occurring 5000ms before look onset for each type of look is presented in Figure S5. Two-dimensional cluster-based permutation analysis revealed no significant clusters of time*frequency points of the difference between attention episodes.

3.3.2.3 Summary

In summary, these results suggest that there is little change in infants' behaviourally ostensive signalling before infant-led mutual attention episodes, compared with adult-led mutual attention. The main finding was a decrease in infant looks to their caregiver in the time before adult-initiated mutual attention. There were no differences in infants' ostensive signalling between infant-led looks that were followed vs not followed by their adult partner.

The neural analyses suggested that there were no differences in neural oscillatory activity before infant-initiated and adult-initiated mutual attention. There were also no differences in neural oscillatory activity between followed vs not followed infant-led attention episodes. There was thus very little evidence that 12-month-old infants proactively initiate joint attention with their partner during shared play.

3.3.3 After look onset: do infants anticipate their gaze being followed?

In this section, we present a similar analysis to section 3.3.2, investigating change in infant behaviour and neural oscilatory activity in the time-period *after* look onset. Again, the section is organised in two parts: first we examine mutual attention, comparing infant-led and adult-led mutual attention episodes (section 3.3.3.1). Second, we compare infant-led attention that was followed vs. not followed by their adult partner (section 3.3.3.2).

3.3.3.1 Infant-led vs adult-led mutual attention

Ostensive signals and infant attention

First, we tested whether infants were more likely to use ostensive signals during the time-period after the start of infant-led, compared to adult-led mutual attention. To investigate this, we conducted the same probability analysis described in section 3.3.2.1, extracting the 5000ms following look-onset from the vocalisation and partner-look time-series. No significant difference in the likelihood of partner looks was observed, but a significant increase in the likelihood of infant vocalisations following adult-led mutual attention was shown (Figure 3.5b). Baseline comparisons suggested that infant vocalisations significantly decreased from baseline in the time after infant-led looks to mutual attention, potentially driving this difference (Figure 3.5b).

We also examined whether infant-led mutual attention episodes tended to be longer lasting than adult-led mutual attention (Figure 3.5e). No significant difference was observed [$t(36)=-1.17$, $p=0.248$, $d=-0.19$]. Finally, we examined the time interval it took caregivers to follow their infant's attention during infant-led looks to mutual attention (Figure 3.5f). This analysis suggested that most looks were followed within 1-2s after look onset [mean=1.49s, SEM=6.91].

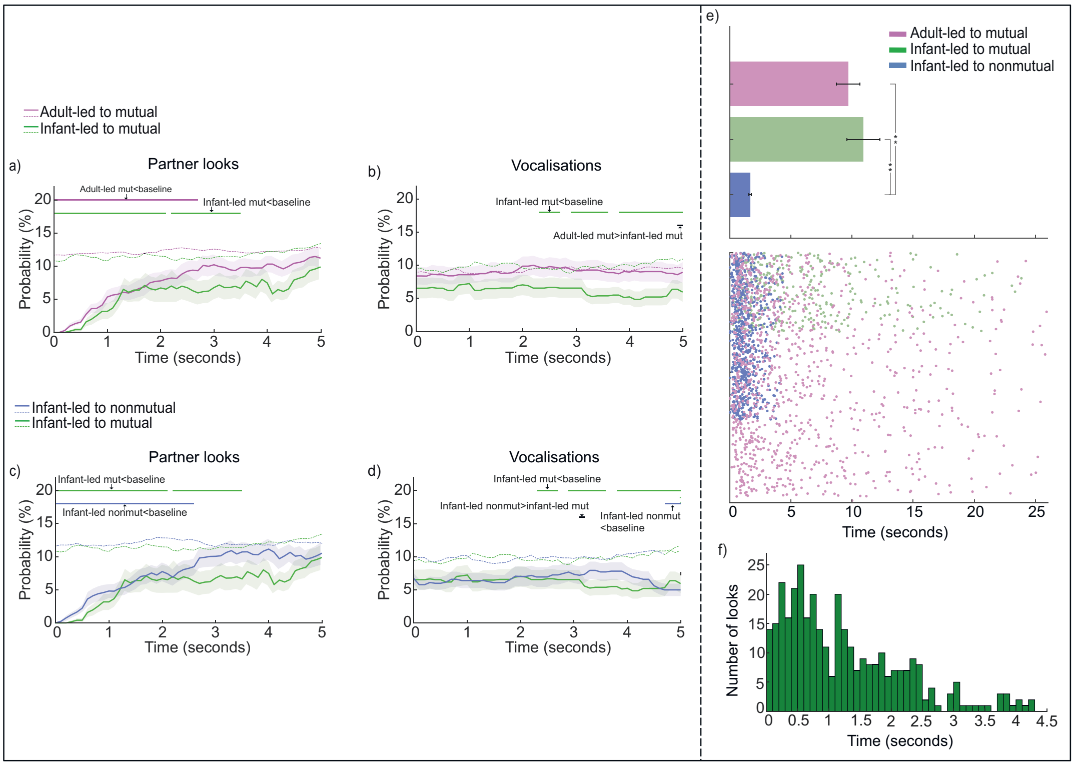

Figure 3. 5 Probability of ostensive signals (partner looks and vocalisations), infant attentiveness, and the time adults took to follow infant attention in the time-period after infant look onset. Probability time-course for: a) Partner looks after infant-led mutual attention vs adult-led mutual attention; b) Vocalisations after infant-led mutual attention vs adult-led mutual attention; c) Partner looks after infant-led mutual attention vs infant-led nonmutual attention; d) Vocalisations after infant-led mutual attention vs infant-led nonmutual attention. In each case shaded areas show the SEM and horizontal black lines show the areas of significant difference, between attention episodes, identified by the cluster-based permutation analysis (Monte-Carlo p value<0.05). Dotted lines show the baseline time-series, plotted for each attention episode. Horizontal coloured lines show the areas of significant difference between each look type and baseline, identified by the cluster-based permutation analysis (Monte-Carlo p value<0.05). e) Length of infant attention towards an object after look onset, for each type of attention episode. The bar plot shows the mean length

of infant attention averaged over participants (error bars show the SEM (n=37)); scatter plot shows the length of each individual look contributing to each object look category, across all participants, after outlier removal. Two-tailed paired t-tests (n=37) compared the difference in the length of infant attention between each type of attention episode (*p<0.05, **p<0.01). f) Histogram shows the distribution of the time it took caregivers to follow infant attention.

Neural oscillatory activity

In this section, we compare differences in infant EEG activity occurring over fronto-central electrodes after look onset for infant-led and adult-led looks to mutual attention (Figure 3.6). Consistent with our hypothesis, infant-led mutual attention episodes led to a decrease in EEG power, particularly in the theta/alpha range towards the end of the 2000ms time-period, as compared to adult-led looks (Figure 3.6a, 3.6b). The 2-dimensional cluster-based permutation analysis identified a significant positive cluster with an average frequency of 7Hz (ranging 5-9Hz), 92-2000ms post look onset (*p*=0.003; Figure 3.6c). Three-dimensional cluster-based permutation analysis also revealed one trend-level positive cluster, with a wide topographical distribution, in the 5-9Hz range (*p*=0.099; see Figure S6).

Again, due to the naturalistic nature of our data, some of the epochs included in these analyses contain additional object and partner looks. Similar to the pre-look analysis, there were too few trials per participant to compare EEG activity occurring during looks without any gaze shifts (see Figure S8). A higher proportion of adult-led looks involved looks to other objects and the partner in the 2000ms time-window, an effect that was driven by a greater number of object looks after the onset of adult-led attention (Figure S8). When the post-look time-period was broken down into 1000ms intervals, however, the difference between infant-led and adult-led looks was only seen in the first 1000ms after look onset (Figure S9,S10), and a high proportion (>70%) of infant- and adult-led looks to mutual attention did not contain any object or partner looks (Figure S7). We nevertheless conducted a secondary analysis, excluding EEG activity for each look, in the time after infants shifted their attention away from the target object, towards another object/ the partner in the 2000ms after look onset. Two-dimensional cluster-based permutation analysis again revealed a significant positive cluster with an average frequency of 7Hz (ranging 5-9Hz), 104-1994ms post look onset (*p*=0.01; Figure S11).

For comparison with the behavioral analysis presented at the beginning of this section, Figure S13 shows EEG activity over the same time-windows (+5000ms). Two-dimensional cluster-based permutation analysis revealed no significant clusters of time*frequency points, comparing between infant-led looks and adult-led looks to mutual attention.

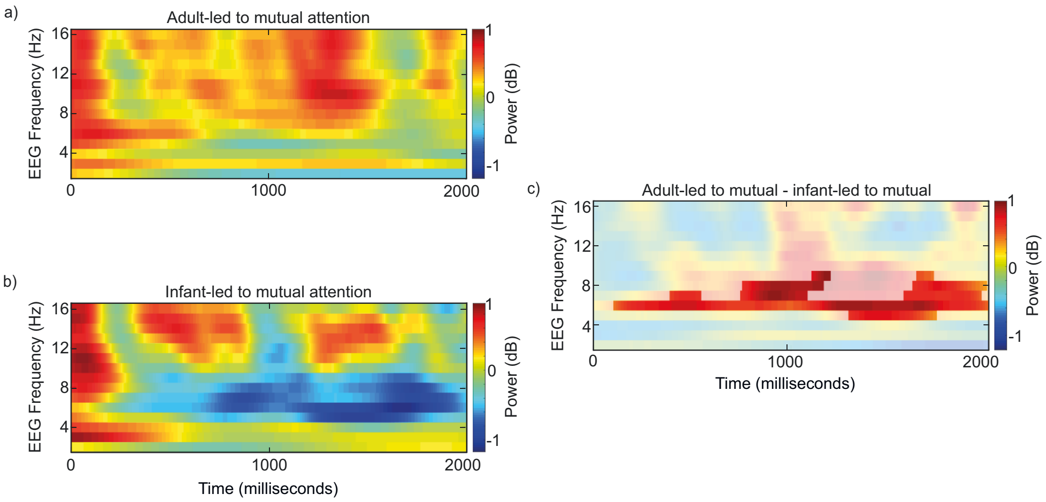

Figure 3. 6 Comparison of infant EEG activity in the 2000ms following infant- and adult-led attention episodes. Time-frequency plots show infant EEG activity (2-16Hz) occurring 2000ms after look onset, for (a) adult-led looks to mutual attention and (b) infant-led looks to mutual attention, over fronto-central electrodes (AF3, AF4, FC1, FC2, F3, F4, Fz). Time 0 indicates infant gaze onset. c) Difference in EEG activity between infant- and adult-led looks to mutual attention (adult-led – infant-led); highlighted area shows the significant positive cluster identified by the cluster-based permutation analysis (p=0.003). The cluster ranges from 5-9Hz, from 92-2000ms post look-onset.

3.3.3.2 Followed vs not followed infant-led attention

Ostensive signals and infant attention

We examined whether ostensive signals differed between infant-led looks that were followed vs. not followed by their adult partner, in the time after look onset. No significant difference in the likelihood of partner looks was observed, but there was a significant increase in the likelihood of infant vocalisations following infant-led nonmutual attention, 3s after look onset (Figure 3.5d). Again, this effect is likely driven by the significant reduction in infant vocalisations from baseline following infant-led looks to mutual attention (Figure 3.3d). Infant-led looks to nonmutual attention lasted a significantly shorter amount of time, compared to infant-led looks to mutual attention, and this difference was marked [$t(36)=6.84$, $p<0.001$, $d=1.13$; Figure 3.5e]. Indeed, mutual attention extended infant attention, irrespective of whether the attention episode was adult- or infant-led, with adult-led mutual attention episodes also lasting significantly longer compared to infant-led nonmutual attention [$t(36)=8.25$, $p<0.001$, $d=1.36$; Figure 3.5e].

Neural oscillatory activity

Corresponding to the significantly shorter object looks during episodes of nonmutual attention, infant-led looks to nonmutual attention included significantly fewer looks that lasted the whole 2000ms after look onset, compared to infant-led looks to mutual attention, resulting in more looks containing object looks and looks to partner combined, both 0-1000ms, and 1000-2000ms after look onset (Figure S7, S8, S9). Due to there being so few infant-led looks resulting in nonmutual attention that lasted the whole 2000ms time-period

(<50%; Figure S7), cluster-based permutation comparing infant-led looks to mutual and nonmutual attention was excluded from analysis.

3.3.3.3 Summary

Consistent with our hypothesis, infant-led mutual attention episodes were accompanied by significantly greater alpha desynchronisation after look onset, compared with adult-led mutual attention. Against our predictions, infants also showed some decrease in their vocalisations after infant-led looks to mutual attention, compared to adult-led looks, and infant-led looks to nonmutual attention, corresponding to a marked decrease from baseline after infant-led looks to mutual attention. No differences in partner looks were observed.

3.4 Discussion

This study investigated whether infants play a proactive role in creating episodes of joint attention during naturalistic tabletop play. In contrast to the results observed using structured, experimental paradigms (Carpenter et al., 1998; Liszkowski, Albrecht, et al., 2008; Liszkowski et al., 2004), our results suggested that, in free-flowing interaction, 12-month-old infants do not readily use their gaze or vocalise before an infant-initiated mutual attention episode; the occurrence of these behaviours throughout the interaction was generally low (Figure 3.2a, 3.2b). Though a significant difference in the probability of partner looks 1s before look onset was identified, baseline comparisons indicated that this was driven by a reduction in infant looks to their partner before adult-initiated looks, rather than an increase before infant-initiated looks (Figure 3.3a). Corresponding to the behavioural findings, and against what would be expected if infants were attempting to proactively drive their caregiver’s attention towards an object, EEG activity at theta frequencies (3-6Hz) did not increase in the 2s before infant-led looks to mutual attention, compared to adult-led looks: cluster-based permutation analysis revealed no significant clusters at any frequency band investigated (Figure 3.4d).

Contrary to our prediction that infants’ proactive engagement with their partner would affect whether a look was followed by the adult, no differences were identified between infant-led looks in ostensive signals (Figure 3.3c, 3.3d), or EEG activity (Figure 3.4e), before look onset. Taken together, the combination of our neural and behavioural results are inconsistent with the idea that infants routinely exert active and intentional control over the allocation of their attention where they lead their partner’s attention, and could suggest that similar processes drive infant attention, when leading a mutual attention episode, and joining the attentional focus of their partner (i.e. adult-led attention).

The null findings reported here are unlikely to be driven by eye-movement related artifact, introduced by temporally variable shifts in infant looking in the time before each look onset. Eye-movement artifacts were removed using ICA decomposition, and, though this does not remove all artifact introduced to the EEG signal (Marriott Haresign et al., 2021), we also show that each look type was equally affected by object and partner looks occurring in the 2000ms preceding look onset (Figure S2). The large sample size included here, particularly for infant EEG research (Noreika et al., 2020), will have also increased signal-to-noise ratio in our data. It is also unlikely that this effect is driven by removal of neural activity during ICA decomposition: the algorithm used to reject ICA components during pre-processing (see Methods), has been shown to be successful in retaining neural signal, especially in comparison to traditional manual rejection techniques (Marriott Haresign et al., 2021).

Furthermore, the secondary analysis conducted, excluding infant neural activity within each epoch, where the infant was not continuously focussed on one object/the partner before look onset also showed no increase in theta power in the time before infant-led looks to mutual attention (Figure S5).

Though infant-led episodes of mutual attention did not appear proactively driven, infants were nevertheless sensitive to whether their look was followed by the adult. In line with hypotheses, in the time-period after look onset, a significant *decrease* in EEG activity was observed over fronto-central electrodes in the alpha band (7Hz), after infant-led looks to mutual attention, compared to adult-led looks (Figure 3.6c). This finding reflects the pattern of neural activity observed in infants whilst observing the predicted outcome of another persons' goal-directed behaviour (Meyer et al., 2022; C. Monroy et al., 2021; Southgate et al., 2010), and is consistent with previous experimental work showing reduced alpha activity where infant gaze was contingently responded to by a video-recorded experimenter (Rayson et al., 2017). Thus, the reduction in EEG activity after infant-led looks to mutual attention may be interpreted as a neural marker of predictive processing during online social interaction, with infants predicting and encoding the behavioural contingency of their partner where they lead a look towards an object, and their partner follows. Against hypotheses, however, infants did not show an increase in looking to their partner in the time after look onset; suggesting that the anticipated contingency of their caregiver was not realised by the infant through observing partner behaviours signalling intention to share attention (Siposova & Carpenter, 2019).

A possible interpretation of our findings is that, rather than shared intentionality, inter-dyadic coordination is largely achieved and perceived by the infant through attending towards their partner’s sensorimotor behaviours. In line with previous findings in naturalistic studies, infants did not readily follow their partner’s gaze (Yu & Smith, 2013). In fact, infants looked to objects more (rather than to their partner) in the 1s time-period before adult-initiated looks. This is consistent with Yu and Smith’s observation that moments infants join their partner’s attention are driven by the partner’s manual activity on objects (Yu & Smith, 2013, 2017). The neural analyses of the current study, that show no increase in endogenous oscillatory activity before infant-led looks, relative to adult-led looks, suggest that similar, external inputs might also drive infant attention where they lead a look towards an object. As well as overt behaviours such as object manipulations and gestural communication, other sensory inputs could also influence shifts in infant gaze. For example, in very early face-to-face interactions, salient events such as pauses in adult vocalisations, and changes in the fundamental frequency of their voice modulate infant attention towards and away from the partners’ face (Crown et al., 2002; Stern et al., 1983). In the current study, analysis of caregivers’ ostensive signalling revealed that partner looks increased and vocalisations decreased in the time before infant-led looks to mutual attention, compared to adult-led looks, and infant-led looks to nonmutual attention (Figure S13).

Entrainment to the low-level sensorimotor dynamics of shared interactions (Sebanz & Knoblich, 2009) could be the mechanism through which infants *perceive* the behavioural contingency of their communicative partner, suggested by the alpha suppression observed after infant-led looks to mutual attention. Research into action-oriented predictive processing suggests that motor intentions actively elicit active predictions about the ongoing consequences of our own actions (Clark, 2013; K. Friston et al., 2012). Perhaps similar processes operate across the dyad during early behavioural coordination, with the infant anticipating the effect of their own action on the behaviour of their partner (Hamilton, 2021;

Konvalinka & Roepstorff, 2012). Again, as well as overt manual behaviours (Yu & Smith, 2013), other fast-changing cues such as temporal and spectral modulations in the partner's vocalisations could also signal behavioural contingency to the infant (Crown et al., 2002; Stern et al., 1983). Interestingly, the probability of infants vocalising falls below baseline in the time after infant-led looks to mutual attention. This finding is possibly indicative of the sensorimotor turn-taking processes occurring after infant-led looks to mutual attention; i.e. that infants are anticipating the behavioural response of their partner in the time after they shift their attention towards a new object.

That said, whilst the findings of the current analysis suggest that infants do not *routinely* show signs of proactively leading their partner's attention during shared interaction, it is still possible that moments of proactive engagement by the infant are infrequent but nevertheless important to the ongoing interactive exchange (Murray, De Pascalis, Bozicevic, et al., 2016). The results of the behavioural analysis, investigating ostensive signals occurring in the time before and after infant-led looks to mutual and nonmutual attention, show that the probability of these behaviours occurring before infants lead a look to their partner is low, but not absent (Figure 3.3a-d; Figure 3.5a-d). An interesting question is whether moments ostensive signals do occur before infants initiate an attention episode with their partner are largely incidental; or whether all or some of these moments occur as a result of the infant attempting to actively engage their adult partner's attention, but remain too rare to make a difference at the statistical level. The mechanisms through which infants perceive the behavioural contingency of their partner during active attention sharing episodes, if they occur, could be functionally different to those engaged during externally driven attention.

Our findings have important implications for how we view and understand the learning processes involved in early joint attention development. Associative learning accounts postulate that infants learn about their environment, and how to act on it, through repeated reinforcement, where the value given to an action is based on previous experience of how that action affected the environment (Deák et al., 2014; Oudeyer & Smith, 2016; L. B. Smith & Breazeal, 2007). In the context of social interaction, infant behaviour is assigned meaning by the adult through consistent and contingent behavioural feedback. Over time, these statistical regularities form the basis for infant representations about the intentions of others, and how their own intentionally-motivated behaviours affect those of their partner (L. B. Smith & Breazeal, 2007). In line with this, the current EEG findings suggest that infants predict and encode the behavioural contingency of their partner to their own actions before they show signs of intentionally initiating joint attention episodes in a routine manner. Caregiver responses to infrequent moments that infants engage in proactive attention-sharing may therefore be particularly important to the development of infants' representations about their own intentionally-motivated behaviour, and, over time, increase the extent to which infants use these behaviourally re-enforced cues to proactively direct the attention of their partner (L. B. Smith & Breazeal, 2007).

This perspective has potential implications for current theories of how infants begin to acquire a language system. A popular view has been that, it is only once infants are able to establish and understand a joint-attentional frame, or 'common ground', between themselves, an object and their partner, that they can begin to engage with and learn from the pragmatic and referential aspects of shared communication (Carpenter et al., 1998; Lieven, 2016; Tomasello & Carpenter, 2007). Our findings, however, consistent with an associative learning framework of joint attention development (L. B. Smith & Breazeal, 2007), suggest that before infants routinely engage in triadic forms of shared attention, infant attunement to their partner's sensorimotor behaviours, and the timing of adult inputs as a function of infant behaviours, may already be contributing towards infants' developing understanding of the principles underlying communication. Recent behavioural work has, in fact, suggested that infants' engagement with objects at the time object labels are presented by their communicative partner could go some way in solving the problem of referential ambiguity (Yu et al., 2021; Yu & Smith, 2012). Combining neural and behavioural methods to explore how infant attunement to action-generated contingencies during naturalistic free-flowing interactions supports early language acquisition should be a key focus for future research.

Whilst this study is the first to show how infant neural activity changes around moments of infant- vs. adult-led episodes of mutual attention during naturalistic interactions, predictive encoding models, investigating the dynamic relationship between infant attention, inter-dyadic behaviour and infant neural activity should be a next step (Jessen et al., 2021). Neural tracking of auditory information to controlled experimental stimuli has been shown in both adults (Zion Golumbic et al., 2013) and more recently, infants (Attaheri et al., 2022; Kalashnikova et al., 2018). Whether and how infants' neural activity dynamically responds to modulations in their partners' behaviours, including features and the timing of caregiver vocalisations, manual activity, and bodily movement, and how this associates with the timing of infant- and adult-led episodes of joint action is yet to be investigated. Examining these questions developmentally will be integral to understanding the development of intentional communication in infancy, and in identifying atypical trajectories (Gaffan et al., 2010).

Our use of naturalistic data is a limitation, as well as a strength, as we were unable to control for how much infants moved their attention between objects in the time before and after look onsets. This not only introduces artifact to the EEG signal, but also means that the extent to which oscillatory activity is influenced by object processing, differs between looks. However, we showed that the number of object and partner looks did not differ in the 2000ms before look onset for either comparison, and differed only in the first 1000ms after look onset, comparing infant- and adult-led looks to mutual attention (Figure S2, S8, S9). Our secondary analyses, excluding infant EEG activity where the infant was not continually focussed on one object or the partner before/after look onset, also showed similar differences in patterns of neural oscillatory activity between looks, reported in the main text (Figure S5, S11). Increased gaze shifts after infant-led looks to nonmutual attention did, however, mean that we were unable to compare infant-initiated looks to mutual and nonmutual attention in the time after look onset. Employing continuous methods of analysis to naturalistic data would overcome this issue (Jessen et al., 2021).

The ability to engage in reciprocally mediated joint attention towards the end of the first year is catalytic to developments in language and social cognition (Carpenter et al., 1998; Yu & Smith, 2012). The findings reported here suggest that at 10-12-months, infants are not yet predominantly proactive in creating and maintaining episodes of joint attention with their adult partner. They are, however, sensitive to whether their behaviour is contingently responded to, potentially forming the basis for the emergence of intentionally-mediated communication.

CHAPTER 4 - Endogenous oscillatory rhythms and interactive contingencies jointly influence infant attention during early infant-caregiver interaction

This chapter is a publication of a pre-print article, about to be submitted for publication, examining the mechanisms that structure infant attention in early interaction, assessing forwards-predictive and reactive changes in caregiver behaviour and infants' endogenous neural activity relative to the duration of infant attention episodes towards objects and their partner (Phillips et al., pre-print). Subheadings, figure placement, figure and table style, and citation style have been adapted to conform to the thesis format. The supplementary materials for this chapter are available in Appendix B.

Abstract

Almost all early cognitive development and learning take place in social contexts. At the moment, however, we know little about the neural and cognitive mechanisms that underpin infants' attention during free-flowing interactions. Recording EEG during naturalistic caregiver-infant interactions, we compare two different accounts. Attentional scaffolding perspectives emphasise the role of the caregiver in structuring the interaction, whilst active learning models focus on motivational factors, endogenous to the infant, that guide their attention. Our results show that intrinsic cognitive processes control infant attention: fluctuations in oscillatory neural activity, indexing endogenous cognitive processes, were a cause and a consequence of changes in infants' attentiveness. In comparison, infant attention was not forwards-predicted by caregiver gaze or vocal behaviours. Instead, caregivers rapidly

modulated their behaviours in response to changes in infant attention and cognitive engagement: greater reactive changes associated with longer infant attention. Our findings suggest attention emerges through interactive but asymmetric, infant-led processes that operate across the caregiver-child dyad.

4.1 Introduction

Almost all early cognitive development and learning takes place in social contexts (Trevarthen, 2001). We know that caregiver behaviours influence where, how, and for how long children allocate their attention in real-world settings (Yu & Smith, 2016), and that individual differences in how caregivers behave while interacting their child can predict later language learning and socio-cognitive development (Donnellan et al., 2020; Henning et al., 2005; Murray et al., 2016). But we currently understand little about how intrapersonal and bidirectional neural mechanisms influence how infants allocate their attention to learn from their environment during naturalistic, free-flowing interactions.

A number of different theoretical models try to explain how social partners influence infants' attention. The first, and probably the oldest, proposes that caregivers directly and didactically scaffold their infant's attention, for example, by building a structure of how they pay attention, and when, and encouraging the child to follow their attentional focus (a process sometimes known as 'attentional scaffolding' (Bornstein, 1985). This might take place through children copying where caregivers are paying attention, second by second, while they complete a shared task (Yu & Smith, 2016). Or, it might happen through adults organisedly and actively using ostensive signalling to guide infant attention (Csibra & Gergely, 2009). To do this, the adult partner might be using salient behaviours (e.g., eye gaze, high pitched speech, etc) to exogenously influence where children allocate their attention. In either of these cases, infant attention is *reactive* to changes in the behaviour of the caregiver (Sebanz & Knoblich, 2009; Yu & Smith, 2013).

Recent micro-behavioural analyses of caregiver and infant gaze behaviour during joint table-top interactions support this perspective, to some extent. Multimodal behavioural inputs by the caregiver are known to support episodes of sustained attention towards objects: for example, infant attention durations lasting over 3 seconds are directly predicted by the amount and timing of caregiver speech and touch to objects (Suarez-Rivera et al., 2019). More indirectly, other research has shown that infant attention is more fast-changing in joint compared to solo play, despite infant attention durations being, overall, longer in joint play (Wass et al., 2018a) - suggesting that endogenous cognitive processes such as attentional inertia (the finding that, the longer a look lasts, the less likely it is to end (Richards & Anderson, 2004)) have less of an influence on infant attention in social contexts. Further research suggests that, rather than following the focus of the adults' gaze, infants most often co-ordinate their attention with the adult through attending towards their partners' object manipulations, which corresponds to the idea that adults use exogenous attention capture to drive infant attention (Yu & Smith, 2013). Other salient behaviours might also be important, but are under-investigated. For example, IDS is known to contain more variability in amplitude and pitch (Spinelli et al., 2017), which increases its auditory salience (Nencheva et al., 2019); but although it is known that children generally pay more attention to IDS (Cooper & Aslin, 1994), no previous research has examined whether caregivers use moment-by-moment variability in the salience of their voice to influence how children allocate attention.

Within this framework, it is possible that, rather than repeated and reactive contingent responsivity to isolated behaviours, temporal dependency between infant and caregiver attention is driven by infant behaviour becoming periodically coupled to the behavioural modulations of their partner (L. Meyer et al., 2019; Wass et al., 2022). Similar to inter-dyadic patterns of vocalisations in adults and marmoset monkeys (Takahashi et al., 2013; Wilson & Wilson, 2005), in early infant-caregiver interactions, vocal pauses in one partner's vocalisations can be predicted from those of the other (Gratier et al., 2015; Jaffe et al., 2001), and, during face-to-face interactions at the end of the first year, caregiver-infant facial affects become temporally aligned (Feldman et al., 1999). Oscillatory entrainment, that is, consistent temporal alignment between fluctuations in caregiver and infant behaviour, could be particularly important in ensuring that salient sensory and information-rich inputs by the caregiver occur at moments infants are most receptive to receiving information (Wass et al., 2022).

An alternative interpretation of these micro-behavioural findings, however, is that, rather than structuring infant behaviour through leading infant attention, caregivers instead scaffold how infants pay attention by following and responding to re-orientations in their infant's attention. This second model suggests that, rather than considering unidirectional caregiver->child influences we should instead be considering bidirectional child<->caregiver influences. In following the focus of their infants' attention at moments that they reorient towards a new object, the caregiver 'catches' and extends infant attention with reactive and dynamic change in their salient ostensive behaviours, to which infants are responsive (Yu & Smith, 2016). The contingent adaptation of the caregiver to modulations in infant attention serves to maintain and extend infant attention, and provides inputs at points where infants anticipate to receive new information (Yu & Smith, 2012). Indeed, from early infancy, caregivers are contingently responsive to modulations in their infant's behaviour. From 2-3 months, caregivers respond differentially to distinct facial affects produced by the infant (Murray, et al., 2016), modulate their vocal feedback to infant babbling (Albert et al., 2018; Goldstein & Schwade, 2008; Yoo et al., 2018); and, towards the end of the first year, provide more labelling responses relative to infant's pointing than to their object-directed vocalisations (Wu & Gros-Louis, 2015).

According to the first model, then, caregivers drive and actively control infants' attention during joint interaction. According to the alternative model, caregivers influence infants' attention by reactively and contingently responding to the infant's attention shifts. But according to the latter model, what drives how infants initially allocate their attention in the first place? In caregivers, the timing of attention shifts can be partially described using an oscillatory, or periodic, structure, reflecting rhythmic attention reorientations that possibly correspond to fluctuations in the central nervous system (Nuthmann et al., 2010; Nuthmann & Henderson, 2010; Nuthmann & Matthias, 2014). Research with infants has also suggested that, even in the first 3 months of life, infants' attention shifting is not purely random (Robertson, 2014), and, in free-viewing paradigms, the allocation of infant attention has been shown to become more periodic and less stochastic over the course of the first year (Renswoude et al., 2019; Stallworthy et al., 2020). Regulatory mechanisms endogenous to the infant could therefore be one mechanism that influences when infants reorient their attention during real-world naturalistic interactions.

By the end of the first year, however, as well as periodic attention reorientations, fluctuations in top-down attentional control processes, thought to be driven by the executive attention system, begin to influence where and when infants shift their attention. For example, research has shown that infants routinely deploy active and effortful information-sampling strategies to maximise their opportunities for learning (Gottlieb et al., 2013; Goupil & Proust, 2023; Kidd et al., 2012; Oudeyer & Smith, 2016; Poli et al., 2020). For example, infants aged 8-9 months optimise information gain by directing their attention towards stimuli that are neither too complex, nor too predictable (Kidd et al., 2012, 2014) and disengage from stimuli that are less informative compared to past observations (Poli et al., 2020). Corresponding to developments in intentionally-mediated forms of joint communication (Tomasello et al., 2007), infants are also thought to begin to use active strategies to directly elicit information from a social partner about their environment. For example, infants aged 12-14 months point in an interrogative manner (Begus & Southgate, 2012; Kovács et al., 2014), and look towards their caregiver to ask for help when uncertain (Bazhydai et al., 2020; Goupil et al., 2016).

These approaches suggest that infants' endogenous engagement or interest forward-predicts their attention patterns. In addition, though, there is an alternative, complementary possibility. Infants' attention shifts may initially happen as reactive or spontaneous behaviours (driven by external attention capture, or, as discussed above, internal regulatory functions; Robertson, 2014; Stephens & Charnov, 1982) that are not forward-predicted by fluctuations in infants' endogenous engagement or interest; processes *after* the attention shift (determined by what information is present at the attended-to location) may drive increases in infants' endogenous engagement or interest which prolong that attention episode. (This distinction is similar to that we discussed above, about whether caregiver behaviours forwards-predict infant attention, or whether caregivers influence infants by reactively responding to their attention shifts, but operates at the individual level.) Consistent with this possibility, dynamic, generative models based on this framework can accurately predict attention patterns at least in younger infants (Robertson, 2014; Robertson et al., 2004).

Dynamic, amplificatory processes that take place after an attention shift can also explain patterns of attention inertia observed in naturalistic settings (Richards & Anderson, 2004).

To examine how fluctuations in endogenous engagement or interest drive and/or maintain infant attention during naturalistic interactions, we can measure theta activity (3-6Hz), which is an oscillatory rhythm associated with intrinsically guided cognitive process in early infancy (Begus & Bonawitz, 2020). In particular, EEG activity in the theta range has been found to increase over fronto-central electrodes during episodes of endogenously controlled attention. For example, theta activity over fronto-central electrodes increases where infants anticipate the next actions of an experimenter, and theta activity occurring in the time before infants look towards an object has been found to predict the length of time infants pay attention to that object during solitary play (Orekhova et al., 2006; Wass et al., 2018b). Recent work has also shown dynamic fluctuations in theta activity over the course of sustained attention episodes: Xie et al., 2018 found that, whilst 10-12 month-old infants viewed cartoon videos, theta activity increased during heart-rate defined periods of attentional engagement (see also Jones et al., 2020).

In summary, therefore, research has examined two separate influences that could support how infants pay attention in social settings. The first type of influence is endogenous engagement or interest. The second is caregivers' exogenous behaviour. But for both of them, it is unclear whether the influences are forwards-predictive or reactive. Does infants' endogenous attention engagement forwards-predict attention, or do fluctuations in engagement that take place after an attention shift predict how long that episode lasts? And do caregivers drive infant attention using salience cues, or do they reactively change their behaviours in response to infant behaviours?

Here, recording EEG from infants during naturalistic interactions with their caregiver, we examined the (inter)-dependent influences of infants' endogenous oscillatory neural activity, and inter-dyadic behavioural contingencies in organising infant attention. First, we examined processes endogenous to the infant that determine the timing of their attention during the interaction (section 4.2). Second, we examine caregiver behaviours (part 4.3).

First, in section 4.1, we test whether oscillatory structures can be derived from the patterns of infant and caregiver looking behaviour at an individual level, by computing the partial auto-correlation function (PACF) for caregiver and infant attention durations. We then test whether infant and caregiver behaviours act as coupled oscillators, by examining the time-course of the cross-correlation function between infant and adult gaze (Takahashi et al., 2013). If true, this would point to the existence of mechanisms of influence between infant and adult gaze that our other analyses, examining forwards- and backwards-predictive relationships, would be unable to detect.

In section 4.2 we then assess whether infants' endogenous cognitive processing forward-predicts infant attention, by using a cross-correlations to estimate the forwards- and backwards-predictive associations between infant theta activity and look durations. In addition, we further examined reactive changes in infant endogenous oscillatory neural activity that take place after the onset of an attention episode towards objects. To do so, we used two analyses: first, using linear-mixed effects models, we examined the direct temporal associations between the infant attention durations and the average levels of infant theta activity during that look. Next, we examined how theta activity changes dynamically across the course of individual looks.

In section 4.3 we examined the (inter)-dependent relationships between caregiver behaviours and infant attention. We examined two aspects of caregiver behaviour in particular. First (section 4.3.1, and 4.3.2), we examined caregiver gaze behaviour, using cross-correlations to test whether increases in caregiver attention towards objects forwards- or backwards-predicted changes in infant attention. In order to test whether any association between infant attention and caregiver behaviour was independent of the relationship between infant attention and their endogenous oscillatory neural activity, we also conducted cross-correlations to examine the associations between caregiver attention and infant theta activity. And we used the same two analyses as used in section 3.2 to examine how caregiver gaze behaviour changes reactively following the onset of an infant attention episode.

Second (section 3.3.3 and 3.3.4) we examined saliency in the caregiver's speech signal by computing the rate of change in the fundamental frequency of their voice (Cooper & Aslin, 1994). For this, we used the same analysis approach. First, we conducted cross-correlations to examine whether changes in caregiver vocal behaviour forwards- or backwards-predict changes in infant attention. Second, we examined how caregiver vocal behaviour changes reactively following the onset of an infant attention episode. Allostatic attentional-structuring models predict reactive change in caregiver behaviour at the onset of infant attention, and over the duration of the look, that associate with the length of infant looking.

4.2 Methods

4.2.1 Participants

Ninety-four caregiver-infant dyads took part in this study. The final overall sample with usable, coded, gaze data was 66 (17 infants were excluded due to recording error or equipment failure, 4 infants were excluded for fussiness and 6 infants were excluded due to poor quality EEG data, and limited coding resources). Of the infants with usable gaze data, 51 had additional vocal data (15 excluded due to recording error/equipment failure: overall exclusion 46%). Of those with gaze data, 60 infants had usable EEG data (a further 6 excluded due to noisy EEG data (see artifact rejection section below); 36% data loss from overall sample. All usable data sets available for each separate analysis were used in the results reported below (e.g. infants with gaze and EEG data but no vocal data are included in analyses exploring the relationship between infant EEG and gaze). In the final sample, the mean age of infants was 11.18 months (SD=1.27; 36 females, 30 males). All caregivers were female. Participants were recruited through baby groups and Childrens' Centers in the Boroughs of Newham and Tower Hamlets, as well as through online platforms such as Facebook, Twitter and Instagram. Written informed consent was obtained from all participants before taking part in the study, and consent to publish was obtained for all identifiable images used. All experimental procedures were reviewed and approved by the University of East London Ethics Committee.

4.2.2 Experimental set-up

Parents and infants were seated facing each other on opposite sides of a 65cm wide table. Infants were seated in a high-chair, within easy reach of the toys (see Figure 4.1b). The shared toy play comprised two sections, with a different set of toys in each section, each lasting ~5 minutes each. Two different sets of three small, age-appropriate toys were used in each section; this number was chosen to encourage caregiver and infant attention to move between the objects, whilst leaving the table uncluttered enough for caregiver and infant gaze behaviour to be accurately recorded (cf. Yu & Smith, 2013).

At the beginning of the play session, a researcher placed the toys on the table, in the same order for each participant, and asked the caregiver to play with their infant just as they would at home. Both researchers stayed behind a screen out of view of caregiver and infant, except for the short break between play sessions. The mean length of joint toy play recorded for play section 1 was 297.28s (SD=54.93) and 323.18s (SD=83.45) for play section 2.

4.2.3. Equipment

EEG signals were recorded using a 32-channel BioSemi gel-based ActiveTwo system with a sampling rate of 512Hz with no online filtering using Actiview Software. The interaction was filmed using three Canon LEGRIA HF R806 camcorders recording at 50 fps. Parent and infant vocalisations were also recorded throughout the play session, using a ZOOM H4n Pro Handy Recorder and Sennheiner EW 112P G4-R receiver.

Two cameras faced the infant: one placed on the left of the caregiver, and one on the right (see Figure 4.1b). Cameras were placed so that the infant's gaze and the three objects placed on the table were clearly visible, as well as a side-view of the caregiver's torso and head. One camera faced the caregiver, positioned just behind the left or right side of the infant's high-chair (counter-balanced across participants). One microphone was attached to the caregiver's clothing and the other to the infant's high-chair.

Caregiver and infant cameras were synchronised to the EEG via radio frequency (RF) receiver LED boxes attached to each camera. The RF boxes simultaneously received trigger signals from a single source (computer running MATLAB) at the beginning of each play section, and concurrently emitted light impulses, visible in each camera. Microphone data was synchronised with the infants' video stream via a xylophone tone recorded in the infant camera and both microphones, which was hand identified in the recordings by trained coders. All systems were extensively tested and found to be free of latency and drift between EEG, camera and microphone to an accuracy of +/- 20 ms.

4.2.4 Video coding

The visual attention of caregiver and infant was manually coded using custom-built MATLAB scripts that provided a zoomed-in image of parent and infant faces (see Figure 4.1b). Coders indicated the start frame (i.e. to the closest 20ms, at 50fps) that caregiver or infant looked to one of the three objects, to their partner, or looked away from the objects or their partner (i.e. became inattentive). Partner attention epsiodes included all looks to the partner's face; looks to any other parts of the body or the cap were coded as inattentive. Periods where the researcher was within camera frame were marked as uncodable, as well as instances where the caregiver or infant gaze was blocked or obscured by an object, or their eyes were outside the camera frame. Video coding was completed by two coders, who were trained by the first author. Inter-rater reliability analysis on 10% of coded interactions (conducted on either play section 1 or play section 2), dividing data into 20ms bins, indicated strong reliability between coders (kappa=0.9 for caregiver coding and kappa=0.8 for infant coding).

4.2.5 Vocalisation coding

The onset and offset times of caregiver and infant vocalisations were identified using an automatic detector. The algorithm detected voiced segments and compared the volume and fundamental frequency detected in each recorded channel to infer the probable speaker (mother vs. infant). Identification of the onset and offset times of the detector then underwent a secondary analysis by trained coders, who identified misidentification of utterances by the automatic decoder, as well as classifying the speaker for each vocalisation. As the decoder did not accurately identify onset and offset times of caregiver and infant during co-vocalisations, and, as these vocalisations could not be included in analyses of the spectral properties of caregiver vocalisations, these were excluded from all analyses. The mean percentage of caregiver vocalisations that were co-vocalisations was less than 20%: 19.43 (SD=12.36; a box plot across all participants is presented in Figure S1). In a previous analysis conducted on a sub-sample of the data (Phillips et al., in press), we have shown that there is no significant change in infant vocalisations, relative to the onset of infant attention episodes, and their vocal beahviour did not distinguish between moments that they either led or followed their partners' attention during the interaction. It is therefore unlikely that inclusion of co-vocalisations in the current analyses would affect the main findings, time-locking caregiver vocalisations to infant attention.

4.2.6 Infant EEG artifact rejection and pre-processing

A fully automatic artifact rejection procedure including ICA was adopted, following procedures from commonly used toolboxes for EEG pre-processing in adults (Bigdely-Shamlo et al., 2015; Mullen, 2012) and infants (Debnath et al., 2020; Gabard-Durnam et al., 2018), and optimised and tested for use with our naturalistic infant EEG data (Georgieva et al., 2020; Marriott Haresign et al., 2021). This was composed of the following steps: first, EEG data were high-pass filtered at 1Hz (FIR filter with a Hamming window applied: order 3381 and 0.25/ 25% transition slope, passband edge of 1Hz and a cut-off frequency at -6dB of 0.75Hz). Although there is debate over the appropriateness of high pass filters when measuring ERPs (see (Widmann & Schröger, 2012), previous work suggests that this approach obtains the best possible ICA decomposition with our data (Dimigen, 2020; Marriott Haresign et al., 2021). Second, line noise was eliminated using the EEGLAB (Bigdely-Shamlo et al., 2015) function *clean_line.m* (Mullen, 2012).

Third, the data were referenced to a robust average reference (Bigdely-Shamlo et al., 2015). The robust reference was obtained by rejecting channels using the EEGLAB *clean_channels.m* function with the default settings and averaging the remaining channels. Fourth, noisy channels were rejected, using the EEGLAB function *clean_channels.m.* The function input parameters 'correlation threshold' and 'noise threshold' (inputs one and two) were set at 0.7 and 3 respectively; all other input parameters were set at their default values. Fifth, the channels identified in the previous stage were interpolated back, using the EEGLAB function eeg_interp.m. Interpolation is commonly carried out either before or after ICA cleaning but, in general, has been shown to make little difference to the overall decomposition (Delorme & Makeig, 2004). Infants with over 21% (7) electrodes interpolated were excluded from analysis. After exclusion, the mean number of electrodes interpolated for infants was 3.37 (*SD*=2.27) for play section 1, and 3 (*SD*=2.16) for play section 2.

Sixth, the data were low-pass filtered at 20Hz, again using an FIR filter with a Hamming window applied identically to the high-pass filter. Seventh, continuous data were automatically rejected in a sliding 1s epoch based on the percentage of channels (set here at 70% of channels) that exceed 5 standard deviations of the mean channel EEG power. For example, if more than 70% of channels in each 1-sec epoch exceed 5 times the standard deviation of the mean power for all channels then this epoch is marked for rejection. This step was applied very coarsely to remove only the very worst sections of data (where almost all channels were affected), which can arise during times when infants fuss or pull the caps. This step was applied at this point in the pipeline so that these sections of data were not inputted into the ICA. The mean percentage of data removed in play section 1 was 11.30 (SD=14.97), and 6.57(SD=6.57) for play section 2.

Data collected from the entire course of the play session (including play section 1 and play section two, as well as two further five minute interactions) were then concatenated and ICAs were computed on the continuous data using the EEGLAB function runica.m. The mean percentage of ICA components rejected was 52.03% (SD=18.38). After ICA rejection, data from each play section were re-split. For representative examples of artifactual ICA components identified in our naturalistic EEG data, see Marriott-Haresign et al. (2021).

4.2.7 Pre-processing of continuous variables

Prior to conducting our main analyses, all primary variables of interest were converted into continuous variables, in order to perform time-lagged and event-locked methods of analysis, relative to infant attention (see Figure 4.1d). All continuous variables were downsampled to match the sampling rate of the video cameras (50Hz).

4.2.7.1 Infant theta activity

First, missing data points were excluded from the continuous time-series. Where one or more of the fronto-central electrodes of an individual infant exceeded 100uV for more than 15% of the interaction, the infant's continuous theta time-series was excluded from analyses. Next, time-frequency decomposition was conducted via continuous morlet wavelet analysis to extract EEG activity occurring at frequencies ranging from 1-16Hz. Specifically, the EEG signal at each channel was convolved with Gaussian-windowed complex sine-waves, ranging from 1-16Hz, in linearly spaced intervals. The width of the guassian was set to 7 cycles. Power was subsequently extracted as the absolute value squared, resulting from the complex signal. After decomposition, to get rid of edge artifacts caused by convolution, the first and last 500ms of the time series were treated as missing data points. Missing data points were then re-inserted into the continuous variable as blank values, and the 500ms before and after these chunks of data also excluded. For each time point, for each frequency, power was expressed as relative power (i.e. the total power at that frequency, divided by the total power over all frequencies). EEG activity was then averaged over frequencies ranging from 3-6Hz, and averaged over fronto-central electrodes (see Figure 4.1).

This continuous, one-dimensional variable was then downsampled from 512 to 50Hz by taking the median theta activity for every 10 samples of data, and, in each second, taking an extra 1 sample for 3 time points and an extra 2 samples for 1 time point. The spacing of these added samples was shuffled for each second of data.

4.2.7.2 Attention durations

An attention episode was defined as a discrete period of attention towards one of the play objects on the table, or to the partner. The end of each attention episode was defined as the moment where the participant first looked away from the target towards another object, towards the partner, or towards another location that was not either the object or the partner (coded as non-target attention). See Figure 4.1d for an example. Parts of the caregiver/infant gaze coded as uncodable were treated as missing data points, as well as the looks occurring in the time just before and after (in order to account for the fact that we do not know how long these looks last).

4.2.7.3 Binary attention variable

For the analyses in section 4.3.1, which examine the temporal oscillatory patterns of attention shifts, we recoded each look alternatively as a 0 or 1 from the first look of the interaction to the last (see Figure 4.1d). These analyses examine therefore the temporal inter-dependencies between attention durations (within an individual and across the dyad), irrespective of where the attention is directed.

4.2.7.4 Continuous attention duration variable

For the analyses in section 4.3.2 and section 4.3.3, which examine the associations between attention durations and other measures, we recoded each look based on the duration in seconds of that look. The durations of each look were then used to produce a continuous look duration variable, irrespective of whether that look was towards the object, partner, or non-target (see Figure 4.1d). These analyses examine therefore the associations between the durations of attention episodes and, respectively, endogenous infant neural activity (section 4.3.2) and caregiver behaviour (4.3.3).

4.2.7.5 Caregiver vocalisation durations

The length of each caregiver vocalisation was computed in seconds and inserted into the video-frame time series for the duration of that vocalisation. Periods where the caregiver was not vocalising (i.e. vocal pauses) were set to missing data points. Times where co-vocalisations occurred were also set to missing data points.

4.2.7.6 Rate of change in the fundamental frequency (F0) of the caregiver's voice

The fundamental frequency of the caregiver's voice was extracted using Praat (Borsema & Weenik, 2019)with floor and ceiling parameters set between 75-600Hz. Caregiver fundamental frequency was placed into the continuous variable only where the coder had identified that section of speech as the caregiver speaking, so that infant vocalisations were not included in the analysis. Due to the caregiver being within variable distance of their microphones, some clipping was identified in a sample of the microphone recordings. A stringent clipping identification algorithm was used (see Figure S2) to remove parts of the microphone data where clipping occurred (Hansen et al., 2021; see SM, section 1.1.1). Vocalisations where any clipping was identified were set to missing data points. Interactions with more than 30% missing vocalisations were excluded from the analyses. Statistics on the number of vocalisations excluded on this basis is presented in Figure S2. Co-vocalisations were again set to missing data points.

Next, unvoiced sounds and periods between vocalisations were interpolated, using MATLAB's interp1 function. To reduce the likelihood of background noise (e.g. toy clacks) affecting the fundamental frequency, the interpolated variable was low-pass filtered at 20Hz using a 9th order butterworth filter. The rate of change in the caregiver's fundamental frequency was computed by taking the sum of the derivative in 1000ms intervals. The start and end points of each interval were then converted to time in camera frames, and the rate of change values inserted for the 50 corresponding frames.

4.2.7.7 Caregiver amplitude modulations

Amplitude modulations in the caregivers' speech were extracted using the NSL toolbox (Chi et al., 2005). First, the speech signal was downsampled to 16kHz. The 128-channel auditory spectrogram, with centre frequencies ranging from 180-7246Hz was then computed (frame length=5ms, time constant=8ms, no nonlinear filtering), and the band-specific envelopes summed across frequencies to obtain the broadband envelope of the speech signal. The amplitude envelope was inserted into the continuous variable only where the coder had identified the caregiver as vocalising: all vocal pauses were treated as missing data points. Clipped vocalisations were also identified using the same method described above, and set as missing values. Finally, the continuous amplitude variable was synchronised to the video frames.

4.2.8 Analysis procedures

4.2.8.1 Partial autocorrelation function

Computation

The partial auto-correlation function (PACF) of the caregiver and infant gaze time series was computed separately, over a range of time intervals, from 100-1000ms. First, the gaze time series was converted to a continuous binary variable, with either a 1 or 0 inserted into the time series for the duration of each attention episode, alternated for each consecutive look. The PACF was then computed by fitting an ordinary least squares regression model, at time-lags ranging from 0 to 10s, in 100ms intervals, controlling for all previous time-lags on each iteration. This analysis was repeated at intervals of 200, 500 and 1000ms.

Shuffled time series

To investigate whether the shape of the PACF reflected the temporal distribution of infant/caregiver attention episodes or more simply the frequency distribution (i.e. infant/caregiver attention episodes frequently last a similar length; Brookshire, 2022), we conducted a permutation procedure, whereby, for each infant, their attention duration time series was shuffled randomly in time to produce a binary gaze time series of shuffled attention durations. The PACF was then computed for this time series in exactly the same way described above. This procedure was subsequently repeated 100 times for each participant, before averaging over all permutations and participants.

Poisson baselines

Poisson baselines were created by computing time series of the Poisson point process with the length of look durations matching the average length of look durations in the actual data (Takahashi et al., 2013). This variable was then converted to a binary look duration variable (see above) for the binary cross-correlation analysis.

4.2.8.2 Cross-correlation analyses

Cross-correlations were computed between infant attention durations, the continuous caregiver variables and infant theta activity. All analyses for the continuous caregiver variables were subsequently repeated relative to infant theta activity.

Computation

First, the time series of each variable were log transformed, and outliers falling 2 inter-quartile ranges above the upper quartile and two inter-quartile ranges below the lower quartile removed. A detrend was then applied to each variable; linear and quadratic bivariate polynomials were fit to each transformed time-series, and the residuals of the model of best fit computed. The cross-correlation between the two variables was then computed at lags -30 to +30s in 500ms intervals. The zero-lagged cross-correlation was first computed between the two variables using a Pearson correlation. The caregiver's time series (or infant theta activity where this was computed relative to infant attention durations) was then moved backwards in time (to compute negative lag correlations), or forwards in time (to compute positive lag correlations), and the Pearson correlation computed between the two time series at each 500ms interval. In this way, we estimated how the association between the two variables changed with increasing time lags. The cross-correlations at each time-lag were then averaged over the two interactions for each participant, and then averaged over all participants.

Significance testing

A cluster-based permutation approach was used to investigate whether the time-lagged cross correlation differed significantly from chance over any time-period. This approach controls for family-wise error rate using a non-parametric Monte Carlo method (Maris & Oostenveld, 2007). To create a random permutation distribution at each time-lag, each participant was randomly paired with another participant, through a process of derangement, and the cross-correlation between the caregiver and infant variables computed, and averaged over participants in exactly the same way described above. This procedure was then repeated 1000 times, resulting in a random permutation distribution at each time lag. Next, the cross-correlation at each time lag in the observed data was compared with the permutation distribution at that time lag, and values falling above the 97.5th centile and below the 2.5th centile were accepted as significant (corresponding to a significance level of 0.05). To examine the likelihood of clusters of significant time points in the observed data occurring by chance, a cluster-threshold was computed using a leave-on-out procedure on the permutation data. On each iteration, one permutation was compared with the 999 other permutations, and significant time-points identified using the same method described above. The largest cluster found on each iteration was identified to create a random permutation distribution of cluster sizes. The clusters identified in the observed data were then compared with this permutation distribution of maximum cluster sizes, and clusters falling above the 95th centile were considered significant (corresponding to a significance level of 0.05).

4.2.8.3 Linear mixed effect models

Linear mixed effect models were used to investigate the relationship between caregiver behaviour, infant theta activity and infant attention duration. First, for each participant, for each attention duration, the continuous caregiver behavioural variable was averaged over the length of the infant episode, to obtain one value for caregiver behaviour per infant attention duration. Next, each variable was log-transformed, and outliers 2 inter-quartile ranges above the upper quartile and two inter-quartile ranges below the lower quartile removed. Finally, linear mixed effects models were fitted, with caregiver behaviour, or infant theta activity as the fixed effect, and infant attention durations as the response variable, with a random effect of participant.

4.2.8.4 Attention onset event-related analysis

Computation

Before event-locking the continuous variables to infant attention, the continuous variable was log-transformed, and outliers removed, applying a similar procedure to that described above. First, the frame of the onset of each infant object look, as well as the duration of that look was extracted from the infant gaze time series. Next, for each continuous variable, the frames occurring five seconds before and five seconds after the onset of each infant look were extracted from the caregiver time series. Given the fact that we were interested in how caregiver behaviour changed around the onset of an attention episode, where the infant shifted gaze again in the 5 second time-period after attention onset, the values in the continuous caregiver variable were set to missing data points. The continuous frames occurring before and after each look were then averaged over looks, for each interaction, resulting in an averaged continuous variable along the time dimension. These values were then averaged over interactions for each participant, before averaging over all participants.

In order to explore the possibility that the length of the infant attention episode might affect how the caregiver's behaviour changed around the onset of that episode, exactly the same analysis was repeated on attention durations of different lengths, in 5 log-spaced intervals, ranging from 0 to the longest attention episode identified across the datasets (118s).

Significance testing

Significance testing followed exactly the same procedures outlined in the cross-correlation analysis section.

4.2.8.5 Modulation during attention episodes

For this analysis, all continuous data variables (caregiver behaviour / infant theta activity) were log transformed and outliers removed (see above). Then, for each infant object look, the continuous caregiver behaviour / infant theta activity was extracted over the length of that attention episode, and divided into 3 equal-spaced chunks. The continuous data variable occurring in the first half of each chunk was then averaged for each attention episode, before being averaged over all episodes for that interaction. Averaged chunks from play section 1 and play section 2 were then averaged together for each participant, and the mean over all participants, for each chunk, computed. A series of Wilcoxon Signed ranks tests assessed whether the chunks differed to each other, compared to that which would be expected by chance. The Benjamini-Hochberg false discovery rate procedure was applied to correct for multiple comparisons ($p< 0.05$; Benjamini Yoav & Hochberg Yosef, 1995).

Similar to the ERP analysis, infant object look durations were divided into 5 log-spaced bins to assess whether modulations in infant endogenous cognitive processing or caregiver behavior differed for episodes lasting different lengths: exactly the same procedure was repeated for each duration bin.

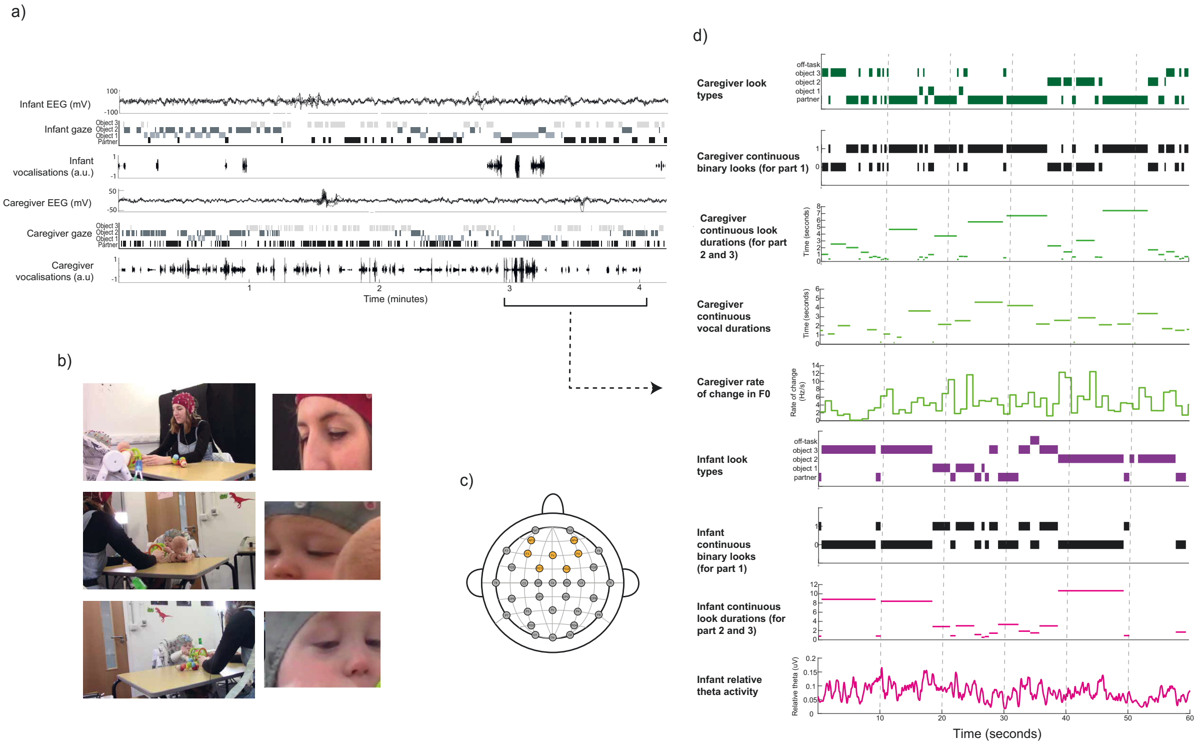

Figure 4. 1 Experimental set-up and example of continuous variables. a) Raw data sample, showing (from top) infant EEG over fronto-central electrodes, after pre-processing, infant gaze behaviour, infant vocalisations, adult EEG over fronto-central electrodes, adult gaze behaviour, adult vocalisations. b) Example camera angles for caregiver and infant (right and left), as well as zoomed-in images of caregiver and infant faces, used for coding. c) Topographical map showing electrode locations on the bio-semi 32-cap; fronto-central electrodes included in the infant time-frequency analysis are highlighted in orange (AF3, AF4, FC1, FC2, F3, F4, Fz). d) Continuous behaviour and EEG variables extracted from the caregiver and infant time-series, showing (from top), caregiver looks to objects, the partner, and off-task looks, caregiver binary attention durations, caregiver continuous attention durations, caregiver vocalisation durations, rate of change in caregiver F0, infant looks to

objects, their partner, and off-task looks, infant binary attention durations, infant continuous look durations, infant relative theta activity.

4.3. Results

The results section is divided into three sections. In section 4.3.1, we first conduct descriptive statistics of infant attention durations, and test for oscillatory structures in caregiver and infant attention. Then, in section 4.3.2, we examine whether endogenous infant neural activity forwards-predicts fluctuations in infant attention, and/or reactively changes in the time after the onset of an attention episode. In section 4.3.3, we assess whether modulations in caregiver gaze and vocal behaviour forwards-predict fluctuations in infant attention, and/or reactively change in the time after infants shift their attention.

4.3.1 Oscillatory structures in caregiver and infant attention

First, as descriptive statistics, we report on the frequency distribution of caregiver and infant attention durations towards objects, the partner, and periods of inattention, dividing attention durations into 100ms bins. Histograms showing the distribution of caregiver and infant attention durations towards objects, partners, and non-targets are displayed in Figure 4.2a. In both distributions the mode is greater than the minimum value, consistent with previous observations that attention shifting is periodic (Saez de Urabain et al., 2017). The caregiver's distribution is also more left-skewed compared to the infants' distribution, reflecting the shorter and more frequent attention durations by the caregiver (c.f Wass et al., 2018a; Yu & Smith, 2012). Finally, consistent with previous reports (Yu & Smith, 2013), caregivers tended to look towards their partner more frequently than infants, with infants attending most frequently to the objects (Fig 2a).

Next, to investigate whether there was an oscillatory component in the caregiver and infant gaze time series, we computed the PACF of a binary attention variable separately for caregiver and infant using 100, 200, 500 and 1000ms lags (see Methods and Fig 1a for more detail). In order to explore whether the PACF reflected the temporal interdependencies between infant/caregiver attention episodes (i.e. how likely an attention episode of a given length was to be followed by another of a similar length), or, more simply, the overall distributions of attention episodes (i.e. how common attention episodes of a given length are overall), the PACF was repeated after shuffling the infant and caregiver attention durations in time (see Methods for permutation procedure).

Figure 2b shows that for the 100ms, 200ms time bins (infant and caregiver) and 500ms time bin (infant only), the lag 1 terms are negative, indicating that attention at time t is negatively predictive of attention at time t+x where x is a short time interval. This pattern is also observed in the baseline data (in which looks have been randomly shuffled in time). It reflects therefore, the overall pattern already shown in the histograms in Figure 4.2a, that short looks (e.g. 100-200ms) are less frequent than longer looks (e.g. 500ms). It can also be seen that, at higher time lags, the observed PACF values are above the baseline rate. This indicates temporal interdependencies between look durations (i.e. that an attention episode of a given length is likely to be followed by another of a similar length), which are not present when the look durations are randomly shuffled to generate the baseline data.

Finally, we replicated previous analyses (Takahashi et al., 2013) to explore whether the inter-dyadic dynamics of caregiver and infant looking behaviour could be modelled as coupled oscillators. We computed the same binary attention variables for the infant as for the caregiver (see Figure 4.1d), and calculated the cross-correlation function between the infant and caregiver binary attention variables. In this way, we examined whether caregiver and infant attention changes at consistent temporal latencies, as would be the case if they were acting as entrained oscillators (Wass et al., 2022). If this is true, then the cross-correlation function should display significant peaks at regular intervals, reflecting these consistent latencies between attention shifts (Takahashi et al., 2013). In order to identify where peaks in the cross-correlation function exceeded chance, we computed Poisson point process timeseries with look duration lengths matching the average look duration in the actual data (see Methods for more details). Figure 4.2c shows the results of this. Cluster-based permutation analysis revealed no significant peaks in the cross-correlation function, compared to baselines created through poisson process.

In summary, oscillatory mechanisms appear to govern both caregiver and infant attention durations; with infant attention durations centring around 1-2s in length, and adults' around 200-500ms. The cross-correlation analysis, however, suggested that caregiver and infant attention shifts do not act as coupled oscillators across the dyad (Figure 4.2c).

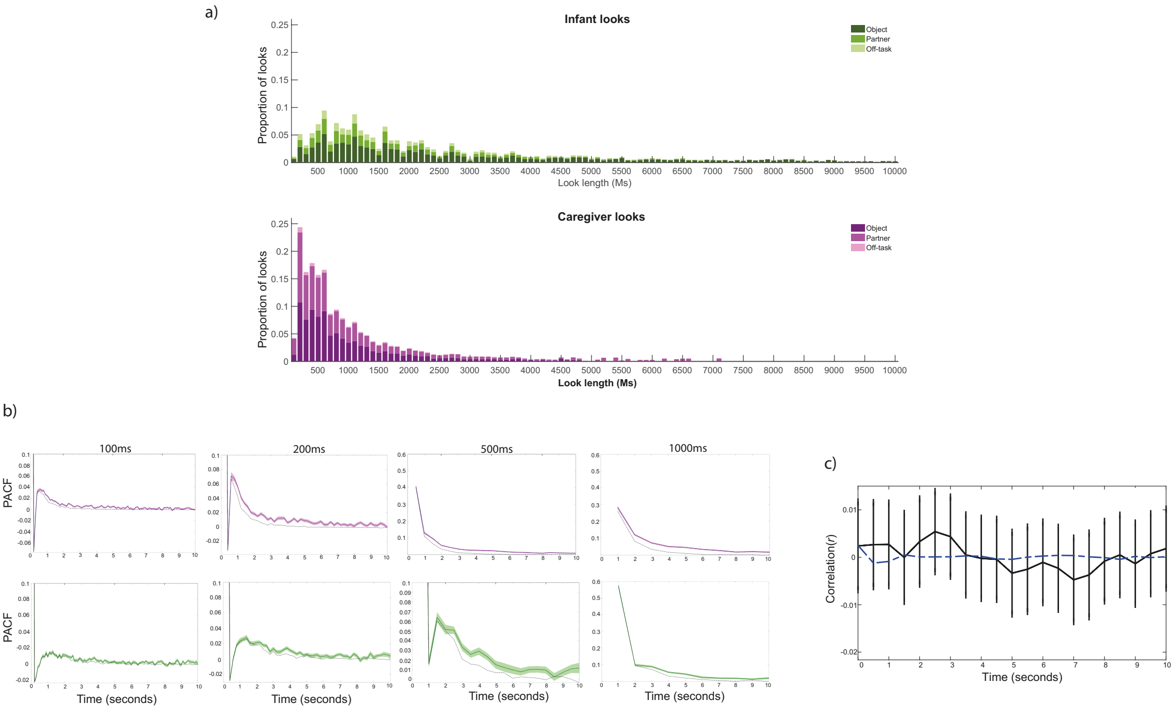

Figure 4. 2 Testing for oscillatory patterns of attention behaviour in infants and caregivers. a) Histogram of caregiver and infant attention episodes to objects, their partner and off-task episodes. Stacked bars show the number of episodes in each category for each 100ms bin for all episodes up to 10s in duration. b) PACF computed at different time lags for caregiver and infant gaze time series. Coloured lines show the PACF for infants (green) and caregivers (pink); shaded areas show the SEM. Dashed black lines show the PACF of shuffled attention duration data. c) Cross-correlation between caregiver and infant binary gaze variables. Black line shows the Spearman correlation coefficient at time-lags ranging from 0-10s; error bars indicated the SEM. Blue dashed line shows the permutation cross-correlation between two time series of poisson point process; one matching the average look rate of caregivers and the other, the average look rate of infants.

4.3.2 Does endogenous infant neural activity forwards-predict infant attention, or reactively change following the onset of a new infant attention episode?

In this section, we investigate the relationship between infant endogenous oscillatory neural activity and infant attention, considering both forwards-predictive relationships and reactive changes in infants' endogenous neural activity after the onsets of attention episodes.

4.3.2.1 Forwards predictive relationship between attention and infant theta activity

To examine whether infant endogenous neural activity significantly forwards-predicted infant attentiveness, we calculated a cross-correlation between the continuous infant attention duration time-series (see Figure 4.1d), including all infant attention episodes to objects, the partner and looks elsewhere, and infant theta activity. Figure 4.3a shows the results of the cross-correlation analysis. This analysis revealed a significant, positive association between the two variables at time-lags ranging from -2 to +6s ($p = 0.004$). This indicates that infant theta power significantly forwards-predicted infant attention durations at lags up to 2 seconds, as well as that infant attention durations significantly forwards-predicted infant theta at lags of up to 6 seconds.

4.3.2.2 Reactive change in infant theta activity following look onset

In addition to the cross-correlation, we also conducted two further analyses to investigate the relationship between theta activity and the duration of attention episodes. First, we calculated a linear mixed effects model to examine the relationship between the lengths of infant attention episode and average theta activity during that episode. This showed a significant, positive association between the two variables (β =0.33; $p<0.001$); scatter plot between the two variables is shown in Figure 4.3b. This indicates that higher average theta power across the attention episode associates with longer attention durations. Second, we explored dynamic change in theta activity relative to the onset of infant attention episodes towards objects. The modulation analysis (Figure 4. 3c) showed that there was little change in infant theta activity over the duration of infant attention episodes, for any duration time-bin: a series of Wilcoxon signed rank tests indicated decreases in infant theta activity for attention episodes lasting 1-3s, but this did not survive Benjamini-Hochberg correction.

In summary, there is a temporally specific relationship between infant attention durations and theta power, with attention durations forwards-predicting theta power more than vice versa (Figure 4.3a). Longer attention episodes are associated with increased average theta activity over the length of the episode (Figure 4.3b), but little dynamical change in theta activity is observed over the course of an attention episode.

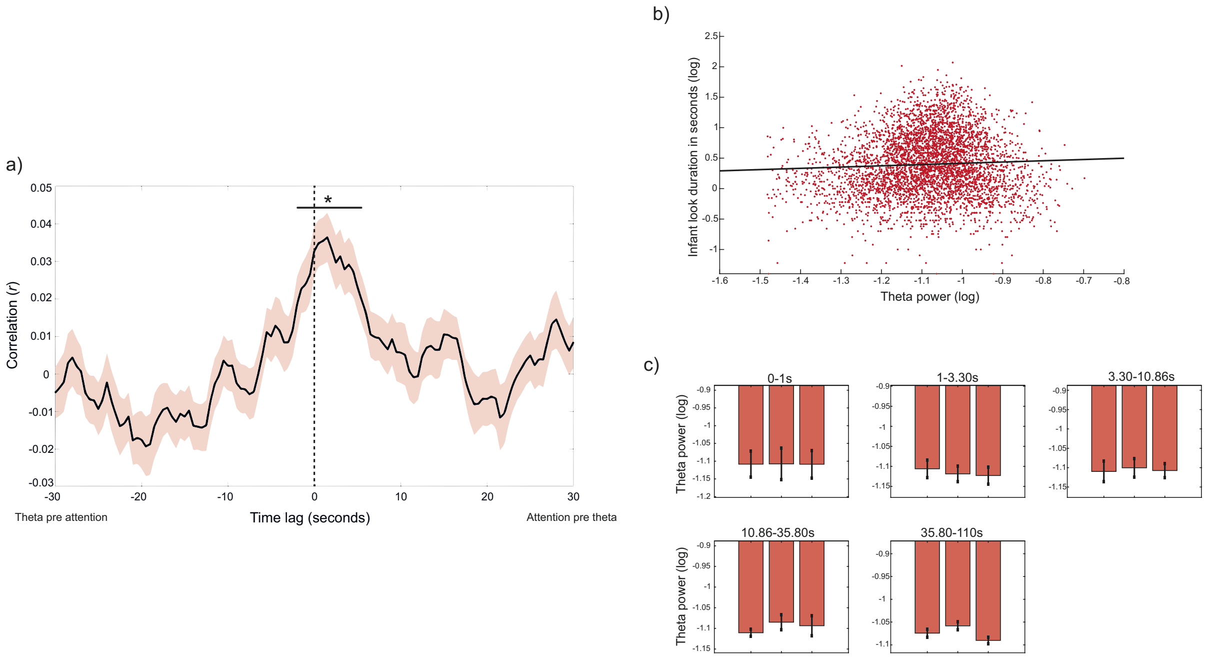

Figure 4. 3 Relationship between infant attention duration and infant theta activity. a) Cross correlation between infant theta activity and infant attention durations. Black line shows the Pearson correlation at each time lag, shaded areas indicate the SEM. Significant time lags identified by the cluster-based permutation analysis are indicated by black horizontal lines (*p <0.05). Cluster-based permutation analysis revealed a significant cluster of time points ranging from -2 to +6 seconds (p = 0.004). b) Linear mixed effects model, predicting infant object attention duration from infant theta activity. Coloured dots show each individual object attention duration; black line shows the linear line of best fit. The model reveals a significant positive association between infant theta activity and attention duration (β=0.33; p<0.001). c) Infant theta activity split into 3 attention chunks across the duration of attention episodes, binned according to episode length. Wilcoxon signed ranks tests explored significant differences between attention chunks, for each duration bin (*p < 0.05).

4.3.3 Do modulations in caregiver behaviours forwards-predict infant attention, or reactively change following the onset of a new infant attention episode?

In this section, we conduct similar analyses to those in section 4.3.2 to examine the relationship between modulations in caregiver behaviours and infant attention, considering, again, both forwards-predictive relationships and reactive changes in caregiver behaviour after the onsets of infant attention episodes. First, we examine the relationships in caregiver gaze behaviours (section 3.3.3.1 and 3.3.3.2); we then repeat these analyses relative to caregiver vocal behaviours, focusing on the rate of change of caregiver F0 as an index of auditory salience (section 3.3.3.4 and 3.3.3.5).

4.3.3.1 Forwards-predictive relationships between infant attention durations and adult attention durations

To examine whether caregiver attentiveness forwards-predicts infant attentiveness, we conducted cross-correlation analyses between the continuous infant and caregiver attention durations towards the objects. In order to test whether any association between infant attentiveness and caregiver attentiveness was independent of the relationship between infant attentiveness and their endogenous oscillatory neural activity shown in Figure 4.3a, we also repeated these analyses relative to infant theta activity. Results are reported in Figure 4.4. The cross-correlation between caregiver and infant attention durations appears to peak after lag zero, but cluster-based permutation analysis revealed no significant clusters of time points, though one cluster verged on significance (p=0.10). The cross-correlation function between caregiver attention durations and infant theta activity revealed a similar pattern (Figure 4.4b), peaking in the period after time 0, and the cluster-based permutation analysis revealed a significant cluster ranging from -1 to 5s (p=0.012). This indicates that caregiver attention durations forwards-predicted infant theta at lags up to 1 second, as well as that infant theta forwards-predicted caregiver attention durations at lags of up to 5 seconds.

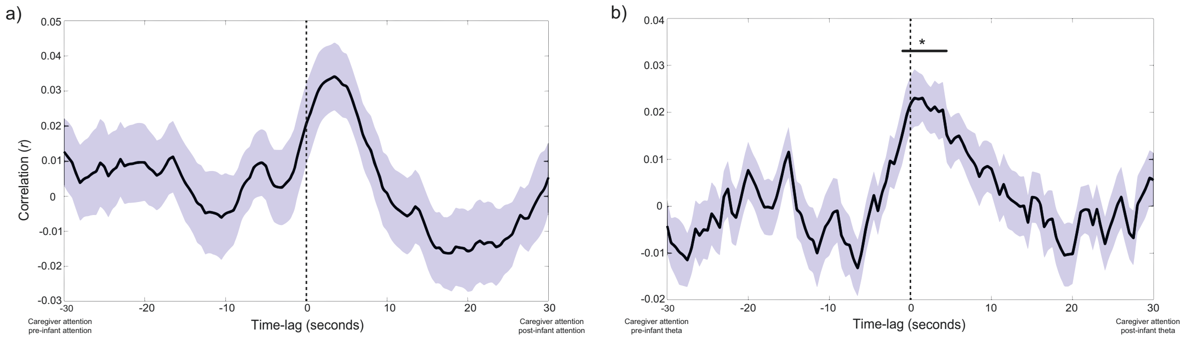

Figure 4. 4 Assessing forwards-predictive associations between caregiver attention durations, infant attention durations, and infant theta activity. Black lines show the Pearson cross-correlation between two variables; shaded areas indicate the SEM. Black horizontal lines show significant clusters of time lags. a) Infant and caregiver attention durations to objects. Cluster based permutation analysis revealed no significant clusters of time points (*p < 0.05), although one cluster verged on significance (p=0.10). b) Infant theta activity and caregiver attention durations to objects. Cluster-based permutation analysis indicated one significant cluster ranging from -1 to 5s (p=0.012).

4.3.3.2 Reactive change in caregiver look durations following infant look onset

To examine reactive change in caregiver attention to objects following the onsets of infant attention episodes to objects, we time-locked caregiver attention durations to infant attention onsets towards objects. Figure 4.5a shows changes in caregiver attention durations around the onset of infant attention towards an object. Cluster-based permutation analysis revealed a significant cluster of time points 0 to 4 seconds post attention onset ($p = 0.009$), indicating that caregiver attention durations significantly decreased after the onset of a new infant attention episode. Figure 4.5b shows the same event-related analysis subdivided by infant attention duration. This revealed that the decrease in caregiver attention durations after infant attention onsets was significant for attention episodes lasting over 3s.

To investigate how caregiver behaviour changed over the course of infant object looks, we next employed the same modulation analysis as described in section 4.3.2, computing differences in mean caregiver attention durations between 3 equal-spaced chunks over the course of an infant object look. This analysis revealed that, in contrast to the first 4 seconds of an infant attention episode during which caregiver attention durations decreased, caregiver attention durations actually significantly increased over the course of the entire attention episode, with a Wilcoxon signed ranks tests indicating a significant difference between the first chunk of an attention episode and the third (Figure 4.5c). Dividing infant attention durations into log-spaced bins again revealed that this effect was driven by attention episodes lasting over 3s (Figure 4.5d). Finally, we computed a linear mixed effect model to examine the relationship between infant object attention durations and caregiver attention durations. Corresponding to the modulation analyses reported above, when we averaged over the course of the entire infant object attention episode, we found that longer infant object attention durations associated with longer caregiver attention durations ($\beta = 0.16$, $p < 0.001$; Figure 4.5e).

4.3.3.3 Summary of caregiver gaze behaviour

In summary, both the continuous and event-related analyses revealed that caregivers dynamically adapted their gaze behaviour in response to changes in infant attentiveness during the interaction. Infant theta activity significantly forwards-predicted caregiver attention durations, suggesting that caregivers dynamically adapt their behaviour according to infant engagement (Figure 4.4b). Caregiver attention durations to objects decreased around the start of a new infant attention episode (Fig 4.5a); but overall, longer infant attention durations associated with longer attention durations by the caregiver (Figure 4.5e). These analyses demonstrate immediate, reactive, change in caregiver behaviour at the onset of infant attention towards an object, as well as slower-changing modulations in their behaviour over the length of an attention episode.

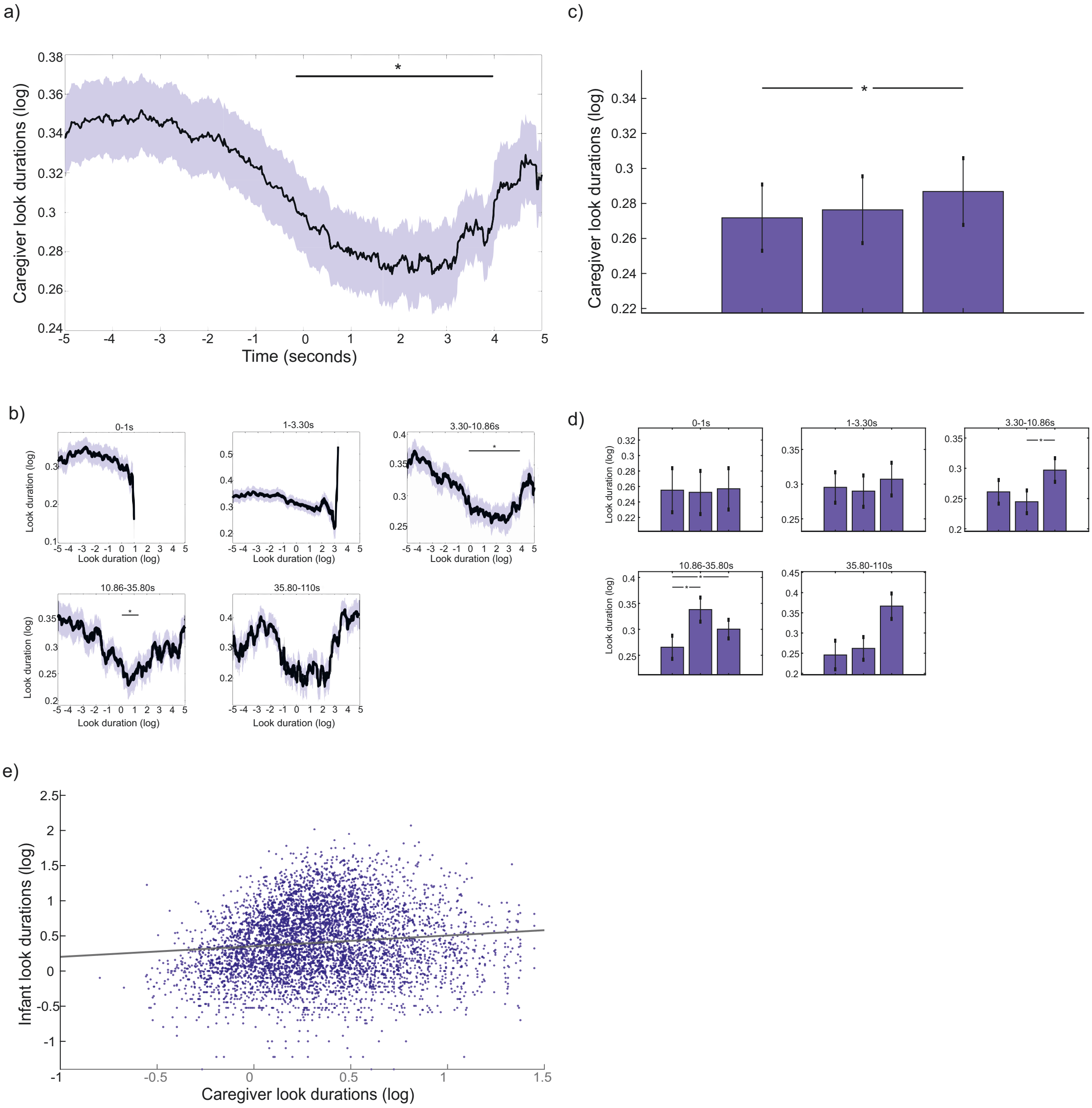

Figure 4. 5 Dynamic event-locked associations between infant object attention and caregiver attention durations. a) Event-related analysis showing change in caregiver attention durations around infant attention onsets to objects: black line shows average caregiver attention durations (log); shaded areas indicate the SEM. Black horizontal line shows areas of significance revealed by the cluster-based permutation analysis ($p < 0.05$). Cluster-based permutation analysis reveals a significant cluster of time points 0 to 4 secs

before and after attention onset ($p = 0.009$). b) Event-related analysis split by infant object attention-duration time bins. Black lines show average caregiver attention durations (log); shaded areas indicate the SEM. Black horizontal lines shows areas of significance revealed by the cluster-based permutation analysis ($p < 0.05$). Permutation analysis again revealed a decrease in caregiver attention durations in the time after attention onset for looks 3-35s long. c) Modulation analysis: each bar shows the median caregiver attention duration, across participants, for each chunk, averaged across all infant object attention durations. Wilcoxon signed ranks tests investigated significant differences between chunks (*$p < 0.05$). d) Same as c), for each infant object attention duration time bin. e) Scatter-plot showing the association between infant object attention durations and adult attention durations. Coloured dots show each individual object attention duration; black line shows the linear line of best fit. A linear mixed effects model revealed a significant association between the two variables ($\beta = 0.16$, $p < 0.001$).

4.3.3.4 Forwards-predictive relationships between infant look durations and caregiver vocal behaviour

Next, we used an identical analysis approach to examine forward-predictive and reactive associations between infant attention and caregiver vocal behaviours. Here, we concentrate on the rate of change in F0 as a marker of auditory saliency in the caregiver's voice. In additional analyses presented in the SM, we also examine caregiver vocal durations, and caregiver amplitude modulations (Figure S4). First, we computed the cross-correlations between rate of change of caregiver F0, infant attention, and endogenous infant oscillatory activity. Results are shown in Figure 4.6. Cluster-based permutation analysis revealed that the time-lagged associations between infant attention and rate of change in caregiver F0 did not exceed chance (both p>0.05). To test whether there was any direct influence of caregiver behaviour on modulations in infant endogenous neural activity, the same analyses were subsequently repeated relative to infant theta activity (Figure 4.6b): cluster-based permutation analysis again suggested no significant association between caregiver vocal saliency and infant endogenous neural activity (p>0.05). The same analyses are presented relative to caregiver vocal durations and amplitude modulations in Figure S4, which showed a similar pattern of results.

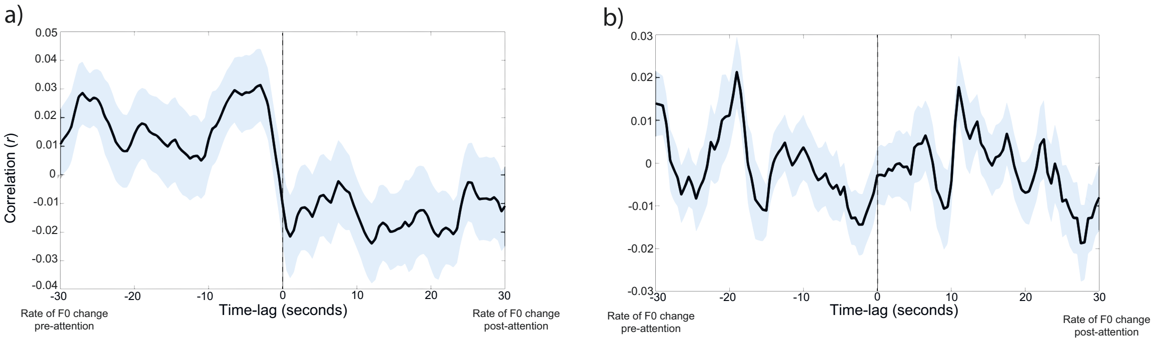

Figure 4. 6 Assessing forwards-predictive associations between caregiver vocal behaviour, infant attention durations, and infant theta activity. Cross correlation between the rate of change in caregiver F0 and a) infant look durations, and b) infant theta activity. Black lines show the Pearson cross-correlation between the two variables; shaded areas indicate the SEM. Black horizontal lines show significant clusters of time lags (*p<0.05). Cluster-based permutation analysis revealed no significant clusters of time-lags.

4.3.3.5. Reactive change in caregiver vocal behaviour following infant look onset

To examine whether caregivers reactively adapted their vocal behaviour to changes in infant attention, we repeated the same analysis presented in section 3.3.2, with rate of change of caregiver F0 as the dependent variable. The event-related analysis revealed no change in the rate of change in caregiver F0 relative to infant attention onsets: cluster-based permutation analysis revealed no change above chance levels ($p>0.05$; Figure 4.7b). This suggests that modulations in the caregiver's speech were not immediately reactive to infant attention onsets towards objects. Over the length of individual attention episodes towards objects, however, linear mixed effects models revealed that longer object looks associated with an increased rate of change in caregiver F0 ($\beta = 0.13$; $p<0.001$; Fig 4.7a,), whilst the modulation analysis revealed a decrease in the rate of change in F0 which appeared largely driven by looks lasting between 3-10s (Figure 4.7c, 4.7d). The exact same analysis relative to caregiver vocal durations and amplitude modulations showed a similar pattern of findings, which is presented in Figure S5.

4.3.3.6 Summary of caregiver vocal behaviour

In summary, longer infant object look durations associated with a greater rate of change in caregiver F0 overall. Caregiver vocal behaviour showed no event-related change relative to infant attention onsets, but longer attention durations were associated with a decrease in the rate of change in F0 over the length of an infant object look.

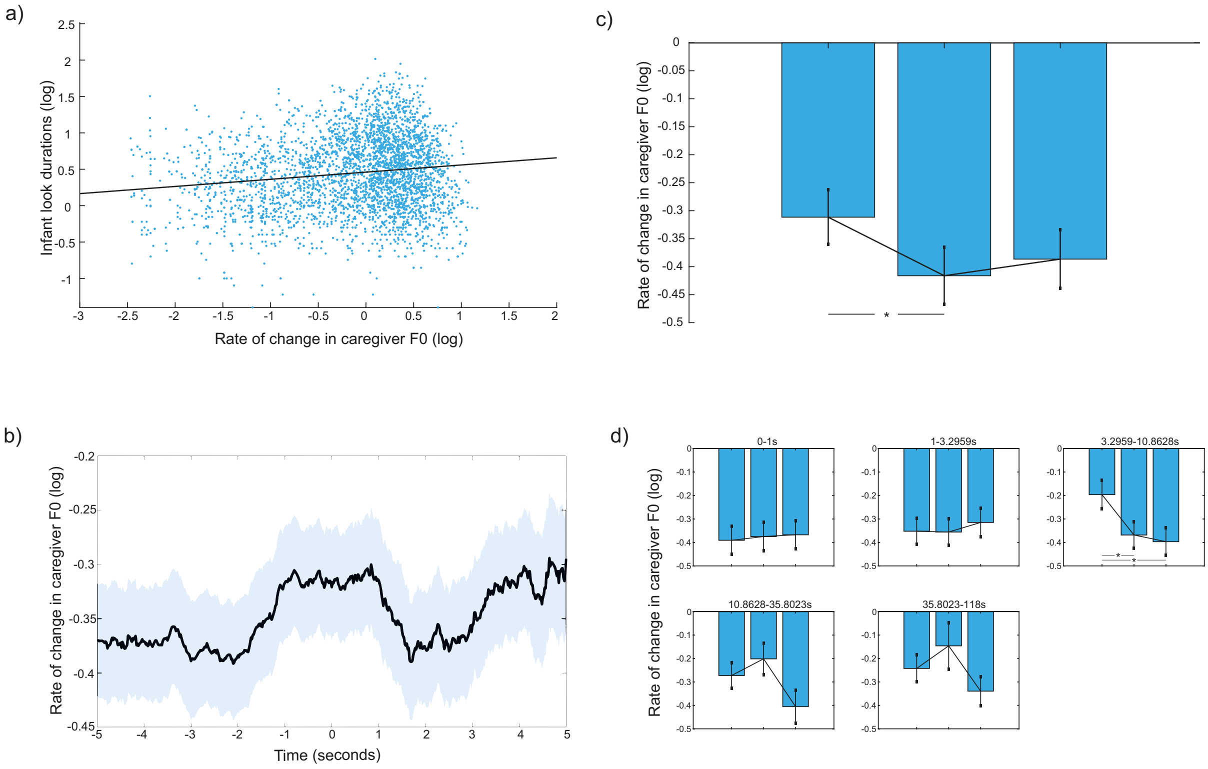

Figure 4. 7 Reactive change in caregiver vocal behaviour relative to infant attention onsets. a) Scatter plot of the association between infant attention duration and rate of change in caregiver fundamental frequency. A linear mixed effects model predicting infant look duration from the rate of change in caregiver F0 revealed a positive association between the two variables ($\beta = 0.13$; $p<0.001$. b) Event related analysis examining reactive change in caregiver F0 in the time after the onset of an infant object look. Black line indicates the average across participants; shaded area indicates SEM. Black horizontal line shows areas of significance revealed by the cluster-based permutation analysis ($p < 0.05$). c) Modulation analysis: each bar shows the median for each chunk across participants; errors bars show the SEM. Wilcoxon signed ranks tests explored significant differences between attention chunks (*$p < 0.05$), d) Same as c), binned by infant attention durations.

4.4. Discussion

We recorded the microdynamics of infant attention whilst they engaged in naturalistic, free-flowing interactions with their caregiver. We conducted three analyses. First, we examined oscillatory structures in caregiver and infant attention, and whether caregiver and infant attention patterns act as coupled oscillators (section 4.3.1). Second, we examined how infants' endogenous neural activity forwards-predicted attention durations, and how it changed reactively relative to the onsets of infant attention episodes towards objects (section 4.3.2). Third, we examined how caregiver gaze and vocal behaviour forwards-predicted infant attention durations, and how it changed reactively to the onsets of infant object looks (4.3.3).

When we examined how intrinsic cognitive processes control infant attention we found evidence for two distinct mechanisms. First, oscillatory mechanisms govern infant attention durations (section 4.3.1, Figure 4.2), with a period centring around 1-2 seconds in length. Second, independently, fluctuations in neural markers of infants' engagement or interest forward-predict their attentiveness towards objects (section 4.2.1). Cross-correlation analyses revealed associations between infant theta activity and infant attention durations, such that increases in infant attention durations forwards-predicted increases in infant theta activity more than *vice versa* (Figure 4.3a). Overall, average theta power during an attention episode correlated with the duration of that episode (Figure 4.3b). Infant theta activity did not, however, show any immediate change at the beginning of an attention episode, or modulate over the length of longer episodes (Figure 4.3c; S3).

Overall these findings suggest, consistent with the predictions of active learning models (Begus & Southgate, 2018; Kidd & Hayden, 2015), that infants' own endogenous cognitive processing is one mechanism that drives and maintains infant attention during online interactions. Infant theta activity associated with the duration of infant looks towards objects, but did not show any immediate, reactive change in the time just after attention onset, or slower change over the duration of the attention episode. The findings presented here could suggest that longer object attention by the infant occurs where infant engagement is higher, pointing to a bidirectional association between infant look durations and infant theta activity, at time scales longer than a single look. Corresponding to this, the cross correlation of the association between all infant look durations and infant theta activity suggested that infant attention durations showed both forwards- and backwards- predictive associations with infant theta activity. In this perspective, these findings can be interpreted relative to active sampling models of visual foraging in early infancy that predict a self-sustaining, bidirectional interaction between exploratory behaviours by the infant, and information gain from the environment (Kidd et al., 2012, 2014; Oudeyer & Smith, 2016; Saez de Urabain et al., 2017).

Next, we examined the association between caregiver behaviours and infant attention. Consistent with previous research (Yu & Smith, 2016) we found that oscillatory mechanisms govern both caregiver and infant attention durations, but that the oscillatory period of infant attention durations is shorter (centring around 1-2 seconds in length) compared with caregivers' (centring around 200- 500ms in length; Figure 2a, 2b). However, when we examined whether infant and caregiver attention patterns act as coupled oscillators, which is one mechanism through which caregiver gaze behaviour might support infant gaze behaviour (Nuthmann & Henderson, 2010), we found no evidence to support this (Figure 2c). This suggests that mechanisms of influence between infant and caregiver attention are more likely to operate as lagged, forwards- or backwards- predictive relationships, as we investigated in section 4.3.3.

We found little to no evidence in support of the hypothesis that adult gaze and vocal behaviours forwards-predict infant attention (4.3.3). Against adult-led attentional structuring perspectives of early interaction, the cross-correlation analyses showed that, overall, fluctuations in infant look durations were not forwards-predicted by changes in caregiver look durations (Figure 4.4a), or modulations in the temporal and spectral properties of their voice (Figure 4.6; Figure S4); rather, changes in infant neural engagement largely forward-predicted changes in caregiver attention durations (Figure 4.4b). This association was likely partially mediated by the weaker and non-significant associations observed between infant attention and caregiver attention (Figure 4.4a). We also found no evidence for co-fluctuations between the rate of change of caregiver F0 (a marker of auditory salience) and infant theta activity (Figure 4.6b).

We did, however, find evidence that caregivers rapidly modulated their behaviours in response to shifts in infant attention. This was particularly evident in adult gaze behaviour, where in addition to the cross-correlation findings (Figure 4.4) our event-locked analyses showed that caregiver attention durations significantly decreased after the onset of a new infant attention episode (Figure 5.5a). Over the duration of longer attention episodes, however, caregiver attention durations significantly increased (Figure 4.5c), so that, overall, the linear mixed effects model revealed that longer infant object looks were associated with longer looks by the adult partner (Figure 4.5e). Whilst linear mixed effects models revealed that infant attention durations co-occurred with longer caregiver vocalisations, along with a greater rate of change in caregiver F0 and an increase in caregiver amplitude (Figure 7a; S5), we observed little dynamic change in caregiver vocal behaviour immediately after attention onset, or over the duration of attention episodes, of any length (Figure7, Figure S5).

Overall, caregiver behaviours were largely reactive to changes in infant attention. The rapid change in caregiver gaze in response to the onset of infant attention towards objects, beginning just before attention onset, suggests that it is unlikely that caregivers are responding to active attention sharing cues produced by the infant (Begus et al., 2014; Kovács et al., 2014; Tomasello et al., 2007). Indeed, similar to previous micro-behavioural studies of 12-month-old infants in shared interactions, infants rarely looked towards their caregiver's face (Figure 4.2a), and, in a previous analysis of this data, infants did not increase looks to their partner's face in the time before leading an episode of joint attention (Phillips et al., in press). It seems therefore unlikely that the relationship between infant attention and fluctuations in their own endogenous cognitive processing is related to intentionally mediated forms of communication by the infant, with the goal of directly eliciting information from their caregiver (Phillips et al., in press; Yu & Smith, 2013).

Instead, caregivers are anticipating shifts in infant attention, and, in line with an allostatic model of inter-personal interaction, 'catching' infants' attention, and monitoring their behaviour (Yu & Smith, 2016). This increase in the rate of caregiver behaviour after look onsets could reflect dynamic up-regulatory processes that serve to maintain infant attention: though not reflected in their vocalisations; other fast-changing salient cues such as hand movements and facial affect could also increase in variability (M. Meyer et al., 2022). The down-regulation of caregiver attention over the course of longer attention episodes by the infant might subsequently index decoupling of caregivers' regulatory processes from infant attention; this is also reflected in the decreased rate of change in caregiver F0 (Figure 6c).

Combined, therefore, our findings suggest that, during interactions at the end of the first year, infant attention is structured through joint but independent influences of caregiver responsivity and regulation, and their own intrinsically motivated engagement.

In this perspective, our results can be interpreted relative to neurocomputational, associative accounts of active learning in early infancy (Kidd & Hayden, 2015; Oudeyer & Smith, 2016). These accounts postulate that contingent changes in the environment in response to actions produced by the infant improves infants' prediction and control over their own behaviour (Friston et al., 2012; Friston, 2019). In the context of shared interaction, consistent and contingent responsiveness by the caregiver to infant attention gives meaning to infants' behaviour, increasing infant engagement and further exploratory behaviours (Oudeyer & Smith, 2016; Smith & Breazeal, 2007). Over time, therefore, infants' experience of repeated interactive contingencies could influence how controlled processes begin to guide their attention, as well as their sensitivity to and engagement in intentionally mediated forms of shared communication (Smith & Breazeal, 2007).

This has implications for how we view and understand the interactive processes that support how infants begin to use and engage with a language system. Previous accounts have emphasised the role of the caregiver in structuring infant learning in joint attentional frames, where they use clear ostensive signals to guide infant attention, and support word-object representations (Lieven, 2016; Tomasello et al., 2007). The present study, however found no evidence that increases in salient cues by the caregiver forward-predicted increases in infant attention durations. Increases in infant attentiveness are instead related to inter-dyadic, sensorimotor processes that are independent of the influence of infants' own endogenous cognitive process. How these fast-acting intra- and inter-individual influences on infant attention support early language acquisition should be a key focus for future research (Yu et al., 2021; Yu & Smith, 2012).

The naturalistic design of our study is a strength as well as a limitation. Of note, we were unable to control how much infants moved during the interaction, which may have contributed to eye movement artifacts time-locked to shifts in infant attention. However, eye-movement artifacts were removed using ICA decomposition, and, though this does not remove all artifact introduced to the EEG signal (Marriott Haresign et al., 2021), the relationships observed between infant attention and theta activity suggest that this did not affect our main findings. If eye-movement artifact influenced the association between infant attention duration and theta activity, shorter attention episodes ought to associate with more theta activity, which was not the case.

In future work it will be important to take a more holistic, computational and multi-modal approach to studying how factors intrinsic to the infant, and the inter-personal behavioural contingencies of the dyad, structure infant attention and behaviour (Xu et al., 2017). For example, studying how inter-related multi-modal patterns of caregiver behaviour, such as body movement (Meyer et al., 2022), facial affect (Murray et al., 2016; Rayson et al., 2017) and vocalisations (Goldstein & Schwade, 2008) support infants' engagement in joint attention, will build on the work that we reported here. In addition to the micro-dynamic analyses that we present here, it will also be important for future work to employ modelling approaches to further investigate infants' neural entrainment to the unidirectional and inter-dyadic action-generated contingencies of shared interaction (Jessen et al., 2019; Jessen et al., 2021). A particular focus of this work should be on studying the temporal latencies at which entrainment and/or behavioural responsivity occur; utilising eye tracking methods will help with this.

Overall, our findings suggest that infant attention in early interaction is asymmetric, related to their own endogenous cognitive processing and to consistent, reactive contingency to changes in their attention by the caregiver. Active learning strategies operate across the dyad; and are likely foundational to early language acquisition and socio-cognitive learning.

CHAPTER 5 - Using continuous methods of analysis to examine speech-brain tracking during naturalistic caregiver-infant object play

This chapter is a methods paper that aims to assess the methodological issues involved in examining speech-brain tracking during naturalistic social interactions. The manuscript is currently in preparation for publication. The supplementary materials for this chapter are presented in Appendix C.

Abstract

Entrainment to slow modulations in the amplitude envelope of infant-directed speech is thought to drive early phonological learning (Leong et al., 2017). Whilst recent research with infants has applied continuous modelling approaches to examining speech-brain tracking, this work has so far been conducted in controlled experimental settings: far from the fast-changing, complex and cluttered environments of everyday interactions. Here we aim to test the applicability of mTRF modelling to examining speech-brain tracking in naturalistic interactions of 9-12-month-old infants interacting with their caregivers. To do so, using a multivariate, backwards modelling approach we test both individual and generic training procedures to examine the effects of data quantity and quality on model fitting. We show that model fitting is most optimal using a subject-dependent approach, and, corresponding to previous findings, we find significant speech-brain tracking at delta modulation frequencies. This work is important in opening up new avenues to examining the dynamic interactive mechanisms that support speech-brain tracking and drive early language learning in

naturalistic settings. In future work, it will be important to explore, in more detail, and across varied naturalistic paradigms, the effects of data quality on model performance.

5.1 Introduction

Infants' ability to learn a language system from inconsistent and irregular environmental input has been a key focus of developmental research (Aslin & Newport, 2012; Yu & Smith, 2012). The neural mechanisms foundational to acquiring language are, however, only just beginning to be understood (Leong & Goswami, 2015). A large body of literature has shown that adults' neural oscillations entrain to the syllabic structure of speech, corresponding to theta rate modulations in the amplitude envelope (Luo & Poeppel, 2007). Recent work has begun to explore when and how infants begin to entrain to the temporal modulation structure of speech (e.g. Attaheri et al., 2022). This work has, however, so far been conducted in structured, experimental paradigms; far from the cluttered, dynamic and complex nature of naturalistic interactions in which infants begin to learn a phonological system (Leong & Goswami, 2015; Yu & Smith, 2012).

Neural tracking of the speech amplitude envelope at different temporal modulation frequencies supports our ability to parse the speech signal into linguistically relevant units (Giraud & Poeppel, 2012; Gross et al., 2013; Poeppel, 2014). In adults, neural tracking has been shown at delta, theta and low gamma frequencies, corresponding to the stress, syllabic and phonetic patterning of the phonological information in speech (Ding et al., 2016; Gross et al., 2013). Oscillatory tracking of a speech signal in the auditory cortex is thought to be hierarchically nested, with slower theta-rate oscillations dynamically modulating the activity of oscillations occurring at higher frequencies (Giraud & Poeppel, 2012). In particular, the phase of theta-rate oscillations is thought to modulate the power and phase of gamma oscillations, supporting the integration of syllabic and phonetic information (e.g. Gross et al., 2013).

Whilst neural tracking at theta-rate modulations is particularly important to speech encoding and intelligibility in adults (Doelling et al., 2014), it is thought that, in early infancy, delta-rate modulations might play a fundamental role in early phonological development (Leong & Goswami, 2015). Recent work has shown that the peak modulation frequency of IDS – i.e. the modulation rate with the most energy, occurs at delta frequencies (~2Hz; Leong et al., 2017). This is in contrast to ADS, where peak modulation rates occur within the theta band (Greenberg et al., 2003; Leong et al., 2017). It has been suggested that the stress-shifted rhythm of infant directed might support neural entrainment to delta-rate temporal speech patterns in infancy, which could be particularly important to locating word boundaries during early language learning (Leong & Goswami, 2015; Power et al., 2013). Supporting this perspective, research on children with dyslexia has shown atypical entrainment to delta-rate amplitude modulations in pre-recorded speech, but typical entrainment at theta frequencies (Power et al., 2013, 2016).

In adults, speech tracking has often been examined by recording the neural response of participants to short, repeated experimental stimuli. For example, one popular method involves presenting participants with a repeated speech stimulus, and calculating the inter-trial phase consistency across neural responses (Luo & Poeppel, 2007; Zion Golumbic et al., 2013) Alternatively, researchers compute the phase locking value between a participants' neural response at a specific frequency and the amplitude envelope of the experimental speech stimuli at the same frequency, and average the phase locking value across segments of short, repeated trials (Doelling et al., 2014; Peelle et al., 2013).

In more recent years, however, the application of continuous methods of analysis to examining ongoing neural responses to a speech signal have become a particular focus (Crosse et al., 2015, 2016, 2021; Jessen et al., 2019; Lalor et al., 2009; Obleser & Kayser, 2019). One popular approach is to model the impulse response function between the continuous speech signal, and ongoing EEG response (Brodbeck & Simon, 2020; Crosse et al., 2016, 2021; Ding & Simon, 2012). These models are referred to as temporal response functions (TRFs), that estimate the linear mapping from an environmental stimulus to the EEG response via a method of regularized regression (Crosse et al., 2016). Through an iterative process of training and testing, a regression model, describing the relationship between the speech amplitude envelope (e.g. 1-8Hz; (Kalashnikova et al., 2018)), and the EEG signal is computed at each electrode over a range of time-lags (Crosse et al., 2016). The resulting TRFs can be interpreted similarly to an ERP component; and the coefficients are directly neurophysiologically interpretable (i.e. the magnitude of a TRF coefficient corresponds to the extent of neural tracking to the stimulus signal at a particular channel, within a certain frequency range; Crosse et al., 2021; Fiedler et al., 2017; Jessen et al., 2021). The accuracy of the model is computed by calculating the correlation between the actual neural response and that predicted by the model (Crosse et al., 2016).

As well as forwards models, backwards, or decoding, mTRF models can also be used to model the linear mapping from the stimulus (speech signal) back to the neural response. This method has become a widely used approach to investigating speech-brain tracking in the adult literature (Bednar & Lalor, 2020; Crosse et al., 2015; Ding & Simon, 2012). Unlike forward encoding models, backwards models do not require the pre-selection of electrodes to enter into the model; instead, each channel is assigned weights at each time lag, based on how much information that channel provides for the reconstruction i.e. channels that track the speech signal to a very little extent and those that track the signal to a great extent will both have high weights because they provide most information to the reconstruction (Crosse et al., 2016)). As a consequence, in contrast to encoding models, TRF weights are not directly neurophysiologically interpretable, but methods for forwards-transforming the weights are available (Haufe et al., 2014).

Two different approaches to model training can be taken: an individual training approach, or a generic approach. Individual model training is the most commonly used method, whereby a TRF model is trained on the trials of an individual participants' data, and then tested on held aside data for that participant. In this way, individual training methods identify consistencies in neural tracking within participants (Jessen et al., 2019; Jessen et al., 2021a). The second approach is a generic training approach: here, a TRF model for each participant is trained by pooling across the models constructed for all other participants. This method is advisable with small or noisier data sets in increasing the amount of data the model is trained on comparative to individual models (Crosse et al., 2021; Jessen et al., 2021). In training models across participants, generic models also inform our understanding of consistencies in neural tracking of the speech signal between participants (Jessen et al., 2021). This is particularly interesting relative to examining speech-brain tracking in naturalistic interactions where each infant listens to a different speech signal (i.e. their caregivers).

Three studies have recently used this method to examine speech-brain tracking to naturalistic stimuli in early infancy. Jessen and colleagues used a forwards modelling generic training approach to compute a TRF model of the relationship between infants' broadband EEG signal (1-40Hz) and the amplitude envelope of a cartoon video that the infants engaged with for just over two minutes. They identified significant speech-brain tracking of the amplitude envelope in their sample of 11-12-month olds, which was also present but more variable with an individual modelling procedure (Jessen et al., 2019). More recently, Attaheri et al., 2022 constructed a backwards mTRF to model the relationship between 4, 7 and 11-month-old infants' continuous neural activity and sung nursery rhymes. Models were computed between the amplitude envelope of the nursery rhymes (filtered at 0.5-15Hz), and infants' EEG responses, filtered into three frequency bands: delta (0.5-4Hz), theta (4-8Hz) and alpha (8-12Hz), using an individual training approach. Alpha band tracking was examined as a control, based on the finding that alpha frequencies do not play a role in computational models of IDS (Leong et al., 2017). Corresponding to the prediction that infants track the amplitude envelope of speech at low modulation frequencies (Leong et al., 2017), across all 3 ages, significant speech-brain tracking was identified at delta and theta frequencies but not alpha frequencies. This finding supports previous reports of significant speech-brain tracking to IDS (0.5-8Hz), but not adult direct speech, among infants aged 7 months using an individual training, forwards TRF modelling procedure (Kalashnikova et al., 2018).

Almost all previous work examining neural tracking to the temporal modulation structure of speech in infant and adult populations has, however, been conducted using structured experimental paradigms, where participants are presented with a continuous speech sound, and their neural response to that speech sound recorded (Attaheri et al., 2022; Bednar & Lalor, 2020; Crosse et al., 2015; Jessen et al., 2019; Kalashnikova et al., 2018). This is very different to the speech infants hear in naturalistic, everyday settings where vocalisations are variable in length, often repetitive and short, and occur in the context of background noise. In contrast to the lengthy speech stimuli often used in structured paradigms, in naturalistic, face-to-face interactions caregiver vocalisations tend to be short, lasting an average of 1s, and ranging between 0.5-9s (Gratier et al., 2015; Jaffe et al., 2001). Examining speech-brain tracking during naturalistic free-flowing interactions is also important to informing our understanding of how infant processing of the temporal modulation structure of speech is related to the ongoing dynamics of shared interactions (joint attention behaviours, for example), as well as its relationship to the semantic timing and complexity of the caregivers' speech inputs (Nencheva & Lew-Williams, 2022). Examining speech-brain tracking during naturalistic interactions, is, however particularly problematic, given the noise inherent to both the infants' EEG signal and the caregivers' speech (Georgieva et al., 2020; Noreika et al., 2020). Naturalistic EEG recordings are particularly affected by movement artifact, and although ICA decomposition removes some of these components, it does not remove all (Marriott Haresign et al., 2021). During shared object play, the caregivers' speech signal is also affected by background noise in the interaction, including toy clacks, as well as infant movement and co-vocalisations (Gratier et al., 2015).

Here, we aim to examine speech-brain tracking during naturalistic interactive caregiver-infant interactions using a backwards mTRF modelling approach. To specifically test whether patterns of speech-brain tracking that are observed in controlled experimental paradigms, transfer to interactive contexts, we match the analysis design of Attaheri et al. (2022). In particular, we compute mTRF models that map the wideband amplitude envelope (0.5-15Hz) to the infant EEG signal at delta (1-4Hz), theta (4-8Hz) and alpha (8-12Hz) frequencies to test whether speech-brain tracking differs significantly from chance at slow modulation frequencies only.

Given that this is the first time continuous methods have been used to examine speech-brain tracking during naturalistic play, we take two different approaches to segmenting the caregivers' speech signal as inputs to the mTRF model. First, we train the mTRF models on continuous data segments, dividing each interaction up into 10 equal-length parts, corresponding to the approach most often taken in experimental designs with adults, and infants, where a continuous stimulus is presented (Jessen et al., 2019). This approach also optimised the amount of data on which the model was trained. Data quality and quantity have been shown to affect the performance of mTRF models, such that noisier data, and models trained on small data sets, often lead to overfitting, resulting in models with poor transferability to new data (Crosse et al., 2021). Due to the fact that, in adult studies, data quality tends to be good and there is a lot of it, extensive testing of how data quality and quantity influence model performance has not been extensively tested (Mesik & Wojtczak, 2022).

Second, we computed mTRF models with each caregiver vocalisation serving as an individual fold in the model (see Sohoglu & Davis, 2020 for a similar approach with adults). Though this approach decreased the amount of data that the models are trained on, it also reduced the amount of noise in the data (i.e. in this way only periods that the caregiver is speaking are included in the model, and this also means that analyses can be run on vocalisations where no background noise or covocalisations occurred within a vocalisation). In this way we computed separate models for all vocalisations lasting 500ms or longer, vocalisations lasting 2000ms or longer, as well as vocalisations lasting 500ms or longer that included no background noise (e.g. toy clacks, and infant vocalisations).

Different to Attaheri et al. (2022), where an individual training method alone was used, due to the noisiness of our data, and the reduced amount of data entered into the vocal chunking models, we also present models computed according to generic training procedures. To test for potential overfitting where individual models are computed, we report predictive accuracy scores obtained during model training and testing to examine whether accuracy scores for training sets, computed during the cross-validation procedure, are much higher than testing sets (see Methods for more details): indicating poor transferability to new data, and therefore insufficient model fitting (Crosse et al., 2021). The generic models also provide an interesting comparison to the individual models in testing for consistencies in speech tracking by infants to their own caregiver's speech signal across participants (Jessen et al., 2019). Something that has not previously been examined.

We hypothesised that speech-brain tracking would be observed at slow amplitude modulation frequencies (delta and theta) for both individual and generic models (Attaheri et al., 2022), and that overfitting of the individual models to the training data would increase, the less data was entered into the model (Crosse et al., 2021).

5.2 Method

5.2.1 Participants

Ninety-four caregiver-infant dyads took part in this study. The final overall sample with usable, coded, gaze data was 46 (32 infants were excluded due to recording error or equipment failure, 4 infants were excluded for fussiness and 12 infants were excluded due to poor quality EEG data). Power analysis was conducted using the G*power tool (Faul et al., 2007): data from a pre-existing data set using mTRF modelling to examine speech-brain tracking, among 10-12-month-old infants, to pre-recorded nursery rhyme speech was used as an estimator of expected effect size (0.571; Siew et al., in prep). Based on an Alpha level of 0.05, the sample size reported here has a power of 95%. For the final, overall sample, the mean age of participants was 11.14 months (SD=1.30; 26 females, 20 males). All caregivers were female. Participants were recruited through baby groups and Children's' Centers in the Boroughs of Newham and Tower Hamlets, as well as through online platforms such as Facebook, Twitter and Instagram. Written informed consent was obtained from all participants before taking part in the study, and consent to publish was obtained for all identifiable images used. All experimental procedures were reviewed and approved by the University of East London Ethics Committee.

5.2.2 Experimental set-up

Parents and infants were seated facing each other on opposite sides of a 65cm wide table. Infants were seated in a high-chair, within easy reach of the toys (see Fig. 1b). The shared toy play comprised two sections, with a different set of toys in each section, each lasting ~5 minutes each. Two different sets of three small, age-appropriate toys were used in each section; this number was chosen to encourage caregiver and infant attention to move between the objects, whilst leaving the table uncluttered enough for caregiver and infant gaze behaviour to be accurately recorded (cf.Yu & Smith, 2013).

At the beginning of the play session, a researcher placed the toys on the table, in the same order for each participant, and asked the caregiver to play with their infant just as they would at home. Both researchers stayed behind a screen out of view of caregiver and infant, except for the short break between play sessions.

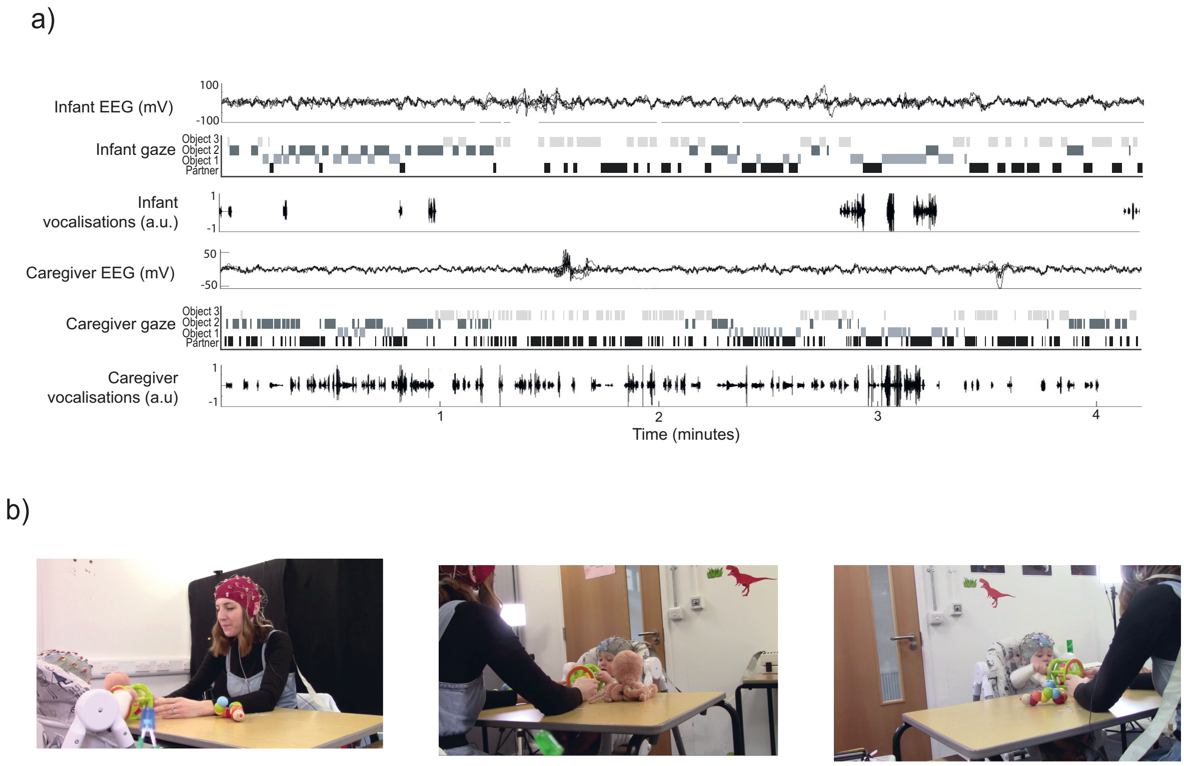

Figure 5. 1 Experimental set-up. a) Raw data sample, showing (from top) infant EEG over fronto-central electrodes, after pre-processing, infant gaze behaviour, infant vocalisations, adult EEG over fronto-central electrodes, adult gaze behaviour, adult vocalisations. b)

Example camera angles for caregiver and infant (right and left), as well as zoomed-in images of caregiver and infant faces, used for coding.

5.2.3 Equipment

EEG signals were recorded using a 32-channel BioSemi gel-based ActiveTwo system with a sampling rate of 512Hz with no online filtering using Actiview Software. The interaction was filmed using three Canon LEGRIA HF R806 camcorders recording at 50 fps. Parent and infant vocalisations were also recorded throughout the play session, using a ZOOM H4n Pro Handy Recorder and Sennheiner EW 112P G4-R receiver.

Two cameras faced the infant: one placed on the left of the caregiver, and one on the right (see Fig. 1b). Cameras were placed so that the infant's gaze and the three objects placed on the table were clearly visible, as well as a side-view of the caregiver's torso and head. One camera faced the caregiver, positioned just behind the left or right side of the infant's high-chair (counter-balanced across participants). One microphone was attached to the caregiver's clothing and the other to the infant's high-chair.

Caregiver and infant cameras were synchronised to the EEG via radio frequency (RF) receiver LED boxes attached to each camera. The RF boxes simultaneously received trigger signals from a single source (computer running MATLAB) at the beginning of each play section, and concurrently emitted light impulses, visible in each camera. Microphone data was synchronised with the infants' video stream via a xylophone tone recorded in the infant camera and both microphones, which was hand identified in the recordings by trained coders.

All systems were extensively tested and found to be free of latency and drift between EEG, camera and microphone to an accuracy of +/- 20 ms.

5.2.4 Vocalisation coding

The onset and offset times of caregiver and infant vocalisations were identified using an automatic detector. The algorithm detected voiced segments and compared the volume and fundamental frequency detected in each recorded channel to infer the probable speaker (mother vs. infant). Identification of the onset and offset times of the detector then underwent a secondary analysis by trained coders, who identified misidentification of utterances by the automatic decoder, as well as classifying the speaker for each vocalisation. As the detector did not accurately identify onset and offset times of caregiver and infant speech during co-vocalisations, and, as periods of vocalisation by the infant during the caregivers' speech would introduce noise to the caregivers' speech signal, as well as the possibility that infants' were entraining to their own speech (Pérez et al., 2021), these segments of data were excluded from the vocal chunking analysis. Whilst the detector was programmed so that a gap of 250ms was placed between vocalisations, after the coder had identified inaccuracies in the detectors' output, the vocalisations were re-processed so that a 1000ms gap was placed between each vocalisation. This step was taken to increase the lengths of the vocalisations included in the vocal chunking analysis.

5.2.5 Infant EEG artifact rejection and pre-processing

A fully automatic artifact rejection procedure including ICA was adopted, following procedures from commonly used toolboxes for EEG pre-processing in adults (Bigdely-Shamlo et al., 2015; Mullen, 2012) and infants (Debnath et al., 2020; Gabard-Durnam et al., 2018), and optimised and tested for use with our naturalistic infant EEG data (Georgieva et al., 2020; Marriott Haresign et al., 2021). This was composed of the following steps: first, EEG data were high-pass filtered at 1Hz (FIR filter with a Hamming window applied: order 3381 and 0.25/ 25% transition slope, passband edge of 1Hz and a cut-off frequency at -6dB of 0.75Hz). Although there is debate over the appropriateness of high pass filters when measuring ERPs (see (Widmann & Schröger, 2012), previous work suggests that this approach obtains the best possible ICA decomposition with our data (Dimigen, 2020; Marriott Haresign et al., 2021). Second, line noise was eliminated using the EEGLAB (Bigdely-Shamlo et al., 2015) function *clean_line.m* (Mullen, 2012).

Third, the data were referenced to a robust average reference (Bigdely-Shamlo et al., 2015). The robust reference was obtained by rejecting channels using the EEGLAB *clean_channels.m* function with the default settings and averaging the remaining channels. Fourth, noisy channels were rejected, using the EEGLAB function *clean_channels.m.* The function input parameters 'correlation threshold' and 'noise threshold' (inputs one and two) were set at 0.7 and 3 respectively; all other input parameters were set at their default values. Fifth, the channels identified in the previous stage were interpolated back, using the EEGLAB function eeg_interp.m. Interpolation is commonly carried out either before or after ICA cleaning but, in general, has been shown to make little difference to the overall decomposition (Delorme & Makeig, 2004). Infants with over 21% (7) electrodes interpolated were excluded from analysis. After exclusion, the mean number of electrodes interpolated for infants was 0.244 (*SD*=1.11) for play section 1, and 3.12 (*SD*=2.16) for play section 2.

Sixth, the data were low-pass filtered at 20Hz, again using an FIR filter with a Hamming window applied identically to the high-pass filter. Seventh, continuous data were automatically rejected in a sliding 1s epoch based on the percentage of channels (set here at 70% of channels) that exceed 5 standard deviations of the mean channel EEG power. For example, if more than 70% of channels in each 1-sec epoch exceed 5 times the standard deviation of the mean power for all channels then this epoch is marked for rejection. This step was applied very coarsely to remove only the very worst sections of data (where almost all channels were affected), which can arise during times when infants fuss or pull the caps. This step was applied at this point in the pipeline so that these sections of data were not inputted into the ICA. The mean percentage of data removed in play section 1 was 13.433 (SD=16.617), and 5.11(SD=7.03) for play section 2.

Data collected from the entire course of the play session (including play section 1 and play section two, as well as two further five minute interactions) were then concatenated and ICAs were computed on the continuous data using the EEGLAB function runica.m. The mean percentage of ICA components rejected was 51.90% (SD=16.92). After ICA rejection, data from each play section were re-split. For representative examples of artifactual ICA components identified in our naturalistic EEG data, see Marriott-Haresign et al. (2021).

5.2.6 mTRF analysis

5.2.6.1 Pre-processing

Removal of clipped segments from the speech signal

Due to the caregiver being within variable distance of their microphones, some clipping was identified in a sample of the microphone recordings. A stringent clipping identification algorithm was used to remove parts of the microphone data where clipping occurred (Hansen et al., 2021). Parts of the signal for which clipping was identified for longer than a period of 1ms were set as missing data points.

Amplitude envelope

The amplitude envelope of the caregivers' speech signal was extracted and filtered below 15Hz. First, the signal was band-pass filtered (using a 3rd order Butterworth filter (forwards and backward)) into 9 frequency bands, with equal spacing along the cochlear, according to Greenwood's (1990) equation. The absolute value of the analytic signal generated by the Hilbert transform was then computed and averaged across frequency bands (Gross et al., 2013). Next, the resulting wholeband amplitude envelope was filtered between 0.5 and 15Hz (low pass filter; 5th order butterworth filter (forwards and backwards), high-pass filter; 3rd order butterworth filter (forwards and backwards)). Finally, at this point, the speech signal was downsampled to the sample rate of the EEG signal (512Hz).

EEG data

After pre-processing, the EEG signal, for each participant, was filtered into delta (1-4Hz), theta (4-8Hz) and alpha (8-12Hz) frequencies. First, missing data points were excluded from the time series. Next, the EEG signal was filtered into the 3 frequency bands of interest using the *pop_eegfiltnew* function in the EEGLab toolbox (Delorme & Makeig, 2004), using zero-phase bandpass Hamming-windowed FIR filters.

5.2.6.2 Separating data into individual folds for training and testing

Continuous chunking analysis

For the continuous analysis, the filtered EEG and speech streams for each interaction (play section 1 and play section 2), for each participant, were split into equal-length folds, and allotted into training and test sets (see Figure 5.2 for a schematic depiction). First, the filtered EEG signal and the amplitude envelope of the caregivers' speech were both synchronised to the video-camera time-series, using the methods described above. Where one signal had been recorded longer than the other, that signal was shortened so that both data streams matched in length. The start and end video frames of each interaction were identified, and the beginning and end of the signals cut off where necessary. The start of each interaction was identified from the xylophone ding in the infants' camera, whilst the end of the interaction was taken as the infants' penultimate look of the interaction.

Next, in order to exclude periods of the interaction where the researcher was visible in the infants' camera frame, and possibly speaking, periods of the infant gaze time series that had been identified as uncodable were set as missing data points in both the EEG and speech signals. This step was taken to exclude periods of the interaction where the researcher was present and possibly speaking to the infant and/or caregiver. Though we did not directly code for periods of the interaction where the researcher was speaking, this procedure is likely to have eliminated most and if not all periods of speech by the researcher given that it was rare for the researcher to speak whilst the interaction was in progress unless they appeared from behind the screen. It is important also to note that uncodable periods in the infant gaze time series could be coded where the infants' eyes were blocked from view by one of the objects on the table; these periods were, however, rare, and only lasted for a short amount of time.

Next, corresponding parts of the EEG (due to missing data) and speech signals (due to clipping) that were set as missing data points in one stream were identified in the other alternate stream and set to missing data points. To increase computational efficiency, both the EEG and speech streams were downsampled to 128Hz. These continuous data streams were then saved for permutation testing (see section below). Both the speech and EEG streams were z-scored (EEG data were z-scored across channels), a step that is recommended when computing generic models for more consistent tuning of model parameters across data sets (Crosse et al., 2021). Finally, corresponding missing data points in both data streams were set to 0 (Jessen et al., 2021).

The speech and EEG signals for play section 1 and play section 2 were then split into 5 equal-length folds, and randomised before one set was placed into the testing set. So that 90% of the folds were allotted to the training sets and 10% to the test set for each participant, one of the testing sets (from play section 1 or play section 2) was placed randomly into the training sets. Given that play section 1 and play section 2 were of variable lengths, the combined training folds were also of variable size. Where a participant had data for only play section 1 or play section 2 (which occurred where, for example, the infants' EEG data was over the acceptable number of interpolated channels for one interaction), one fold was kept as the test set and the rest as the training sets. This step was taken to ensure that folds for each participant were of similar lengths.

Vocal chunking analysis

For the vocal chunking analysis, each caregiver vocalisation was cut out of the EEG and speech time-series and treated as a separate data fold (see Figure 5.2 for a schematic depiction). First, the start and end of each caregiver vocalisation was identified in the synchronised time-series described above (up to the point that the data was downsampled to 128Hz) and cut out of the corresponding time-series. Where any missing data points occurred in either the EEG or speech segment for one vocalisation, this vocalisation was removed from analysis. Next, the periods of vocalisation identified in both time-series were concatenated and z-scored, and then converted back into discrete vocalisations for both the EEG and speech signals. Finally, vocalisations for each play section were randomised and allotted to training and test sets, so that 90% went into the training set and 10% into the test set. By grouping together vocalisations of similar lengths, the randomisation procedure ensured that vocalisations in the training and test sets were of similar sizes. For participants with play section 1 and play section 2, training and test sets from each play section were concatenated, randomised and adjusted to ensure that 10% of the vocalisations were in the test set and 90% in the training set. The procedure was repeated for vocalisations with and without background noise lasting over 500ms and over 2000ms, resulting in four separate sets of vocalisation chunks. Where caregivers had less than 5 vocalisations for a respective set, that dyad was excluded from the analysis.

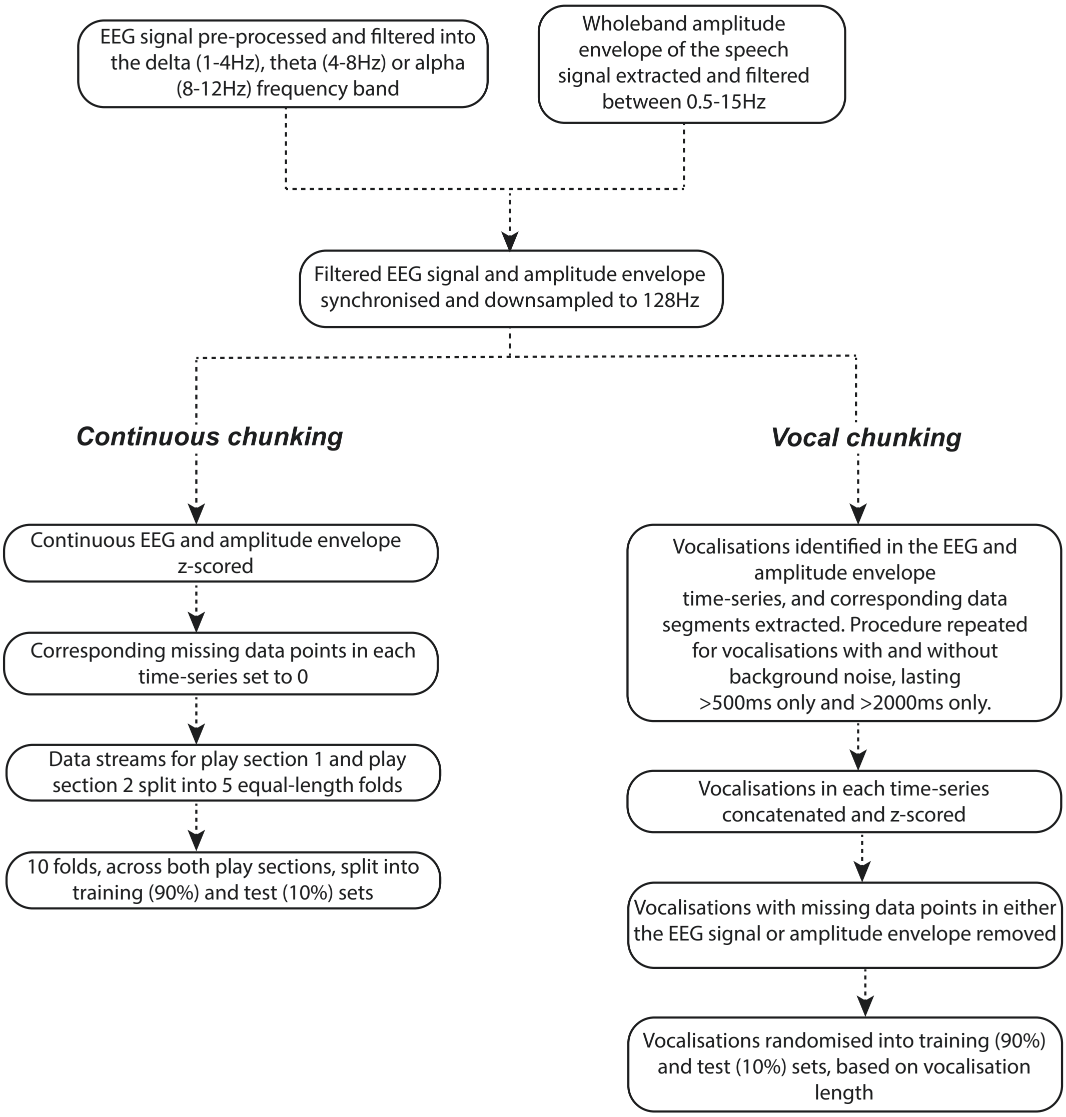

Figure 5. 2 Schematic representation of the pre-processing steps to split data into individual folds for model training and testing. Procedures for the continuous chunking analysis (left) and vocal chunking analysis (right) are depicted.

5.2.6.3 mTRF computation

mTRF analysis was carried out using the mTRF toolbox (Cross et al., 2016). The backwards mTRF model can be expressed by the following equation, where the TRF (g) represents the linear mapping from the neural response $r(\tau, n)$ back to the speech stimulus $s(t)$:

$$\acute{s}(t) = \sum_{n} \sum_{\tau} r(t + \tau, n) g(\tau, n)$$

where, $\acute{s}(t)$ is the estimated stimulus envelope, $g(\tau, n)$ is the TRF, and $r(\tau, n)$ is the EEG signal, over each time lag (τ), at each channel (n) included in the model. The TRF (g) is estimated using a method of regularized regression:

$$g = (R^T R + \lambda I)^{-1} R^T s$$

where R is the EEG data at each time lag, and s is the zero-lagged stimulus. $R^T R$ is the autocovariance matrix of the neural response which is divided out from the model. In dividing out the autocovariance of the EEG response from the model, inter-channel redundancies are no longer included in the model, and, as a result, each TRF weight at each channel represents the amount of information that weight provides for the re-construction (Crosse et al., 2021). Both individual and generic models were computed over time-lags from -50 to 250ms (Attaheri et al., 2022).

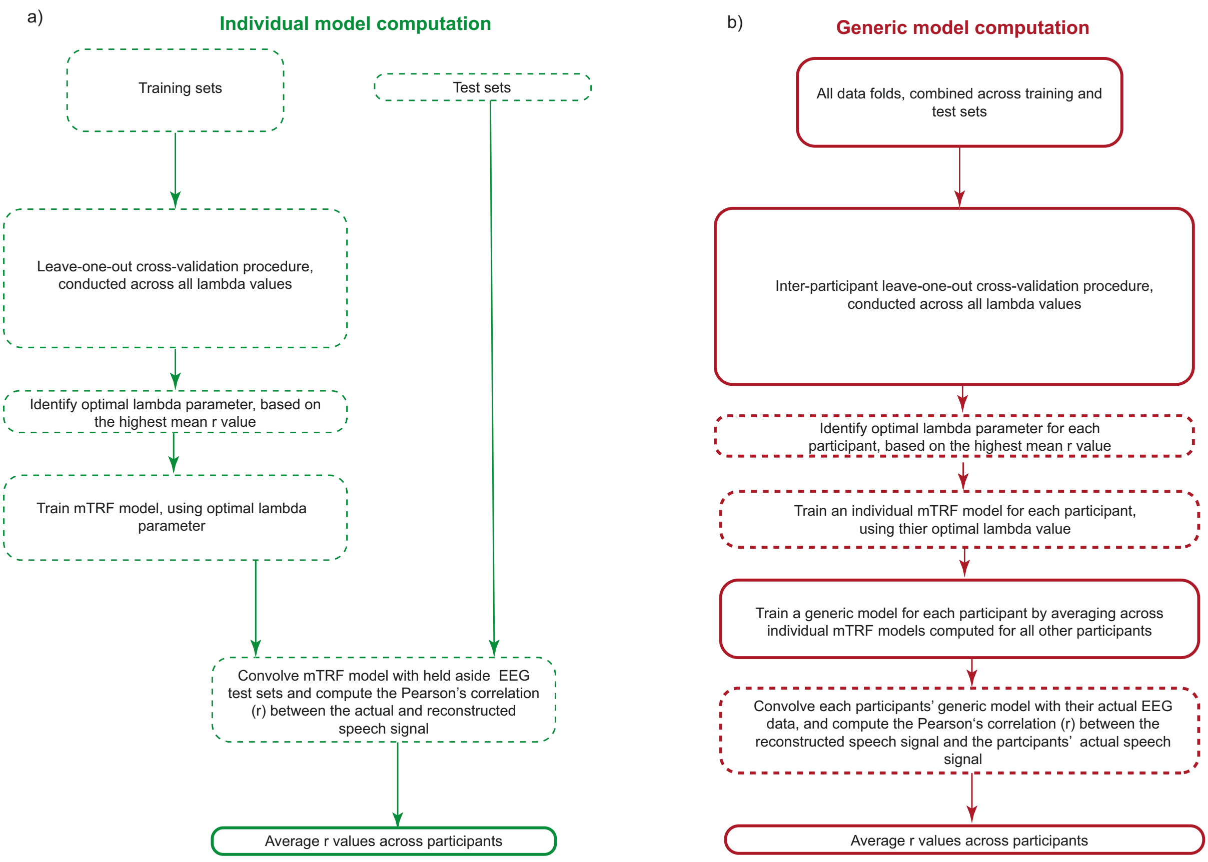

Figure 5. 3 Schematic representation of mTRF model computation. a) Individual model computation. b) Generic model computation. Dashed boxes indicate procedures conducted on the data of each individual participant; hard-line boxes indicate procedures conducted across the data of all participants. The proportional size of the boxes reflects relative computation time.

Individual model training and testing

For individual models, continuous folds or vocalisations were split into training and test sets, following procedures outlined in the previous section (see Figure 5.3 for a schematic depiction). First, to obtain the most optimal lambda (λ) parameter for the mTRF model for each individual participant, models were trained over λ values ranging from 10^{-7} to 10^{7} (Jessen et al., 2021), using a 'leave-one-out' cross-validation procedure. The cross-validation procedure was implemented using the *mTRFcrossval* function in the mTRF toolbox. Using the data allotted to the training sets only, for each lambda parameter (λ), a TRF was calculated for every trial, bar one (left-out) trial, and then averaged over training trials. The predicted speech signal of the left-out trial was then calculated by convolving the EEG response of that trial with the averaged TRF at each channel, which was then compared to the actual speech signal of the left-out trial using Pearson's correlation (r). The same procedure was iterated over the number of trials available for that participant and the resulting r value for each iteration averaged. This yielded a predictive power estimate (r) of the model for each lambda parameter (λ) tested.

The most optimal ridge parameter for that participant (i.e. the parameter yielding the model associated with greatest predictive power) was identified (Crosse et al., 2016) and used to train the model on the training set using *mTRFtrain.* This function computes an mTRF model for each training set and then averages across models. Finally, the model was tested using the data left aside in the test set, using *mTRFpredict*: the averaged training model was convolved with the neural response to reconstruct the speech stimulus, and the predictive accuracy between the actual speech signal and that predicted by the model computed via Pearson's correlation (r). The resulting r value was taken as the overall measure of speech-brain tracking for that participant. Model values were subsequently averaged across participants.

To compare predictive accuracy values during model training and testing, the predictive accuracy values for the lambda parameter yielding the best predictive accuracy for each participant during cross-validation were compared with the predictive accuracy scores obtained during model testing (Crosse et al., 2021)

Generic model training and testing

For the generic models, a model was trained and tested for each participant, by averaging over participant-specific mTRF models across all other participants (see Figure 5.3 for a schematic depiction). Whilst there are different approaches to training and testing procedures in computing generic models (Crosse et al., 2021; Di Liberto & Lalor, 2017), we follow the procedure outlined by Jessen et al. (2021).

First, to identify the best λ parameter for each participant (using the same parameters as the individual model), an inter-participant cross-validation procedure was conducted. For each λ value, for each participant, an mTRF model was estimated using the *mTRFtrain* function in the mTRF toolbox (see above for description). Then, for each participant, a generic model was computed by averaging over the models of all other participants. The predictive accuracy of the averaged mTRF model for the left-out participants' speech data was computed using the *mTRFpredict* function. This procedure was repeated for every participant at each λ value, and the λ term yielding the best predictive accuracy for each participant identified. Second, an mTRF model for each participant was computed using the *mTRFtrain* function, inputting their most optimal λ value. To train a generic model for each participant, the models resulting from the optimal λ terms were averaged across all other participants and the predictive accuracy of the averaged model tested on the target participants' speech data, using *mTRFpredict*. Resulting r values were subsequently averaged across participants.

5.2.6.4 Permutation testing

Continuous analysis

To create a random permutation distribution for the continuous models, each participants' synchronised and downsampled speech stream (see section 5.2.5.2**)** was randomly paired with another participants' EEG stream. Similar to the procedures outlined above, the longer of the two data streams was shortened so that both data streams were of the same length. Corresponding parts of the EEG and speech signals that were set as missing data points in one stream were identified in the other alternate stream and set to missing data points. Each stream was then z scored, and corresponding missing data points set to 0. Exactly the same procedure for computing the continuous folds, and allotting folds to training and test sets was then repeated in the same way as the main analysis.

A generic or individual mTRF model was then generated and the overall r value computed (see section 5.2.5.3). This same procedure was repeated 100 times, and r values averaged over the 100 iterations. To test whether the observed values in the actual data differed from chance, a t-test was computed between the observed values and the averaged permutation values, after outlier removal (Attaheri et al., 2022).

Vocal chunking analysis

To create a random permutation distribution, again, each participants' synchronised and downsampled speech stream was randomly paired with another participants' EEG stream. Here, each vocalisation, extracted from the participants' synchronised speech stream was paired with a segment of the randomly paired participants' EEG signal, that matched the length of the vocalisation. Again, where either the speech segment or EEG segment contained any missing data points, this vocalisation was excluded from analysis. To optimise the number of vocalisations included in the permutation distribution, where a segment of the randomly paired EEG data contained missing data points, the algorithm searched a further 3 times for a segment of data without any missing data values: this number was capped to deal with the fact that some caregivers had some vocalisations that lasted over 1 minute. The speech and EEG segments were then z scored and allotted to training and test sets, following exactly same procedures outlined above.

Again, a generic or individual mTRF model was then generated and the overall r value computed. This same procedure was repeated 100 times, and r values averaged over the 100 iterations. To test whether the observed values in the actual data differed from chance a t-test was computed between the observed values and the averaged permutation values, after outlier removal (Attaheri et al., 2022).

5.3 Results

The results section is organised in 3 sections. First, we present descriptive statistics relating to the temporal characteristics of the caregivers' speech. Second, we present the results of the individual mTRFs, before the results of the generic mTRF models are described.

5.3.1 Descriptives

First, we examined the temporal characteristics of caregiver vocal behaviours during the naturalistic interactions with their infants. The results of these analyses are presented in Figure 5.4. Figure 5.4a shows the proportion of time that caregivers and infants vocalised on their own, as well as the proportion of time they vocalised at the same time (co-vocalisations). As mentioned in the Methods section, due to the automatic decoding algorithm used to identify periods that the caregivers and infants were speaking, co-vocalisations were not divided into sections of caregiver and infant speech. Inspection of Figure 5.4b and 5.4c shows that the number of vocalisations lasting over 2000ms (M=23.70, SD=26.08) was much lower compared to those lasting over 500ms (M=59.12, SD=26.08), and the number of clean vocalisations were even lower per participant dyad: vocalisations lasting over 500ms (M=16.43, SD=11.02); vocalisations lasting over 2000ms (M=4.78, SD=4.66). Given the particularly low number of vocalisations that were clean and lasting over 2000ms, analyses including these vocalisations only were not conducted. For all vocalisations over 2000s, 2 participants were removed from analyses because they contributed fewer than 5 vocalisations. For clean vocalisations over 500ms, 7 participants were excluded from analyses for the same reason.

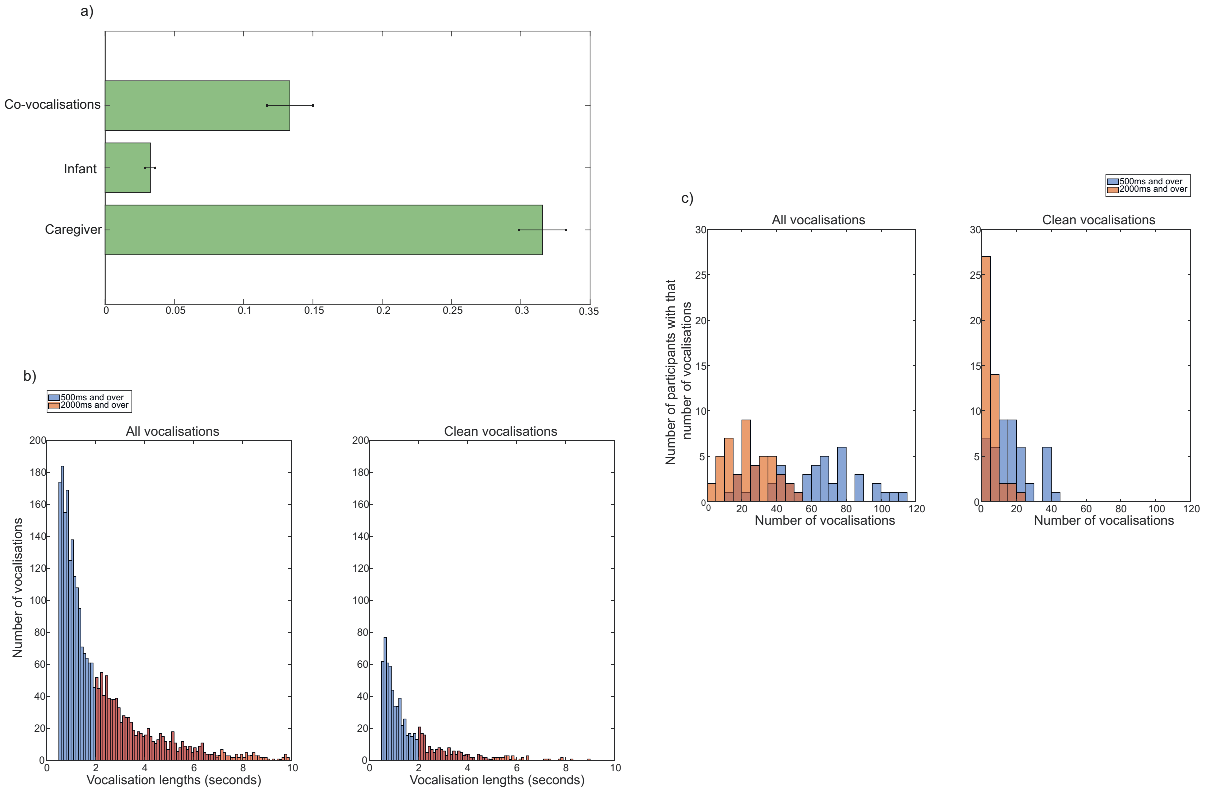

Figure 5. 4 Temporal characteristics of caregiver vocal behaviours. a) Proportion of time caregivers and infants spent vocalising, and co-vocalising. b) Histogram of the lengths of vocalisations for all vocalisations and clean vocalisations only, over 500ms (blue) and 2000ms (red), after outlier removal. c) Number of participants with a certain number of vocalisations for all vocalisations and clean vocalisations only lasting over 500ms (blue) and over 2000ms (orange).

5.3.2 Individual mTRF models

In this section, we outline the results of the individual mTRF models obtained for each frequency band, and consider the findings relative indexes of how well the trained model transfers to new data (see Methods for further description).

Continuous individual models

The results of the continuous individual models are presented in Figure 5.5. The number of outliers removed for each analysis are presented in Table 5.1. Figure 5.5a shows the mean predictive accuracy scores across participants for each frequency band. A series of two-tailed paired sample t-tests indicated that only the predictive accuracy values of the delta rate model were significantly higher than the values obtained in the permutation distribution (see Table 5.1). Testing whether the models differed to chance with Wilcoxon Signed Ranks tests, before outlier removal, yielded a similar pattern of results, with predictive accuracy values significantly above chance for the delta model only (Figure S1).

Across the different frequency bands, the lengths of the folds included in the training and test sets are similar, and these lengths are consistent across frequency bands (Figure 5.5b). Inspection of the error bars (Figure 5.5b) also reveals little variability between participants, suggesting that the models for each participant were trained and tested on continuous folds of a similar length. The predictive accuracy values computed in model testing, were, however, lower compared to those obtained during cross-validation, for all frequency bands (Figure 5.5c), which could be an indication that the models are over-fit during model training and therefore do not transfer well to the model at test. That said, for all frequencies, the difference in predictive accuracy values is small.

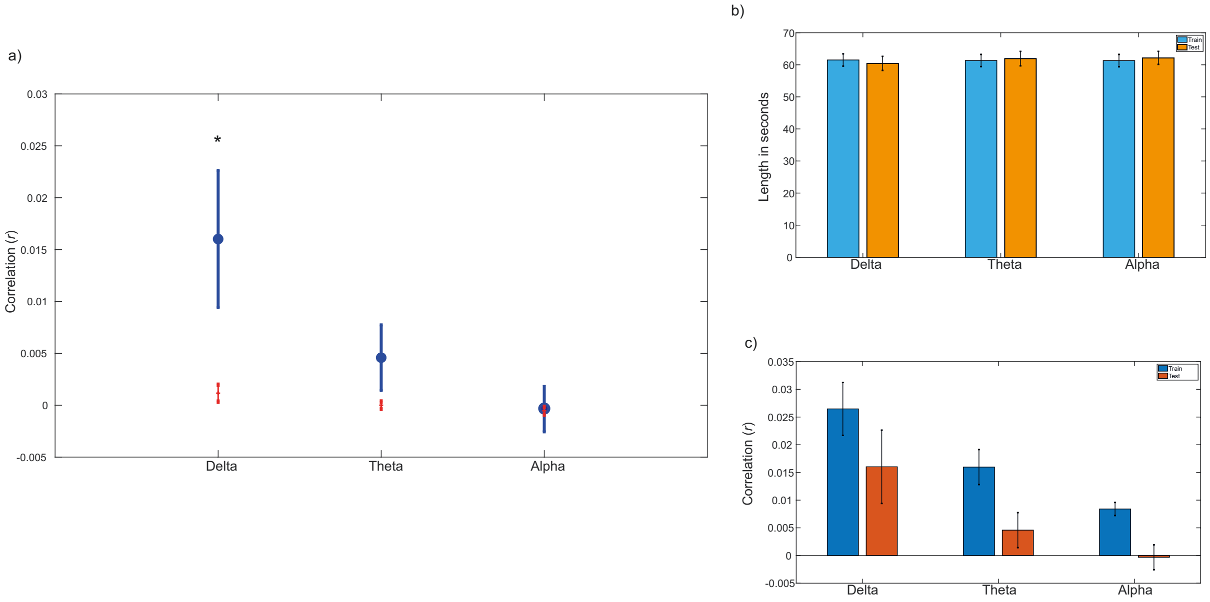

Figure 5. 5 Continuous individual models. a) Grand average predicative accuracy values for each frequency band (Pearson's r). Blue circles show the mean across participants for each frequency band. Blue lines indicate the SEM of the averaged predictive accuracy values. Red horizontal lines show the averaged permutation values, averaged across permutations and participants. Red vertical lines indicate the SEM. Two-tailed paired sample tests compared the observed r values to chance (*p<0.05). b) Lengths of the folds used in training and testing for each frequency band. Blue bars show the mean lengths of the training sets, and orange bars show the mean lengths of the test sets. Black lines show the SEM. c) Averaged predictive accuracy for training and testing sets. Blue bars show the training sets and orange bars show the test sets. Black lines indicate the SEM.

Vocalisation individual models

Next, we examined the results of the individual mTRF models computed with individual vocalisations serving as the model folds. First, we tested the performance of individual models, trained and tested on vocalisations lasting 500ms or longer. The results of this analysis are shown in Figure 5.6. Inspection of Figure 5.6a shows that, across frequencies, the lengths of the training and test sets were similar, and there was little variability in this across participants. Comparing the predictive accuracy of the models during training and testing (Figure 5.6b) reveals, that, again, predictive accuracies were much lower for the testing sets compared to the training sets. Relative to Figure 5.5b, where the data for the continuous individual models is presented, this difference is much greater, and the predictive accuracy (r) values for both training and testing sets are much higher in comparison to those observed where continuous data chunks are used. The results of the individual mTRF model are presented in Figure 5.6c. A series of paired sample t-tests revealed that mean predictive accuracy values did not fall significantly above their permutation distribution at any frequency band examined (see Table 5.1 for the results of the t-tests and information on outlier removal).

Next, we examined the performance of the individual models conducted with vocalisations lasting 2000s or longer. As mentioned in the Descriptives section (Figure 5.4b), excluding vocalisations between 500-2000ms led to the inclusion of much fewer vocalisations in the mTRF model. Corresponding to this, the lengths of the vocalisations included in the training and test sets are much longer, compared to the analysis using all vocalisations over 500ms (Figure 5.6d). Comparing the predictive accuracy values in the training and test sets indicated that predictive accuracies at test were substantially lower than those computed during model training (Figure 5.6e). Though the predictive accuracies, across the models at different frequencies, showed a similar pattern to those obtained in the continuous analysis, correlation values were much lower, and no values were significantly above chance (see Table 5.1). Finally, we examined models for clean vocalisations lasting 500ms or longer (Figure 5.6, bottom row), including even fewer vocalisation again (see section 5.3.1). Here the difference in predictive accuracy values between training and testing values is very large, and the predictive accuracy values at test are uninterpretable; falling significantly below the permutation distribution for each frequency examined (Table 5.1).

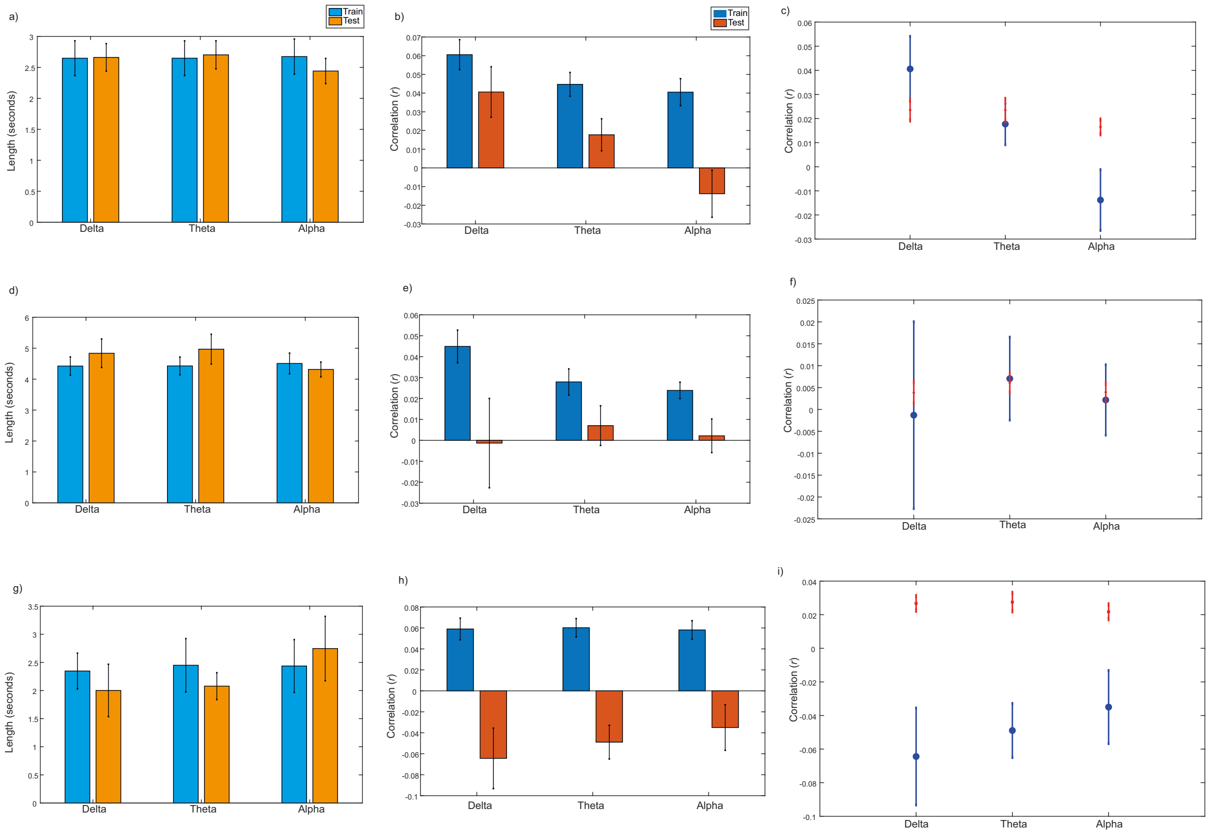

Figure 5. 6 Vocal chunk individual models. First column shows the lengths of the vocalisations included in the training and tests sets for a) vocalisations over 500ms, d) vocalisations over 2000ms, g) clean vocalisations over 500ms. Blue bars show the mean for the training sets and orange show the mean for the tests sets. Black lines indicate the SEM. Second column shows the averaged predictive accuracy for training and testing sets for b) vocalisations over 500ms, e) vocalisations over 2000ms, h) clean vocalisations over 500ms. Blue bars show the training sets and orange bars show the testing sets. Black lines indicate the SEM. Last column shows the grand average predicative accuracy values for each frequency band (Pearson's r), for c) vocalisations over 500ms, f) vocalisations over 2000ms, i) clean vocalisations over 500ms. Blue circles show the mean across participants for each frequency band. Blue lines indicate the SEM of the averaged predictive accuracy values. Red horizontal lines show the averaged permutation values, averaged across permutations and

participants. Red vertical lines indicate the SEM. Two-tailed paired sample tests compared the observed r values to chance (*p<0.05).

	df	t-statistic	*p-values*	Outliers
Continuous				
Delta	43	2.09	0.043	2
Theta	44	1.40	0.170	1
Alpha	41	0.09	0.930	4
Vocalisations over 500ms				
Delta	43	1.31	0.521	2
Theta	42	0.65	0.521	3
Alpha	42	2.23	0.03	3
Vocalisations over 2000ms				
Delta	41	0.24	0.814	2
Theta	41	0.10	0.920	2
Alpha	41	0.20	0.839	2
Clean vocalisations over 500ms				
Delta	37	3.07	0.004	1
Theta	36	4.41	<0.001	2
Alpha	33	2.36	0.0240	5

Table 5.1 Individual model results of the paired sample t-tests of the difference between observed predictive accuracy scores and the permutation distribution for each frequency band, for each analysis.

5.3.3 Generic mTRF models

In this section, we examine the predictive accuracy of the mTRF models yielded by the subject-independent generic models.

Continuous generic models

The results of the continuous generic models are presented in Figure 5.7a. Similar to the individual models derived from the continuous data, inspection of Figure 5.7a indicates higher predictive accuracy values at delta and theta frequencies, compared to alpha frequencies. Corresponding to this, a series of paired sample t-tests revealed that only the r value of the theta frequency model fell significantly above chance (see Table 5.2 for the results of the t-tests and information on outlier removal). Of note, however, and in line with previous research with infants using a generic modelling approach (Jessen et al., 2019), the predictive accuracies of the models at all frequency bands are low, and much lower in comparison to the individual model (Figure 5.7a). Comparing the observed values to the permutation distribution using Wilcoxon signed ranks tests, before outlier removal, revealed significant speech-brain tracking at delta and theta frequencies (Figure S2).

Vocalisation generic models

The generic models conducted with vocalisations lasting over 500ms showed a similar pattern: with higher predictive accuracies at theta and delta frequencies, though their predictive accuracies were much lower and more variable (Figure 5.7b). None of the models at any of the frequency bands investigated differed significantly from the permutation distribution (Figure 5.7b). Inspection of Figure 5.7c shows that where models included vocalisations only lasting over 2000ms, predictive accuracies were even lower, with even greater variability, and this was the same for models conducted with clean vocalisations lasting over 500ms (Figure 5.7d). Again, no r values for these models differed significantly to the permutation distribution (for further information on statistical testing see Table 5.2).

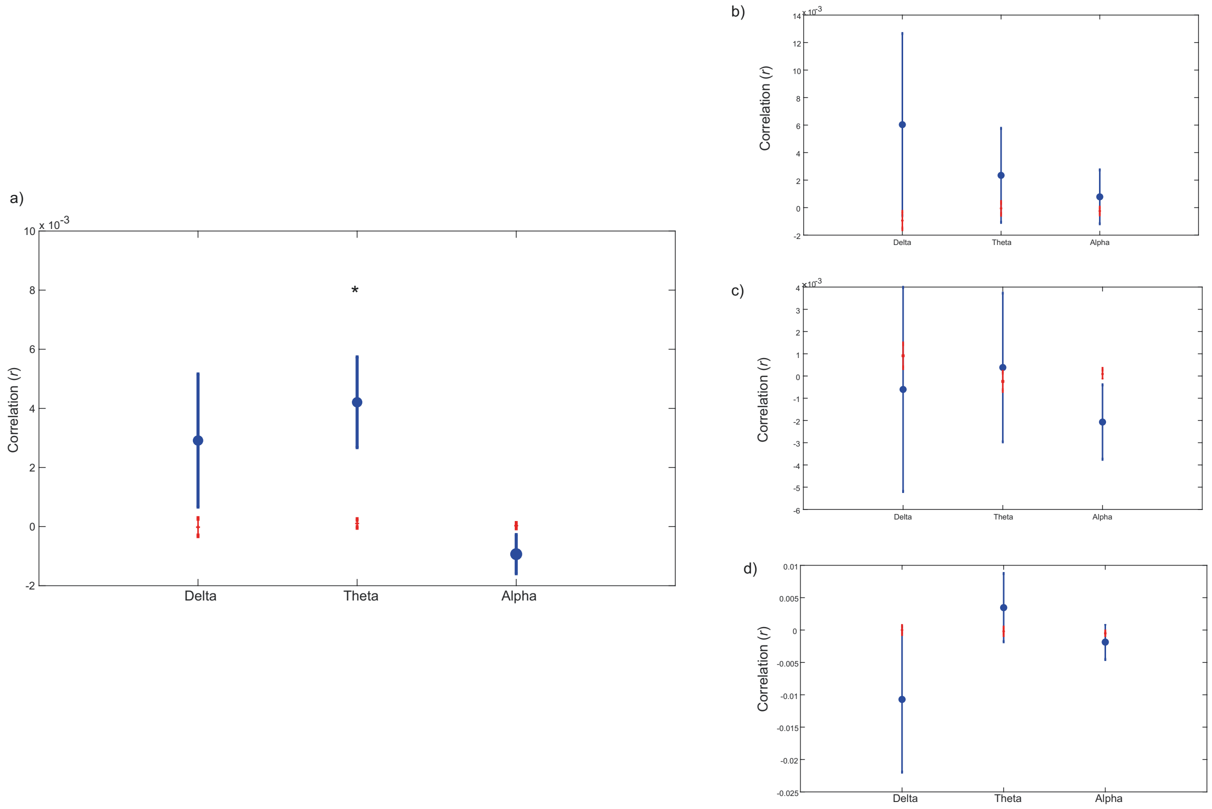

Figure 5. 7 Continuous and vocal chunk generic models. Grand average predicative accuracy values for each frequency band (Pearson's r) for a) continuous data, b) vocalisations over 500ms, c) vocalisations over 2000ms and d) clean vocalisations over 500ms. Blue circles show the mean across participants for each frequency band. Blue lines indicate the SEM of the averaged predictive accuracy values. Red horizontal lines show the averaged permutation values, averaged across permutations and participants. Red vertical lines indicate

the SEM. Two-tailed paired sample tests compared the observed r values to chance (*p<0.05).

	df	t-statistic	*p-values*	Outliers
Continuous				
Delta	43	1.32	0.195	2
Theta	41	2.46	0.018	4
Alpha	40	1.35	0.184	5
Vocalisations over 500ms				
Delta	45	1.04	0.305	0
Theta	45	0.66	0.511	0
Alpha	44	0.51	0.614	1
Vocalisations over 2000ms				
Delta	41	0.31	0.758	2
Theta	42	0.18	0.857	1
Alpha	40	1.20	0.236	3
Clean Vocalisations over 500ms				
Delta	37	0.93	0.360	1
Theta	38	0.70	0.485	0
Alpha	34	0.47	0.638	4

Table 5. 2 Generic model results of the paired sample t-tests of the difference between observed predictive accuracy scores and the permutation distribution for each frequency band, for each analysis.

5.4 Discussion

In this study we examined the applicability of continuous methods of analysis for assessing speech-brain tracking during naturalistic caregiver-infant table-top play. Whilst these methods are increasingly being used in adult and infant populations to assess speech-brain tracking relative to pre-recorded stimuli, no previous work has applied these methods to continuous, free-flowing interactions. Investigating neural tracking of the speech signal in naturalistic contexts is crucial to our understanding of how infants process the phonological information of speech during everyday communication (Attaheri et al., 2022; Leong, Kalashnikova, et al., 2017), as well the association of these micro-processes to the ongoing behavioural dynamics of the interaction. We conducted analyses similar to those outlined in Attaheri et al. (2022), employing a backwards modelling mTRF approach, that constructs a model from the infants' continuous neural response back to the speech signal. To maximise the amount of data included in the models, we first conducted training and testing procedures on continuous folds of the interaction, and then on specific vocalisations of certain lengths. To train the models, both individual and generic methods were employed.

For the individual models, where continuous folds were entered into the model, predictive accuracy values were greater during model training compared to testing, but this difference was small. This suggests that where individual mTRF models are trained and tested on continuous periods of data obtained by dividing the interactions up into folds of equal lengths, model over-fitting to the training data is low (Crosse et al., 2021). Given that the models were trained on large quantities of data (~10 minutes per dyad; see Figure 5.5), it is likely that the higher values obtained in model training are the result of the noisiness of the EEG data and the speech signal, where, with the continuous analysis, all background noise (toy clacks and infant vocalisations) are included in the caregiver's amplitude envelope. Corresponding, to the findings of Attaheri et al. (2022), as well as predictions based on the computational modelling of IDS (Leong & Goswami, 2015), significant speech-brain tracking was observed at delta frequencies, but not at theta or alpha frequencies.

In the models that were trained and tested on individual vocalisation chunks of 500ms or more, the predictive accuracy values showed the same overall pattern but none of the values fell significantly above chance. In comparison to the continuous models, the difference in the predictive accuracy values acquired in model training and testing was much greater (Figure 5. 6), suggesting that model overfitting to the training data was more problematic in this analysis compared to the continuous training method. Also of note, the predictive accuracy values for these models are much higher in comparison to the models trained on the continuous data sets, and unlike the continuous models, the permutation distributions fall above 0. For models constructed on vocalisation chunks over 2000ms and clean vocalisations over 500ms, where far fewer vocalisations were included in model training, comparison of the training and test values indicated large differences, and predictive accuracies were very low, and wholly uninterpretable for the clean vocalisations (Figure 5.6).

Overall, the results of the individual models suggest that, with a subject-dependent training procedure, most reliable speech tracking values are obtained where the interaction is divided up into continuous folds, and, therefore, where the greatest amount of data is used in model training.

Given the issues of data quality and quantity inherent to naturalistic data recordings, in the second part of this paper, we computed mTRF models using generic training procedures, according to Jessen et al. (2019), which train models by utilising all the available data for each participant. Whilst generic models reduce the likelihood of overfitting with smaller and more noisy data sets, the predictive power of these models is often lower in being trained across participants, with data sets of variable quality (Jessen et al., 2021). The results of the generic models showed that, in both types of training procedures (continuous inputs and vocalisation chunks of variable lengths) the predictive accuracy values were extremely low and much more variable in comparison to previous studies of speech-brain tracking conducted with infants (Jessen et al., 2019; Jessen et al., 2021). Significant speech-brain tracking was observed at theta frequencies where the models were trained on continuous data folds, and the delta model fell significantly above the permutation distribution before outlier removal (Figure S2). Given the extremely low predictive accuracy of the models, however, the mTRF models are unlikely to be an accurate representation of the true neural response (Crosse et al., 2021).

The particularly low and variable predictive accuracy values obtained with the generic models are likely driven by the fact that, in our naturalistic interactions, infants were interacting with their caregiver and therefore listening to different speech signals: comparative to controlled paradigms where each infant listens to the same stimulus. This will drive variation in model fitting in two ways: the amount of noise in the speech signal is likely to vary substantially across participants, and aspects of the noise and characteristics of the speech signal could also drive differences in how infants track the amplitude envelope. Furthermore, considering infant-centred factors, whereas in controlled paradigms infants are often seated on their caregiver’s lap and encouraged to stay still as much as possible throughout the recording, in naturalistic interactions infants are free-moving, meaning that the amount of movement and, therefore, the amount of movement-related artefact affecting the EEG signal will also vary substantially across participants (Marriott Haresign et al., 2021).

Overall, the results of the individual and generic models indicate that the use of individual training procedures, conducted on continuous segments of the interaction may be most optimal to examining speech-brain tracking during naturalistic social interactions. The difference between the predicative accuracy scores for training and test sets was smallest for the continuous data models, compared to the vocal chunk models, and variability in predictive accuracy values was also low. Computing models using generic training procedures led to particularly low and likely uninterpretable speech-tracking values.

The results of the continuous individual models are consistent with the prediction that delta rate modulations play a particularly important role in facilitating infants' early ability to parse the speech signal (Leong, Kalashnikova, et al., 2017; Leong & Goswami, 2015). Leong et al., (2017), for example, showed that rather than containing most power at theta rate modulations (as in the case of ADS), IDS is associated with peak modulation frequencies at delta-rate modulations, with the peak frequency increasing from 7-11 months. The results of the individual models also correspond to Attaheri et al.'s (2022) findings, where speech-brain tracking was found to be significantly above chance at both delta and theta frequencies but not alpha frequencies: and tracking at delta rate frequencies was significantly higher in comparison to theta frequencies at all 3 ages examined. That we find speech-brain tracking to delta frequencies only could be a result of the use of nursery rhyme stimuli in Attaheri's experiment which would be inherently more rhythmic compared to the spoken IDS of the caregivers in our study, likely resulting in a higher proportion of stressed syllables (Leong, Kalashnikova, et al., 2017; Leong & Goswami, 2015).

Given that the inputs to the mTRF model with continuous analysis procedures contained frequent segments of noise unrelated to the speech of the caregiver, including toy clacks and infant vocalisations, it will be important to conduct further analyses to support that the neural tracking by infants to delta-rate frequencies is driven by the infant tracking the amplitude modulations in the caregiver's speech. The mTRF modelling approach is particularly robust to the inclusion of missing data points (even up to 30-40%; Jessen et al., 2021). One option would therefore be to insert 0 values at points in the amplitude envelope that the caregivers are not speaking. That said, inspection of Figure 5.4a reveals that, in our data sets, caregivers vocalised on their own for an average of 35% of the interaction, meaning that, even with the robustness of the mTRF models to missing data points, this analysis would not be possible. To examine this in more detail, therefore, in future work, it will be important to apply mTRF modelling procedures to naturalistic interactions where the caregiver speaks more often and the interaction is less affected background noise (e.g. caregiver puppet shows with their infant). Another particularly important avenue for future work examining speech-brain tracking in naturalistic contexts is in testing the effects of different aspects of signal noise in the performance of the mTRF analysis, with infant populations .

Overall, this methods paper has demonstrated the applicability of continuous modelling approaches to examining speech-brain entrainment by infants to their own caregivers' speech signal during naturalistic, free-flowing interactions. The results are important in showing that the finding of significant speech-brain tracking in early infancy to controlled, often purposefully rhythmic stimuli, in experimental paradigms transfers to everyday, naturalistic interactions. Developing these methods further will open up new avenues for investigating how speech-brain tracking by infants associates with the joint attention processes of real-time interactions (e.g. mutual gaze and shared attention towards objects), as well as the association of speech-brain tracking with the timing of caregiver object labelling, and the complexity of their inputs (Nencheva & Lew-Williams, 2022).

CHAPTER 6 – General discussion

6.1 Summary and integration of findings

Taking a multi-method approach to the study of early caregiver-infant interactions, this thesis aimed to examine the sub-second processes that drive infant attention during joint interactions with their caregiver, considering endogenous cognitive processes intrinsic to the infant, as well the interactive contingencies generated through dynamic inter-personal process. To do this, infants' neural oscillatory activity was recorded during online social interactions with their caregivers. Using time-locked and continuous methods of analysis, this thesis advances our current understanding of the fast-acting processes that support the coordination of infant attention with a social partner, with fundamental implications for how we conceptualise the learning mechanisms involved in the development of intentionally mediated cognitive process and their role in driving early language acquisition.

In Chapter 2, two approaches to understanding the processes that drive infant attention during naturalistic shared interactions were outlined. Traditional and active learning socio-pragmatic approaches had previously emphasised that, towards the end of the first year, infants are proactive in creating episodes of shared attention with a social partner (Tomasello et al., 2007). In this perspective, infant attention is thought to be guided by intentionally mediated communication: infants begin to understand that other peoples' behaviour is intentional and that they can influence the internal representations of others through modulations in their own intentionally guided behaviour (Carpenter et al., 1998; Donnellan et al., 2020). This perspective was contrasted with dynamic systems approaches to understanding infant attention during early interactions, where fast-acting sensorimotor contingencies are thought to drive the moment-to-moment development and maintenance of shared perceptions and cognitions (L. B. Smith, 2005; Yu & Smith, 2017). A series of studies conducted by Yu and Smith were outlined, where micro-behavioural analytical techniques were employed to examine the fast-acting sensorimotor contingencies that drive behavioural coordination in early triadic interactions. This work showed that, rather than active attention sharing behaviours, coordinated attention was most often achieved and perceived by infants through attending towards their partners' sensorimotor cues (Franchak et al., 2011; Yu & Smith, 2013, 2017).

It was highlighted, however, that, given similar behaviours can occur across different levels of attentional and intentional engagement, to understand the endogenous cognitive mechanisms and interactive contingencies that drive infant attention more fully, we need to apply neurocognitive techniques to the study of early inter-personal coordination. The rest of the chapters in this thesis subsequently combined event-locked and continuous methods of analysis to examine how sub-second changes in infants' endogenous neural activity associated with intra and inter-dyadic behavioural process.

In Chapter 3, we took a multi-method approach to examine whether, consistent with the socio-pragmatic perspective, infants showed patterns of neural and behavioural activity indicative of proactive attention-sharing in naturalistic, free-flowing interactions with their caregiver. Event-locked methods of analysis were employed to examine the micro-second fluctuations in infant neural activity 2000ms before and 2000ms after infant looks towards objects that resulted in mutual attention (where the caregiver joined the infants' look towards an object) and nonmutual attention (where the caregiver did not join the infants' attentional focus). Results indicated that infants, aged 9-12 months, were not routinely proactive in creating episodes of shared attention with their partner: in the 2000ms before leading their partners' attention, infant EEG activity at theta frequencies did not increase and infants did not increase their use of ostensive cues. This finding is line with previous reports by Yu and Smith, where the gaze patterns of caregivers and infants were recorded with head-mounted eye trackers (Yu & Smith, 2013, 2017).

Analysing the time-period after look onset, however, revealed that infants were nevertheless sensitive to moments that their adult partner followed their gaze towards an object: infants' EEG activity showed a significant decrease at alpha frequencies in the time after infant-led looks to mutual attention, compared to adult-led looks. Based on previous reports of the association between alpha desynchronisation and predictive processing in early infancy (C. Monroy et al., 2021; Rayson et al., 2019), this finding was interpreted to suggest that infants predict and encode the behavioural contingency of their partner in the time after they lead a look towards an object and the partner contingently followed that look. Interestingly, infant ostensive signals in the time after look onset showed no differences between looks: suggesting that the behavioural contingency of the partner was not realised by the infant through attending towards their partners' intentionally communicative behaviours.

The findings of Chapter 3 highlight the importance of examining fluctuations in neurocognitive process that are otherwise unobservable behaviourally. Combining the results of the neural and behavioural analyses, the findings suggest that infant attention is largely *reactive* to the sensorimotor behaviours of their partner: the similarity in neural and behavioural activity before infant- and adult-led look onset suggested that modulations in the sensorimotor cues of their partner might guide infant attention at moments they both lead or follow their partners' attention towards an object. It was postulated that the same interactive sensorimotor process might also be the mechanisms that signal behavioural contingency to the infant in the time after they lead their partners' attention towards an object. Findings were interpreted relative to embodied associative learning perspectives of the development of intentionally mediated communication in early infancy. These accounts assert that consistent and contingent feedback to infant behaviour gives meaning to infant behaviours, forming the basis for infant representations about the intentions of others, and how their own intentionally-motivated behaviours affect those of their partner (L. B. Smith & Breazeal, 2007).

Most important then, the findings of Chapter 3 indicate that, although infants are not engaging in active attention-sharing behaviours (Yu & Smith, 2013), they are, nevertheless, making online predictions relative to the inter-dynamic action-generated contingencies of the interaction. In doing so, they point towards the importance of cross-dyadic sensorimotor learning mechanisms in driving the development of intentionally guided communication, which have largely been omitted from traditional socio-pragmatic accounts of the development of early joint attention (Tomasello et al., 2007). It will be important for future work to examine longitudinal associations between infants' early sensitivity to the contingent responsivity of their caregiver and the development of intentionally produced communicative behaviours. Modelling how associations between infants' endogenous cognitive processing and their attention develop even over the course of an interaction, as a function of the responsivity of the caregiver could also be an important next step. For example, Goldstein and Shwade (2008) have shown that, where caregivers consistently and contingently respond to their 9-month olds' vocalisations, infants increase the semantic complexity of their vocalisations over the duration of an interactive play episode, matching the phonological complexity of the speech inputs by the caregiver. Similar processes could operate relative to moments that infants engage in episodes of attention-sharing with their caregiver, with repeated and contingent feedback to infant-led attention associating with immediate and long-term increases in infants' control over their own attention and behaviour (Oudeyer & Smith, 2016).

Chapter 4 built on these findings; testing dynamic associations between endogenous cognitive processes and exogenous interactive contingencies in driving the duration of infant attention episodes. Here, continuous methods of analysis were employed to examine the intra- and inter-individual processes that structured the timing of infant attention: testing, in particular, modulations in caregivers' vocal saliency as a sensory-motor cue driving infant attention. At the intra-individual level, cross-correlation analyses showed that both fluctuations in infant attentiveness (indexed by the durations of infant attention episodes) forward-predicted longer attention durations, and, to a greater extent, increases in look durations forward-predicted modulations in their theta activity. Against the interpretation of the findings from Chapter 3, that infant attention is largely *reactive* to the sensorimotor behaviours of their partner, however, Chapter 4 showed that infant attention durations were not forwards-predicted by modulations in the adult partners' behaviours. Neither change in caregiver gaze, nor the temporal or spectral properties of their voice associated with modulations in infant attentiveness. Instead, changes in caregiver behaviours were largely reactive to fluctuations in infant attention, and associated with the length of infant attention episodes: caregivers increased the rate of change in their gaze behaviour in the immediate time after infants shifted their attention towards a new object, and both the durations of caregiver looks, and modulations in their vocal saliency predicted the length of infant attention towards objects.

Taken together, the findings of Chapters 3 and 4 suggest that, rather than considering forwards-predictive associations between intra-individual process and infant attention, we need to consider the micro-processes that operate over the duration of infant attention episodes and, in particular, how these processes operate across the dyad (Richards & Anderson, 2004; Wass, 2021b). In Chapter 3, results indicated that infants predicted and encoded the behavioural contingency of their partner where they led a look that resulted in an episode of mutual attention. The findings of Chapter 4 further suggest that dynamic and reactive change in caregiver behaviours in response to re-orientations in infant attention drive increases in infant attentiveness. But processes operating in the time before an attentional reorientation by the infant do not distinguish its form (Chapter 3), or predict infant attentiveness (Chapter 4).

As outlined in the Introduction section to Chapter 4, previous work examining infant attention whilst they observe dynamic screen-based stimuli has shown that infant attention can largely be characterised by patterns of attentional inertia: the finding that, the longer a look lasts, the less likely that look is to end (Richards & Anderson, 2004). It is thought that this pattern is driven by endogenous amplificatory mechanisms that operate between the infant and their environment: decreases in distractibility drive increases in engagement which drive further decreases in distractibility, decreasing the likelihood that the look will end (Richards, 2011). In naturalistic interactions similar processes could operate across the dyad, where modulations in caregiver behaviours in response to changes in infant attention, associate with increases in infant engagement, which drive further modulations in caregiver behaviours, and, possibly, other, fast-acting processes of inter-dyadic coordination (Wass, 2021a). In Chapter 4 caregivers were found to upregulate their behaviours in the time immediately following infant looks towards objects, but, over the course of longer looks, down-regulated their gaze behaviour, as well as the rate of change in the F0 of their voice.

Interpreting these results in the context of the cross-correlation between infant attention durations and theta activity shown in Chapter 4 has potentially important implications for our understanding of the pattern and function of the endogenous processes that mediate infant behaviour in shared interaction, and how these processes are emergent from the dynamic sensorimotor contingencies of inter-personal coordination (Wass et al., 2018b; Yu & Smith, 2017). In Chapter 4, the cross-correlation between infant attention durations and theta activity showed that increases in infant attentiveness largely forward-predicted increases in infant theta activity. Combined with the findings of reactive change in caregiver behaviour to infant attention episodes, these results could suggest that contingent responding by the caregiver to re-orientations in infant attention, dynamically increase infants' endogenous control over their attention over the duration of a sustained attention episode, and over the course of successive attention episodes.

In future work it will be important to examine, in more detail, the fast-acting sensorimotor processes that operate over the course of infant look durations to maintain infant attention, and how these processes associate with infants' endogenous neural activity. In particular, given the finding that infant attention is most often directed towards objects in shared interactions, and use the position of the caregivers' hand to coordinate their attention with their partners' (Yu & Smith, 2012, 2013, 2017), the dynamic hand and body movement of the caregiver during infant attention episodes should be a key focus. As mentioned in the introduction, variability in the hand actions of video-recorded stimuli has been shown to increase infant engagement (M. Meyer et al., 2022). Examining the relationship of these fast-acting cues to infant neural activity over the duration of a look will be particularly important. Indeed, through continuous modelling procedures, Chapter 5 shows that it is possible to measure speech-brain entrainment by the infant to amplitude modulations in the caregivers' speech signal. This finding opens up new avenues to examining the role of brain-behaviour entrainment in maintaining infant attention episodes, which could also be examined across different levels of fast-changing behavioural process, such as oscillatory structures in the caregivers' hand and body movements (Jessen et al., 2019).

Given, however, that, as shown in Chapter 4, infant attention patterns are characterised by longer attention episodes interspersed by shorter looks to their partner and objects; as well as processes operating within an attention episode, we need to examine the emergent patterns of behavioural contingency that characterise the organisation of infant attention and how this associates with fluctuations in infants' endogenous cognitive process. As a result of the focus of the intra-individual and uni-modal associations with infant theta activity in Chapter 4, this was not examined in the present thesis. In line with embodied accounts of the active learning mechanisms that operate across the dyad in early infancy, consistent and contingent responding by the caregiver could lead to in-the moment increases in infants' endogenous control over their own attention (Oudeyer & Smith, 2016). This could, in turn promote further active sampling strategies and effortfully controlled behaviours by the infant to maximise their opportunities for learning through their attunment to the low-level sensorimotor contingencies of the interaction (Kidd et al., 2012; McQuillan et al., 2020).

Chapter 5 assessed the applicability of continuous methods of analysis to examining speech-brain tracking during naturalistic infant-caregiver interactions. In Chapter 2, it was outlined that examining entrainment by the infant to amplitude modulations in the caregivers' speech is important to understanding infants' sensitivity to the phonological information in speech, and how this process operates relative to aspects of inter-dyadic coordination and caregiver behavioural inputs (Leong, Byrne, Clackson, Harte, et al., 2017; Yu & Smith, 2013). Our results indicated that, with continuous methods of data segmentation, which optimised the amount of data inputted to the model, it was possible to examine speech-brain tracking using subject-dependent modelling procedures. The overall results suggested, consistent with theories of early phonological processing, that speech-brain tracking was significantly above chance at delta frequencies only: tracking at theta and alpha frequencies did not differ to their permutation distributions. As mentioned above, the ability to examine fine-grained entrainment by the infant to oscillatory structures in the caregiver's behaviour will be crucial to examining, in more detail, the patterns of behavioural contingencies that drive fluctuations in infants own endogenously mediated attention and behaviour, and that support online prediction by infants during the interaction (K. J. Friston, 2019; Hamilton, 2021; Rayson et al., 2019). How these processes interact to support the development of intentionally mediated behaviour and language acquisition should be a key focus for future research. Below, we outline the implications of the findings of all three empirical chapters to current theories of early language development.

6.2 Implications for theories of early language development

In Chapter 2, the emphasis of the socio-pragmatic model on the development of intentionally-mediated understanding in driving early language learning was outlined (Lieven, 2016). This account argues that it is only once infants understand that a social partners' behaviours are intentionally mediated, and they can engage in the co-creation of shared intentional representations, that they learn from the referential and pragmatic aspects of communication

(Tomasello et al., 2007). It is thought that clear ostensive cues by adults and infants create direct and clean object-word mappings between the caregivers' speech inputs and the object to which they are both attending (Baldwin, 1991; Donnellan et al., 2020; Lieven, 2016). Recent active learning approaches have further suggested that infants might ostensively cue adult attention in order to directly elicit information from them about their environment (Begus et al., 2014).

Concurrent with a number of micro-behavioural findings, Chapters 3 and 4 show that infants do not often engage in active attention-sharing behaviours, suggesting, as outlined in Chapter 2, that other low-level sensorimotor processes, and statistical regularities in the caregivers' inputs might drive infant language learning during naturalistic interactions (Yu & Smith, 2012). Here we outlined four different models of the low-level processes that could drive infant attention during shared interactions: each with different implications for how infants learn from the speech inputs of their caregivers'. Integrating findings across neural and behavioural analyses in Chapters 3-5, we were able to delineate between these different models by examining the endogenous mechanisms and interactive contingencies that structure the allocation of infant attention and behaviour on a moment-by-moment basis in shared interactions.

As mentioned above, we found no support for the suggestion that infant attention is controlled by the sensorimotor cues of their partner (Chapter 4); arguing against the idea that the timing, regularity and saliency of the caregivers' speech inputs are the result of adult-directed processes alone. The other two models implicated both adult- and infant-directed factors in driving language learning. The first implicates infant interest in driving and maintaining infant attention towards objects, and in caregivers providing object labels at these moments. For example, it has been argued that moments of object dominance in the childs' field of view, that reduce the referential uncertainty of the caregivers' speech input, are as the result of infant interest in that object (Yu & Smith, 2012). However, somewhat against this perspective, in Chapter 3, analysis of infant neural activity in the time before they led their partners' attention towards an object showed no increase in infant theta activity relative to where they followed their partners' attention. This suggests that we may not be able consider language learning purely a result of interest-input associations; we need to consider, instead, how speech inputs occur relative to the dynamic patterning of interactive contingencies throughout the caregiver-infant interaction.

Indeed, the findings of Chapters 3 and 4 are most consistent with the last model outlined in Chapter 2 where infant attention is structured through dynamic interactive contingencies that regulate infant attention through allostatic process (S. V. Wass, 2021a). As outlined above, Chapter 4 shows that longer infant look durations associate with reactive and dynamic changes in the caregivers' salient sensorimotor behaviours, and combined with the cross-correlation of infant attention to theta activity, suggest that repeated interpersonal contingences within and between joint attentional episodes could drive periods of increased attentional and intentional engagement by infants. The placement of speech inputs by the caregiver at moments of increased endogenous control by infants over their own attention, could, in turn, be particularly important to infant learning (McQuillan et al., 2020; Yu & Smith, 2012).

These findings emphasise the need to examine caregiver speech inputs, not relative to the moment of interactive contingency where an object label is produced (Goldstein et al., 2010; Wu & Gros-Louis, 2015), but relative to the dynamic contingencies leading up to that moment, and that associate with increased endogenous control by infants over their behaviour. Similarly, as well as contingent responsivity by adults and infants to each others' behavioural cues, continuous neural tracking of the oscillatory structures in caregiver behaviours could increase infant sensitivity to both the attentional focus of their partner and their speech input around moments of object labelling. For example, in interactions with 14-month-old infants, object labels provided by caregivers occur at peaks in the amplitude envelope (Messer, 1981; Nencheva & Lew-Williams, 2022), and, during book-sharing interactions, caregivers produce object labels with exaggerated pitch intonation, which are often placed at the end of sentences (Fernald & Mazzie, 1991). How neural tracking of the caregivers' speech signal during naturalistic interactions associates with coordination by the infant to the focus of their caregivers' speech inputs is a crucial question for future research.

6.3 Limitations and future directions

The findings of the present thesis show the potential for neurocognitive methods of analysis in studying early infant-caregiver interactions for furthering our understanding of the endogenous cognitive mechanisms and bidirectional contingencies that give rise to the structure of infant attention during shared interactions that, using behavioural methods alone, would be unidentifiable.

As outlined in the Chapters 3, 4 and 5, however, examining the relationship of infant neural activity to their naturalistic interactive behaviours does face some limitations. One limitation considered in the analysis and interpretation of the findings in all 3 chapters is the issue of movement-related artifact in the EEG signal, which is particularly problematic in naturalistic datasets where the infant is freely moving (Georgieva et al., 2020; Marriott Haresign et al., 2021). In all 3 chapters movement-related artifacts were removed using an automated algorithm for identifying artifactual components with ICA decomposition (Marriott Haresign et al., 2021). This, however, does not remove all movement-related artifact from the data, which is often time-locked to the behaviour events of interest (e.g. re-orientations of infant attention towards objects and their partner (Marriott Haresign et al., 2021; Noreika et al., 2020)). That said, in Chapters 2 and 3, through detailed behavioural analysis of infant attention, and consideration of the direction of the relationships between infant attention and their neural activity, we showed that it was unlikely that the reported effects were driven by artifact in the EEG signal. In Chapter 5, however, this was more difficult, and a key area for future work should be in examining the influence of EEG artifact in testing speech-brain tracking in naturalistic interactions. As the field of naturalistic EEG recording moves forward, more sophisticated techniques for artifact removal are being developed (Marriott-Haresign et al., 2021).

Another of the limitations associated with recording EEG from infants during live, free flowing naturalistic interactions is that some infant behaviours occur too infrequently to examine associations of EEG activity to infant behaviour at the individual level. In Chapter 3, for example, it was shown that the probability of infant vocalising during the course of the interaction was low, which is common in lab-based experiments (Gratier et al., 2015). In the discussion section of Chapter 3 it was considered that, whilst we did not find an increase in infant theta activity in the time before all infant-led looks towards objects, it could be the case that certain infant behaviours are more endogenously guided than others, and that, caregiver responses to these behaviours might be particularly important to infants' socio-communicative understanding and language development (Murray et al., 2016). As well as infant vocalisations, gestures, and specific body movements would be particularly relevant to the current investigation (Lucca & Wilbourn, 2018).

In the sections above, it was argued that the patterning of action generated contingencies throughout the course of the interaction, as well as infants' attunement to micro-oscillatory processes of the caregiver, could drive moments of endogenously driven behaviour and cognitive process, unidentifiable in the unimodal and intra-personal relationships examined in Chapter 4. Similar to much previous work conducted in social interactions (Yu & Smith, 2013, 2017), this thesis examined a wide, but, nevertheless, limited number of interactive behaviours. In Chapter 3, for example we conducted analyses on the temporal structure of caregiver and infant gaze and vocal behaviours, and, in Chapters 4 and 5, the spectral qualities in caregiver speech were also examined relative to infant attention and neural activity. Researchers are often limited by resources for coding the micro-dynamics of interactions, given the time and resources it takes to code these fine-grained behaviours.

To advance the findings of the current thesis, we need to move towards continuous modelling and data-driven methods of analysis, that utilise automated behavioural coding packages and online motion-tracking software (Pereira et al., 2008) to examine the dynamic associations between infant brain activity and inter-dyadic behaviour. A key focus of this work should be on examining multimodal behavioural process in supporting the development of early joint attention and cognition, and in tracking the longitudinal relationships between these influences over time (Gaffan et al., 2010). For example, one particularly interesting avenue for future work would be in using data-driven approaches to examine the intra- and inter-dyadic behavioural process associated with specific patterns of neural activity in infants (Bayet et al., 2020).

The results presented in Chapter 5 show that it is possible to measure infant neural tracking of their caregivers' speech signal during shared naturalistic interactions, using continuous modelling approaches. Similar methods should be utilised in future work to examine how infants track multi-modal patterns of caregiver behaviour that operate across multiple time scales, as well as their neural responsivity to the contingencies that these multimodal behaviours create (Jessen et al., 2019). For example, in mTRF analyses, it is possible to regress numerous behavioural variables on to the infants' continuous EEG signal, opening exciting opportunities to examine how infants attune to multiple levels of interactive process. This work will aid further insight into the attentional mechanisms through which early interactions support infants' developing capacity for attentional control over their own behaviours, as well as their role in driving early language development.

6.4 Overall conclusions

The contribution of this thesis is in taking a new, multi-method approach, combining neural and behavioural techniques, to examining the intra-individual and bidirectional processes that characterise early triadic caregiver-infant interactions. In all 3 empirical chapters, this work has provided insight into the sub-second neural activity associated with the timing of infant attention, as well as their sensitivity to the dynamic interactive contingencies that drive and maintain inter-personal coordination on a moment-by-moment basis. In particular, the findings of the present thesis advance our understanding of the endogenous cognitive activity and predictive processing strategies that structure infant attention in shared interactions. From this, the work generates new perspectives on the learning mechanisms involved in the development of intentionally mediated behaviours, with cascading implications for our understanding of how inter-dyadic process supports early development (Murray, De Pascalis, Bozicevic, et al., 2016; Murray et al., 1993; Murray & Trevarthen, 1986; Tamis-LeMonda et al., 2001). These are mechanisms that have only previously been considered as possibilities in explaining the timing of infant attention and the organisation of inter-dyadic co-ordination using behavioural methods alone (Yu & Smith, 2012).

This, however, is only the first step. The development of well-designed, and purposefully constructed set-ups to examining the continuous associations between intra-personal neural activity and inter-dyadic sensorimotor process in naturalistic interactions, with attention to multi-modal behaviours across multiple time-scales, as well as advances in sophisticated analytical deigns will advance the findings reported here. Through the integration of different micro-analysis methods this work will be crucial to furthering our understanding of how infants begin to learn from and engage with a language system.

References

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

Albert, R. R., Schwade, J. A., & Goldstein, M. H. (2018). The social functions of babbling: Acoustic and contextual characteristics that facilitate maternal responsiveness. *Developmental Science*, *21*(5). https://doi.org/10.1111/desc.12641

Aslin, R. N., & Newport, E. L. (2012). Statistical Learning: From Acquiring Specific Items to Forming General Rules. *Current Directions in Psychological Science*, *21*(3), 170–176. https://doi.org/10.1177/0963721412436806

Attaheri, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., Olawole-Scott, H., Boutris, P., Gibbon, S., Williams, I., Grey, C., Flanagan, S., & Goswami, U. (2022). Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *NeuroImage*, *247*, 118698. https://doi.org/10.1016/j.neuroimage.2021.118698

Baldwin, D. A. (1991). Infants’ Contribution to the Achievement of Joint Reference. *Child Development*, *62*(5), 875. https://doi.org/10.2307/1131140

Bayet, L., Zinszer, B. D., Reilly, E., Cataldo, J. K., Pruitt, Z., Cichy, R. M., Nelson, C. A., & Aslin, R. N. (2020). Temporal dynamics of visual representations in the infant brain. *Developmental Cognitive Neuroscience*, *45*, 100860. https://doi.org/10.1016/j.dcn.2020.100860

Bazhydai, M., Westermann, G., & Parise, E. (2020). “I don’t know but I know who to ask”: 12-month-olds actively seek information from knowledgeable adults. *Developmental Science*, *23*(5). https://doi.org/10.1111/desc.12938

Bednar, A., & Lalor, E. C. (2020). Where is the cocktail party? Decoding locations of attended and unattended moving sound sources using EEG. *NeuroImage*, *205*, 116283. https://doi.org/10.1016/j.neuroimage.2019.116283

Begus, K., & Bonawitz, E. (2020). The rhythm of learning: Theta oscillations as an index of active learning in infancy. *Developmental Cognitive Neuroscience*, *45*, 100810. https://doi.org/10.1016/j.dcn.2020.100810

Begus, K., Gliga, T., & Southgate, V. (2014). Infants Learn What They Want to Learn: Responding to Infant Pointing Leads to Superior Learning. *PLoS ONE*, *9*(10), e108817. https://doi.org/10.1371/journal.pone.0108817

Begus, K., Gliga, T., & Southgate, V. (2016). Infants' preferences for native speakers are associated with an expectation of information. *Proceedings of the National Academy of Sciences*, *113*(44), 12397–12402. https://doi.org/10.1073/pnas.1603261113

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function: Infant pointing serves an interrogative function. *Developmental Science*, *15*(5), 611–617. https://doi.org/10.1111/j.1467-7687.2012.01160.x

Begus, K., & Southgate, V. (2018). Curious Learners: How Infants' Motivation to Learn Shapes and Is Shaped by Infants' Interactions with the Social World. In M. M. Saylor & P. A. Ganea (Eds.), *Active Learning from Infancy to Childhood* (pp. 13–37). Springer International Publishing. https://doi.org/10.1007/978-3-319-77182-3_2

Begus, K., Southgate, V., & Gliga, T. (2015). Neural mechanisms of infant learning: Differences in frontal theta activity during object exploration modulate subsequent object recognition. *Biology Letters*, *11*(5), 20150041. https://doi.org/10.1098/rsbl.2015.0041

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

Benjamini Yoav, & Hochberg Yosef. (1995). Controlling for false discovary rate: A practical and powerful approach to multiple testing. *Journal of the Royal Society. Series B (Methodological)*, *57*.

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, *9*. https://doi.org/10.3389/fninf.2015.00016

Bigelow, A. E., & Power, M. (2016). Effect of Maternal Responsiveness on Young Infants' Social Bidding-Like Behavior during the Still Face Task: Maternal Responsiveness and Infants' Social Bidding-Like Behavior. *Infant and Child Development*, *25*(3), 256–276. https://doi.org/10.1002/icd.1974

Bigelow, A. E., & Rochat, P. (2006). Two-Month-Old Infants' Sensitivity to Social Contingency in Mother-Infant and Stranger-Infant Interaction. *Infancy*, *9*(3), 313–325. https://doi.org/10.1207/s15327078in0903_3

Bornstein, M. H. (1985). How infant and mother jointly contribute to developing cognitive competence in the child. *Proceedings of the National Academy of Sciences*, *82*(21), 7470–7473. https://doi.org/10.1073/pnas.82.21.7470

Bornstein, M. H., & Azuma, H. (1990.). *Maternal Responsiveness to Infants in Three Societies: The United States, France, and Japan*. 15.

Borsema, P, & Weenik, D. (2019). *Praat: Doing phonetics by computer* (61.09) [Computer programme].

Bourvis, N., Singer, M., Saint Georges, C., Bodeau, N., Chetouani, M., Cohen, D., & Feldman, R. (2018). Pre-linguistic infants employ complex communicative loops to engage mothers in social exchanges and repair interaction ruptures. *Royal Society Open Science*, *5*(1), 170274. https://doi.org/10.1098/rsos.170274

Braithwaite, E. K., Jones, E. J. H., Johnson, M. H., & Holmboe, K. (2020). Dynamic modulation of frontal theta power predicts cognitive ability in infancy. *Developmental Cognitive Neuroscience*, *45*, 100818. https://doi.org/10.1016/j.dcn.2020.100818

Brodbeck, C., & Simon, J. Z. (2020). Continuous speech processing. *Current Opinion in Physiology*, *18*, 25–31. https://doi.org/10.1016/j.cophys.2020.07.014

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*(01). https://doi.org/10.1017/S030500090700829X

Brooks, R., & Meltzoff, A. N. (2015). Connecting the dots from infancy to childhood: A longitudinal study connecting gaze following, language, and explicit theory of mind. *Journal of Experimental Child Psychology*, *130*, 67–78. https://doi.org/10.1016/j.jecp.2014.09.010

Brookshire, G. (2022). Putative rhythms in attentional switching can be explained by aperiodic temporal structure. *Nature Human Behaviour*, *6*(9), 1280–1291. https://doi.org/10.1038/s41562-022-01364-0

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. https://doi.org/10.2307/1166214

Chi, T., Ru, P., & Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *The Journal of the Acoustical Society of America*, *118*(2), 887–906. https://doi.org/10.1121/1.1945807

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice*. The MIT Press.

Cohn, J. F., & Tronick, E. Z. (1987). *Mother-Infant Face-to-Face Interaction: The Sequence of Dyadic States at 3, 6, and 9 Months*. 10.

Colombo, J. (2001). The Development of Visual Attention in Infancy. *Annual Review of Psychology*, *52*(1), 337–367. https://doi.org/10.1146/annurev.psych.52.1.337

Cooper, R. P., & Aslin, R. N. (1994). Developmental Differences in Infant Attention to the Spectral Properties of Infant-Directed Speech. *Child Development*, *65*(6), 1663. https://doi.org/10.2307/1131286

Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent Visual Speech Enhances Cortical Entrainment to Continuous Auditory Speech in Noise-Free Conditions. *Journal of Neuroscience*, *35*(42), 14195–14204. https://doi.org/10.1523/JNEUROSCI.1829-15.2015

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00604

Crosse, M. J., Zuk, N. J., Di Liberto, G. M., Nidiffer, A. R., Molholm, S., & Lalor, E. C. (2021). Linear Modeling of Neurophysiological Responses to Speech and Other Continuous Stimuli: Methodological Considerations for Applied Research. *Frontiers in Neuroscience*, *15*, 705621. https://doi.org/10.3389/fnins.2021.705621

Crown, C. L., Feldstein, S., Jasnow, M. D., Beebe, B., & Jaffe, J. (2002). The Cross-Modal Coordination of Interpersonal Timing: Six-Week-Olds Infants' Gaze with Adults' Vocal Behavior. *31*, *1*, 23.

Csibra, G. (2010). Recognizing Communicative Intentions in Infancy. *Mind & Language*, *25*(2), 141–168. https://doi.org/10.1111/j.1468-0017.2009.01384.x

Csibra, G., & Gergely, G. (n.d.). *Social learning and social cognition: The case for pedagogy*. 15.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Custode, S. A., & Tamis-LeMonda, C. (2020). Cracking the code: Social and contextual cues to language input in the home environment. *Infancy, 25 (6)*.

Deák, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science*, *17*(2), 270–281. https://doi.org/10.1111/desc.12122

Debnath, R., Buzzell, G. A., Morales, S., Bowers, M. E., Leach, S. C., & Fox, N. A. (2020). The Maryland analysis of developmental EEG (MADE) pipeline. *Psychophysiology*, *57*(6). https://doi.org/10.1111/psyp.13580

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Di Liberto, G. M., & Lalor, E. C. (2017). Indexing cortical entrainment to natural speech at the phonemic level: Methodological considerations for applied research. *Hearing Research*, *348*, 70–77. https://doi.org/10.1016/j.heares.2017.02.015

Dimigen, O. (2020). Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. *NeuroImage*, *207*, 116117. https://doi.org/10.1016/j.neuroimage.2019.116117

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164. https://doi.org/10.1038/nn.4186

Ding, N., & Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, *107*(1), 78–89. https://doi.org/10.1152/jn.00297.2011

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85*, 761–768. https://doi.org/10.1016/j.neuroimage.2013.06.035

Donnellan, E., Bannard, C., McGillion, M. L., Slocombe, K. E., & Matthews, D. (2020). Infants’ intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language: A longitudinal investigation of infants’ vocalizations, gestures and word production. *Developmental Science*, *23*(1). https://doi.org/10.1111/desc.12843

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*(14), 9602–9605. https://doi.org/10.1073/pnas.152159999

Farroni, T., Johnson, M. H., Brockbank, M., & Simion, F. (2000). Infants’ use of gaze direction to cue attention: The importance of perceived motion. *Visual Cognition*, *7*(6), 705–718. https://doi.org/10.1080/13506280050144399

Farroni, T., Mansfield, E. M., Lai, C., & Johnson, M. H. (2003). Infants perceiving and acting on the eyes: Tests of an evolutionary hypothesis. *Journal of Experimental Child Psychology*, *85*(3), 199–212. https://doi.org/10.1016/S0022-0965(03)00022-5

Feldman, R. (2007). Parent–Infant Synchrony: Biological Foundations and Developmental Outcomes. *Current Directions in Psychological Science*, *16*(6), 340–345. https://doi.org/10.1111/j.1467-8721.2007.00532.x

Feldman, R., Greenbaum, C. W., & Yirmiya, N. (1999). Mother-infant affect synchrony as an antecedent of the emergence of self-control. *Developmental Psychology*, *35*(1), 223–231. https://doi.org/10.1037//0012-1649.35.1.223

Feldman, R., & Mayes, L. C. (1999). The cyclic organization of attention during habituation is related to infants' information processing. *Infant Behavior and Development*, *22*(1), 37–49. https://doi.org/10.1016/S0163-6383(99)80004-6

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development*, *59*(5), i. https://doi.org/10.2307/1166093

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental psychology*, *27*(2), 209.

Fiedler, L., Wöstmann, M., Graversen, C., Brandmeyer, A., Lunner, T., & Obleser, J. (2017). Single-channel in-ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech. *Journal of Neural Engineering*, *14*(3), 036020. https://doi.org/10.1088/1741-2552/aa66dd

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

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

Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as Hypotheses: Saccades as Experiments. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00151

Friston, K. J. (2019). Waves of prediction. *PLOS Biology*, *17*(10), e3000426. https://doi.org/10.1371/journal.pbio.3000426

Frith, C. D., & Frith, U. (2007). Social Cognition in Humans. *Current Biology*, *17*(16), R724–R732. https://doi.org/10.1016/j.cub.2007.05.068

Gabard-Durnam, L. J., Mendez Leal, A. S., Wilkinson, C. L., & Levin, A. R. (2018). The Harvard Automated Processing Pipeline for Electroencephalography (HAPPE): Standardized Processing Software for Developmental and High-Artifact Data. *Frontiers in Neuroscience*, *12*, 97. https://doi.org/10.3389/fnins.2018.00097

Gaffan, E. A., Martins, C., Healy, S., & Murray, L. (2010). Early Social Experience and Individual Differences in Infants' Joint Attention. *Social Development*, *19*(2), 369–393. https://doi.org/10.1111/j.1467-9507.2008.00533.x

Georgieva, S., Lester, S., Noreika, V., Yilmaz, M. N., Wass, S., & Leong, V. (2020). Toward the Understanding of Topographical and Spectral Signatures of Infant Movement Artifacts in Naturalistic EEG. *Frontiers in Neuroscience*, *14*, 352. https://doi.org/10.3389/fnins.2020.00352

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Nugent, T. F., Herman, D. H., Clasen, L. S., Toga, A. W., Rapoport, J. L., & Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences*, *101*(21), 8174–8179. https://doi.org/10.1073/pnas.0402680101

Goldin-Meadow, S. (2007). Pointing sets the stage for learning language—and creating language. *Child development*, *78*(3), 741-745.

Goldin-Meadow, S., Goodrich, W., Sauer, E., & Iverson, J. (2007). Young children use their hands to tell their mothers what to say. *Developmental Science*, *10*(6), 778–785. https://doi.org/10.1111/j.1467-7687.2007.00636.x

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, *100*(13), 8030–8035. https://doi.org/10.1073/pnas.1332441100

Goldstein, M. H., & Schwade, J. A. (2008). Social Feedback to Infants' Babbling Facilitates Rapid Phonological Learning. *Psychological Science*, *19*(5), 515–523. https://doi.org/10.1111/j.1467-9280.2008.02117.x

Goldstein, M. H., Schwade, J., Briesch, J., & Syal, S. (2010). Learning While Babbling: Prelinguistic Object-Directed Vocalizations Indicate a Readiness to Learn: LEARNING WHILE BABBLING. *Infancy*, *15*(4), 362–391. https://doi.org/10.1111/j.1532-7078.2009.00020.x

Goswami, U., & Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives? *Laboratory Phonology*, *4*(1). https://doi.org/10.1515/lp-2013-0004

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: Computational and neural mechanisms. *Trends in Cognitive Sciences*, *17*(11), 585–593. https://doi.org/10.1016/j.tics.2013.09.001

Goupil, L., & Proust, J. (2023). Curiosity as a metacognitive feeling. *Cognition*, *231*, 105325. https://doi.org/10.1016/j.cognition.2022.105325

Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don't know. *Proceedings of the National Academy of Sciences*, *113*(13), 3492–3496. https://doi.org/10.1073/pnas.1515129113

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01167

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—A syllable-centric perspective. *Journal of Phonetics*, *31*(3–4), 465–485. https://doi.org/10.1016/j.wocn.2003.09.005

Gros-Louis, J., West, M. J., & King, A. P. (2014). Maternal Responsiveness and the Development of Directed Vocalizing in Social Interactions. *Infancy*, *19*(4), 385–408. https://doi.org/10.1111/infa.12054

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLoS Biology*, *11*(12), e1001752. https://doi.org/10.1371/journal.pbio.1001752

Grossmann, T., Lloyd-Fox, S., & Johnson, M. H. (2013). Brain responses reveal young infants' sensitivity to when a social partner follows their gaze. *Developmental Cognitive Neuroscience*, *6*, 155–161. https://doi.org/10.1016/j.dcn.2013.09.004

Hamilton, A. F. de C. (2021). Hyperscanning: Beyond the Hype. *Neuron*, *109*(3), 404–407. https://doi.org/10.1016/j.neuron.2020.11.008

Hansen, J. H. L., Stauffer, A., & Xia, W. (2021). Nonlinear waveform distortion: Assessment and detection of clipping on speech data and systems. *Speech Communication*, *134*, 20–31. https://doi.org/10.1016/j.specom.2021.07.007

Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: A neuromagnetic study. *Proceedings of the National Academy of Sciences*, *95*(25), 15061–15065. https://doi.org/10.1073/pnas.95.25.15061

Hari, R., Henriksson, L., Malinen, S., & Parkkonen, L. (2015). Centrality of Social Interaction in Human Brain Function. *Neuron*, *88*(1), 181–193. https://doi.org/10.1016/j.neuron.2015.09.022

Hasson, U., & Frith, C. D. (2016). Mirroring and beyond: Coupled dynamics as a generalized framework for modelling social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), 20150366. https://doi.org/10.1098/rstb.2015.0366

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, *16*(2), 114–121. https://doi.org/10.1016/j.tics.2011.12.007

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067

Henning, A., Striano, T., & Lieven, E. V. M. (2005). Maternal speech to infants at 1 and 3 months of age. *Infant Behavior and Development*, *28*(4), 519–536. https://doi.org/10.1016/j.infbeh.2005.06.001

Hoehl, S., Michel, C., Reid, V. M., Parise, E., & Striano, T. (2014). Eye contact during live social interaction modulates infants' oscillatory brain activity. *Social Neuroscience*, *9*(3), 300–308. https://doi.org/10.1080/17470919.2014.884982

Hood, B. M., Willen, J. D., & Driver, J. (1998). ADULT'S EYES TRIGGER SHIFTS OF VISUAL ATTENTION IN HUMAN INFANTS. *PSYCHOLOGICAL SCIENCE*, *9*(2), 4.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture Paves the Way for Language Development. *Psychological Science*, *16*(5), 367–371. https://doi.org/10.1111/j.0956-7976.2005.01542.x

Jaffe, J., Beebe, B., & Feldstein, S. (2001). *Rhythms of Dialogue in Infancy: Coordinated Timing in Development*. *66*(2), 1–149.

Jessen, S., Fiedler, L., Münte, T. F., & Obleser, J. (2019). Quantifying the individual auditory and visual brain response in 7-month-old infants watching a brief cartoon movie. *NeuroImage*, *202*, 116060. https://doi.org/10.1016/j.neuroimage.2019.116060

Jessen, S., Obleser, J., & Tune, S. (2021). Neural tracking in infants – An analytical tool for multisensory social processing in development. *Developmental Cognitive Neuroscience*, *52*, 101034. https://doi.org/10.1016/j.dcn.2021.101034

Jones, E. J. H., Goodwin, A., Orekhova, E., Charman, T., Dawson, G., Webb, S. J., & Johnson, M. H. (2020). Infant EEG theta modulation predicts childhood intelligence. *Scientific Reports*, *10*(1), 11232. https://doi.org/10.1038/s41598-020-67687-y

Kalashnikova, M., Peter, V., Di Liberto, G. M., Lalor, E. C., & Burnham, D. (2018). Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Scientific Reports*, *8*(1), 13745. https://doi.org/10.1038/s41598-018-32150-6

Kidd, C., & Hayden, B. Y. (2015). The Psychology and Neuroscience of Curiosity. *Neuron*, *88*(3), 449–460. https://doi.org/10.1016/j.neuron.2015.09.010

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple Nor Too Complex. *PLoS ONE*, *7*(5), e36399. https://doi.org/10.1371/journal.pone.0036399

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks Effect in Infant Auditory Attention. *Child Development*, n/a-n/a. https://doi.org/10.1111/cdev.12263

Kilner, J. M., Vargas, C., Duval, S., Blakemore, S.-J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience*, *7*(12), 1299–1301. https://doi.org/10.1038/nn1355

Klimesch, W., Sauseng, P., Hanslmayr, S., Gruber, W., & Freunberger, R. (2007). Event-related phase reorganization may explain evoked neural dynamics. *Neuroscience & Biobehavioral Reviews*, *31*(7), 1003–1016. https://doi.org/10.1016/j.neubiorev.2007.03.005

Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: From entrainment to joint action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1499), 2021–2031. https://doi.org/10.1098/rstb.2008.0006

Konvalinka, I., & Roepstorff, A. (2012). The two-brain approach: How can mutually interacting brains teach us something about social interaction? *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00215

Kovács, Á. M., Tauzin, T., Téglás, E., Gergely, G., & Csibra, G. (2014). Pointing as Epistemic Request: 12-month-olds Point to Receive New Information. *Infancy*, *19*(6), 543–557. https://doi.org/10.1111/infa.12060

Lalor, E. C., Power, A. J., Reilly, R. B., & Foxe, J. J. (2009). Resolving Precise Temporal Processing Properties of the Auditory System Using Continuous Stimuli. *Journal of Neurophysiology*, *102*(1), 349–359. https://doi.org/10.1152/jn.90896.2008

Leong, V., Byrne, E., Clackson, K., Georgieva, S., Lam, S., & Wass, S. (2017). Speaker gaze increases information coupling between infant and adult brains. *Proceedings of the National Academy of Sciences*, *114*(50), 13290–13295. https://doi.org/10.1073/pnas.1702493114

Leong, V., Byrne, E., Clackson, K., Harte, N., Lam, S., Barbaro, K. de, & Wass, S. (2017). *Infants' neural oscillatory processing of theta-rate speech patterns exceeds adults'* [Preprint]. Neuroscience. https://doi.org/10.1101/108852

Leong, V., & Goswami, U. (2015b). Acoustic-Emergent Phonology in the Amplitude Envelope of Child-Directed Speech. *PLOS ONE*, *10*(12), e0144411. https://doi.org/10.1371/journal.pone.0144411

Leong, V., Kalashnikova, M., Burnham, D., & Goswami, U. (2017). The Temporal Modulation Structure of Infant-Directed Speech. *Open Mind*, *1*(2), 78–90. https://doi.org/10.1162/OPMI_a_00008

Lieven, E. (2016). Usage-based approaches to language development: Where do we go from here? *Language and Cognition*, *8*(3), 346–368. https://doi.org/10.1017/langcog.2016.16

Liszkowski, U., Albrecht, K., Carpenter, M., & Tomasello, M. (2008). Infants' visual and auditory communication when a partner is or is not visually attending. *Infant Behavior and Development*, *31*(2), 157–167. https://doi.org/10.1016/j.infbeh.2007.10.011

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science*, *7*(3), 297–307. https://doi.org/10.1111/j.1467-7687.2004.00349.x

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108*(3), 732–739. https://doi.org/10.1016/j.cognition.2008.06.013

Liszkowski, U., & Tomasello, M. (2011). Individual differences in social, cognitive, and morphological aspects of infant pointing. *Cognitive Development*, *26*(1), 16–29. https://doi.org/10.1016/j.cogdev.2010.10.001

Lucca, K., & Wilbourn, M. P. (2018). Communicating to Learn: Infants' Pointing Gestures Result in Optimal Learning. *Child Development*, *89*(3), 941–960. https://doi.org/10.1111/cdev.12707

Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron*, *54*(6), 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Marriott Haresign, I., Phillips, E. A. M., Whitehorn, M., Lamagna, F., Eliano, M., Goupil, L., Jones, E. J. H., & Wass, S. V. (2022). Gaze onsets during naturalistic infant-caregiver interaction associate with 'sender' but not 'receiver' neural responses, and do not lead to changes in inter-brain synchrony. *BioRxiv*, 2022.05.27.493545. https://doi.org/10.1101/2022.05.27.493545

Marriott Haresign, I., Phillips, E., Whitehorn, M., Noreika, V., Jones, E. J. H., Leong, V., & Wass, S. V. (2021). Automatic classification of ICA components from infant EEG using MARA. *Developmental Cognitive Neuroscience*, *52*, 101024. https://doi.org/10.1016/j.dcn.2021.101024

Mason, G. M. (2019). The role of multisensory development in early language learning. *Journal of Experimental Child Psychology*, 17.

Mason, G. M., Kirkpatrick, F., Schwade, J. A., & Goldstein, M. H. (2019). The Role of Dyadic Coordination in Organizing Visual Attention in 5-Month-Old Infants. *Infancy*, *24*(2), 162–186. https://doi.org/10.1111/infa.12255

McQuillan, M. E., Smith, L. B., Yu, C., & Bates, J. E. (2020). Parents Influence the Visual Learning Environment Through Children's Manual Actions. *Child Development*, *91*(3). https://doi.org/10.1111/cdev.13274

Mesik, J., & Wojtczak, M. (2022). *The effects of data quantity on performance of temporal response function analyses of natural speech processing* [Preprint]. Neuroscience. https://doi.org/10.1101/2022.06.07.495139

Messer, J.D. (1981). The identification of names in maternal speech to infants. *Journal of Psycholinguistic Research*, *10*, 69–77.

Meyer, L., Sun, Y., & Martin, A. E. (2019). Synchronous, but not entrained: Exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience*, 1–11. https://doi.org/10.1080/23273798.2019.1693050

Meyer, M., Braukmann, R., Stapel, J. C., Bekkering, H., & Hunnius, S. (2016). Monitoring others' errors: The role of the motor system in early childhood and adulthood. *British Journal of Developmental Psychology*, *34*(1), 66–85. https://doi.org/10.1111/bjdp.12101

Meyer, M., Schaik, J. E., Poli, F., & Hunnius, S. (2022). How infant-directed actions enhance infants' attention, learning, and exploration: Evidence from EEG and computational modeling. *Developmental Science*. https://doi.org/10.1111/desc.13259

Miller, J. L., & Gros-Louis, J. (2013). Socially guided attention influences infants' communicative behavior. *Infant Behavior and Development*, *36*(4), 627–634. https://doi.org/10.1016/j.infbeh.2013.06.010

Miller, J. L., Hurdish, E., & Gros-Louis, J. (2018). Different patterns of sensitivity differentially affect infant attention span. *Infant Behavior and Development*, *53*, 1–4. https://doi.org/10.1016/j.infbeh.2018.10.001

Monroy, C., Chen, C., Houston, D., & Yu, C. (2021). Action prediction during real-time parent-infant interactions. *Developmental Science*, *24*(3). https://doi.org/10.1111/desc.13042

Monroy, C. D., Meyer, M., Schröer, L., Gerson, S. A., & Hunnius, S. (2019). The infant motor system predicts actions based on visual statistical learning. *NeuroImage*, *185*, 947–954. https://doi.org/10.1016/j.neuroimage.2017.12.016

Mullen, T. (2012). *CleanLine EEGLAB plugin*. San Diego, CA: Neuroimaging Informatics Toolsand Resources Clearinghouse (NITRC).

Mundy, P., & Newell, L. (2007). Attention, Joint Attention, and Social Cognition. *Current Directions in Psychological Science*, *16*(5), 269–274. https://doi.org/10.1111/j.1467-8721.2007.00518.x

Murray, L., De Pascalis, L., Bozicevic, L., Hawkins, L., Sclafani, V., & Ferrari, P. F. (2016). The functional architecture of mother-infant communication, and the development of infant social expressiveness in the first two months. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep39019

Murray, L., De Pascalis, L., Tomlinson, M., Vally, Z., Dadomo, H., MacLachlan, B., Woodward, C., & Cooper, P. J. (2016). Randomized controlled trial of a book-sharing intervention in a deprived South African community: Effects on carer-infant interactions, and their relation to infant cognitive and socioemotional outcome. *Journal of Child Psychology and Psychiatry*, *57*(12), 1370–1379. https://doi.org/10.1111/jcpp.12605

Murray, L., Kempton, C., Woolgar, M., & Hooper, R. (1993). Depressed Mothers’ Speech to Their Infants and its Relation to Infant Gender and Cognitive Development. *Journal of Child Psychology and Psychiatry*, *34*(7), 1083–1101. https://doi.org/10.1111/j.1469-7610.1993.tb01775.x

Murray, L., & Trevarthen, C. (1986). The infant's role in mother–infant communications. *Journal of Child Language*, *13*(1), 15–29. https://doi.org/10.1017/S0305000900000271

Muthukumaraswamy, S. D., Johnson, B. W., & McNair, N. A. (2004). Mu rhythm modulation during observation of an object-directed grasp. *Cognitive Brain Research*, *19*(2), 195–201. https://doi.org/10.1016/j.cogbrainres.2003.12.001

Nencheva, M. L., & Lew-Williams, C. (2022). Understanding why infant-directed speech supports learning: A dynamic attention perspective. *Developmental Review*, *66*, 101047. https://doi.org/10.1016/j.dr.2022.101047

Nencheva, M. L., Piazza, E. A., & Lew-Williams, C. (2019). The real-time dynamics of child-directed speech: Using pupillometry to evaluate children's processing of natural pitch contours. *The Journal of the Acoustical Society of America*, *145*(3), 1765–1765. https://doi.org/10.1121/1.5101465

Noreika, V., Georgieva, S., Wass, S., & Leong, V. (2020). 14 challenges and their solutions for conducting social neuroscience and longitudinal EEG research with infants. *Infant Behavior and Development*, *58*, 101393. https://doi.org/10.1016/j.infbeh.2019.101393

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision*, *10*(8), 20–20. https://doi.org/10.1167/10.8.20

Nuthmann, A., & Matthias, E. (2014). Time course of pseudoneglect in scene viewing. *Cortex*, *52*, 113–119. https://doi.org/10.1016/j.cortex.2013.11.007

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, *117*(2), 382–405. https://doi.org/10.1037/a0018924

Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. *Trends in Cognitive Sciences*, *23*(11), 913–926. https://doi.org/10.1016/j.tics.2019.08.004

Orekhova, E. (1999). Theta synchronization during sustained anticipatory attention in infants over the second half of the first year of life. *International Journal of Psychophysiology*, *32*(2), 151–172. https://doi.org/10.1016/S0167-8760(99)00011-2

Orekhova, E., Stroganova, T., Posikera, I., & Elam, M. (2006). EEG theta rhythm in infants and preschool children. *Clinical Neurophysiology*, *117*(5), 1047–1062. https://doi.org/10.1016/j.clinph.2005.12.027

Oudeyer, P.-Y., & Smith, L. B. (2016). How Evolution May Work Through Curiosity-Driven Developmental Process. *Topics in Cognitive Science*, *8*(2), 492–502. https://doi.org/10.1111/tops.12196

Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development*, *13*(4), 539–545. https://doi.org/10.1016/0163-6383(90)90022-Z

Parise, E., & Csibra, G. (2013). Neural Responses to Multimodal Ostensive Signals in 5-Month-Old Infants. *PLoS ONE*, *8*(8), e72360. https://doi.org/10.1371/journal.pone.0072360

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cerebral Cortex*, *23*(6), 1378–1387. https://doi.org/10.1093/cercor/bhs118

Pereira, A. F., Smith, L. B., & Yu, C. (2008). Social coordination in toddler's word learning: Interacting systems of perception and action. *Connection Science*, *20*(2–3), 73–89. https://doi.org/10.1080/09540090802091891

Pérez, A., Davis, M. H., Ince, R. A. A., Zhang, H., Fu, Z., Lamarca, M., Ralph, M. L., & Monahan, P. J. (2021). *Timing of brain entrainment to the speech envelope during speaking and listening: Implications for inter-brain synchronisation* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/tzhn4

Phillips, E., Goupil, L., Haresign, I. M., Bruce-Gardyne, E., Csolsim, F.-A., Whitehorn, M., Leong, V., & Wass, S. (in press). *Proactive or reactive? Neural oscillatory insight into the leader-follower dynamics of early infant-caregiver interaction* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/cg38a

Piazza, E. A., Hasenfratz, L., Hasson, U., & Lew-Williams, C. (2018). *Infant and adult brains are coupled to the dynamics of natural communication* [Preprint]. Neuroscience. https://doi.org/10.1101/359810

Piazza, E. A., Iordan, M. C., & Lew-Williams, C. (2017). Mothers Consistently Alter Their Unique Vocal Fingerprints When Communicating with Infants. *Current Biology*, *27*(20), 3162-3167.e3. https://doi.org/10.1016/j.cub.2017.08.074

Piccardi, E. S., & Gliga, T. (2022). Understanding sensory regulation in typical and atypical development: The case of sensory seeking. *Developmental Review*, *65*, 101037. https://doi.org/10.1016/j.dr.2022.101037

Piccardi, E. S., Johnson, M. H., & Gliga, T. (2020). Explaining individual differences in infant visual sensory seeking. *Infancy*, *25*(5), 677–698. https://doi.org/10.1111/infa.12356

Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology*, *28*, 142–149. https://doi.org/10.1016/j.conb.2014.07.005

Poli, F., Serino, G., Mars, R. B., & Hunnius, S. (2020). Infants tailor their attention to maximize learning. *Science Advances*, *6*(39), eabb5053. https://doi.org/10.1126/sciadv.abb5053

Power, A. J., Colling, L. J., Mead, N., Barnes, L., & Goswami, U. (2016). Neural encoding of the speech envelope by children with developmental dyslexia. *Brain and Language*, *160*, 1–10. https://doi.org/10.1016/j.bandl.2016.06.006

Power, A. J., Mead, N., Barnes, L., & Goswami, U. (2013). Neural entrainment to rhythmic speech in children with developmental dyslexia. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00777

Rayson, H., Bonaiuto, J. J., Ferrari, P. F., Chakrabarti, B., & Murray, L. (2019). Building blocks of joint attention: Early sensitivity to having one's own gaze followed. *Developmental Cognitive Neuroscience*, 100631. https://doi.org/10.1016/j.dcn.2019.100631

Rayson, H., Bonaiuto, J. J., Ferrari, P. F., & Murray, L. (2017). Early maternal mirroring predicts infant motor system activation during facial expression observation. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-12097-w

Redcay, E., Kleiner, M., & Saxe, R. (2012). Look at this: The neural correlates of initiating and responding to bids for joint attention. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00169

Renswoude, D. R., Visser, I., Raijmakers, M. E. J., Tsang, T., & Johnson, S. P. (2019). Real-world scene perception in infants: What factors guide attention allocation? *Infancy*, *24*(5), 693–717. https://doi.org/10.1111/infa.12308

Richards, J. E. (2011). Infant attention, arousal and the brain. In *Infant Perception and Cognition*. Oxford University Press.

Richards, J. E., & Anderson, D. R. (2004). Attentional inertia in children's extended looking at television. In *Advances in Child Development and Behavior* (Vol. 32, pp. 163–212). Elsevier. https://doi.org/10.1016/S0065-2407(04)80007-7

Robertson, S. S. (2014). Empty-headed dynamical model of infant visual foraging: Empty-Headed Model. *Developmental Psychobiology*, *56*(5), 1129–1133. https://doi.org/10.1002/dev.21165

Robertson, S. S., Guckenheimer, J., Masnick, A. M., & Bacher, L. F. (2004). The dynamics of infant visual foraging. *Developmental Science*, *7*(2), 194–200. https://doi.org/10.1111/j.1467-7687.2004.00338.x

Saez de Urabain, I. R., Nuthmann, A., Johnson, M. H., & Smith, T. J. (2017). Disentangling the mechanisms underlying infant fixation durations in scene perception: A computational account. *Vision Research*, *134*, 43–59. https://doi.org/10.1016/j.visres.2016.10.015

Sebanz, N., & Knoblich, G. (2009). Prediction in Joint Action: What, When, and Where. *Topics in Cognitive Science*, *1*(2), 353–367. https://doi.org/10.1111/j.1756-8765.2009.01024.x

Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology*, *18*(9), 668–671. https://doi.org/10.1016/j.cub.2008.03.059

Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: Infants' preference for object-directed gaze. *Cognition*, *108*(2), 303–319. https://doi.org/10.1016/j.cognition.2008.02.009

Senju, A., Johnson, M. H., & Csibra, G. (2006). The development and neural basis of referential gaze perception. *Social Neuroscience*, *1*(3–4), 220–234. https://doi.org/10.1080/17470910600989797

Siew, J., Phillips, E., Warreyn, P., Moerman, F., Van Lierde, T., Marriott Haresign, I., Noens, I. & Steyaert, J. (in prep). Cortical tracking at 10- and 14-months in infants with and without an elevated likelihood for autism: links with language.

Siposova, B., & Carpenter, M. (2019). A new look at joint attention and common knowledge. *Cognition*, *189*, 260–274. https://doi.org/10.1016/j.cognition.2019.03.019

Smith, L. B. (2005). Cognition as a dynamic system: Principles from embodiment. *Developmental Review*, *25*(3–4), 278–298. https://doi.org/10.1016/j.dr.2005.11.001

Smith, L. B., & Breazeal, C. (2007). The dynamic lift of developmental process. *Developmental Science*, *10*(1), 61–68. https://doi.org/10.1111/j.1467-7687.2007.00565.x

Smith, L., & Gasser, M. (2005). The Development of Embodied Cognition: Six Lessons from Babies. *Artificial Life*, *11*(1–2), 13–29. https://doi.org/10.1162/1064546053278973

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Sohoglu, E., & Davis, M. H. (2020). Rapid computations of spectrotemporal prediction error support perception of degraded speech. *ELife*, *9*, e58077. https://doi.org/10.7554/eLife.58077

Southgate, V., Johnson, M. H., Karoui, I. E., & Csibra, G. (2010). Motor System Activation Reveals Infants' On-Line Prediction of Others' Goals. *Psychological Science*, *21*(3), 355–359. https://doi.org/10.1177/0956797610362058

Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, *44*, 1–18. https://doi.org/10.1016/j.dr.2016.12.001

Stallworthy, I. C., Sifre, R., Berry, D., Lasch, C., Smith, T. J., & Elison, J. T. (2020). Infants' gaze exhibits a fractal structure that varies by age and stimulus salience. *Scientific Reports*, *10*(1), 17216. https://doi.org/10.1038/s41598-020-73187-w

Stephens, D. W., & Charnov, E. L. (1982). Optimal foraging: Some simple stochastic models. *Behavioral Ecology and Sociobiology*, *10*(4), 251–263. https://doi.org/10.1007/BF00302814

Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language*, *10*(1), 1–15. https://doi.org/10.1017/S0305000900005092

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology*, *55*(1), 96–109. https://doi.org/10.1037/dev0000628

Takahashi, D. Y., Narayanan, D. Z., & Ghazanfar, A. A. (2013). Coupled Oscillator Dynamics of Vocal Turn-Taking in Monkeys. *Current Biology*, *23*(21), 2162–2168. https://doi.org/10.1016/j.cub.2013.09.005

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal Responsiveness and Children's Achievement of Language Milestones. *Child Development*, *72*(3), 748–767. https://doi.org/10.1111/1467-8624.00313

Thut, G. (2006). -Band Electroencephalographic Activity over Occipital Cortex Indexes Visuospatial Attention Bias and Predicts Visual Target Detection. *Journal of Neuroscience*, *26*(37), 9494–9502. https://doi.org/10.1523/JNEUROSCI.0875-06.2006

Tomasello, M. (2016). The ontogeny of cultural learning. *Current Opinion in Psychology*, *8*, 1–4. https://doi.org/10.1016/j.copsyc.2015.09.008

Tomasello, M. (2022). The social-pragmatic theory of word learning. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 401–413. https://doi.org/10.1075/prag.10.4.01tom

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, *10*(1), 121–125. https://doi.org/10.1111/j.1467-7687.2007.00573.x

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A New Look at Infant Pointing. *Child Development*, *78*(3), 705–722. https://doi.org/10.1111/j.1467-8624.2007.01025.x

Trevarthen, C. (2001). Intrinsic motives for companionship in understanding: Their origin, development, and significance for infant mental health. *Infant Mental Health Journal*, *22*(1–2), 95–131. https://doi.org/10.1002/1097-0355(200101/04)22:1<95::AID-IMHJ4>3.0.CO;2-6

Tronick, E. D., Als, H., & Brazelton, T. B. (1977). Mutuality in Mother-Infant Interaction. *Journal of Communication*, *27*(2), 74–79. https://doi.org/10.1111/j.1460-2466.1977.tb01829.x

Van Egeren, L. A., Barratt, M. S., & Roach, M. A. (2001). Mother-infant responsiveness: Timing, mutual regulation, and interactional context. *Developmental Psychology*, *37*(5), 684–697. https://doi.org/10.1037//0012-1649.37.5.684

Wass, S. V. (2021a). Allostasis and metastasis: The yin and yang of childhood self-regulation. *Development and Psychopathology*, 1–12. https://doi.org/10.1017/S0954579421000833

Wass, S. V. (2021b). The origins of effortful control: How early development within arousal/regulatory systems influences attentional and affective control. *Developmental Review*, *61*, 100978. https://doi.org/10.1016/j.dr.2021.100978

Wass, S. V., Clackson, K., Georgieva, S. D., Brightman, L., Nutbrown, R., & Leong, V. (2018a). 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. https://doi.org/10.1111/desc.12667

Wass, S. V., Noreika, V., Georgieva, S., Clackson, K., Brightman, L., Nutbrown, R., Covarrubias, L. S., & Leong, V. (2018b). Parental neural responsivity to infants' visual attention: How mature brains influence immature brains during social interaction. *PLOS Biology*, *16*(12), e2006328. https://doi.org/10.1371/journal.pbio.2006328

Wass, S. V., Perapoch Amadó, M., & Ives, J. (2022). Oscillatory entrainment to our early social or physical environment and the emergence of volitional control. *Developmental Cognitive Neuroscience*, *54*, 101102. https://doi.org/10.1016/j.dcn.2022.101102

Wass, S. V., Whitehorn, M., Marriott Haresign, I., Phillips, E., & Leong, V. (2020). Interpersonal Neural Entrainment during Early Social Interaction. *Trends in Cognitive Sciences*, *24*(4), 329–342. https://doi.org/10.1016/j.tics.2020.01.006

Widmann, A., & Schröger, E. (2012). Filter Effects and Filter Artifacts in the Analysis of Electrophysiological Data. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00233

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, *12*(6), 957–968. https://doi.org/10.3758/BF03206432

Wu, Z., & Gros-Louis, J. (2015). Caregivers provide more labeling responses to infants' pointing than to infants' object-directed vocalizations. *Journal of Child Language*, *42*(3), 538–561. https://doi.org/10.1017/S0305000914000221

Xie, W., Mallin, B. M., & Richards, J. E. (2018). Development of infant sustained attention and its relation to EEG oscillations: An EEG and cortical source analysis study. *Developmental Science*, *21*(3), e12562. https://doi.org/10.1111/desc.12562

Xu, T. L., Abney, D. H., & Yu, C. (2017). *Discovering Multicausality in the Development of Coordinated Behavior*. 6.

Yoo, H., Bowman, D. A., & Oller, D. K. (2018). The Origin of Protoconversation: An Examination of Caregiver Responses to Cry and Speech-Like Vocalizations. *Frontiers in Psychology*, *9*, 1510. https://doi.org/10.3389/fpsyg.2018.01510

Yu, C., & Smith, L. B. (2012a). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262. https://doi.org/10.1016/j.cognition.2012.06.016

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262. https://doi.org/10.1016/j.cognition.2012.06.016

Yu, C., & Smith, L. B. (2013). Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLoS ONE*, *8*(11), e79659. https://doi.org/10.1371/journal.pone.0079659

Yu, C., & Smith, L. B. (2016). The Social Origins of Sustained Attention in One-Year-Old Human Infants. *Current Biology*, *26*(9), 1235–1240. https://doi.org/10.1016/j.cub.2016.03.026

Yu, C., & Smith, L. B. (2017). Multiple Sensory-Motor Pathways Lead to Coordinated Visual Attention. *Cognitive Science*, *41*, 5–31. https://doi.org/10.1111/cogs.12366

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, *22*(1), e12735. https://doi.org/10.1111/desc.12735

Yu, C., Zhang, Y., Slone, L. K., & Smith, L. B. (2021). The infant's view redefines the problem of referential uncertainty in early word learning. *Proceedings of the National Academy of Sciences*, *118*(52), e2107019118. https://doi.org/10.1073/pnas.2107019118

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., Poeppel, D., & Schroeder, C. E. (2013). Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a "Cocktail Party". *Neuron*, *77*(5), 980–991. https://doi.org/10.1016/j.neuron.2012.12.037

APPENDIX A – Supplementary Materials for: Proactive or reactive? Neural oscillatory insight into the leader-follow dynamics of early caregiver-infant interaction

1.1 Methods

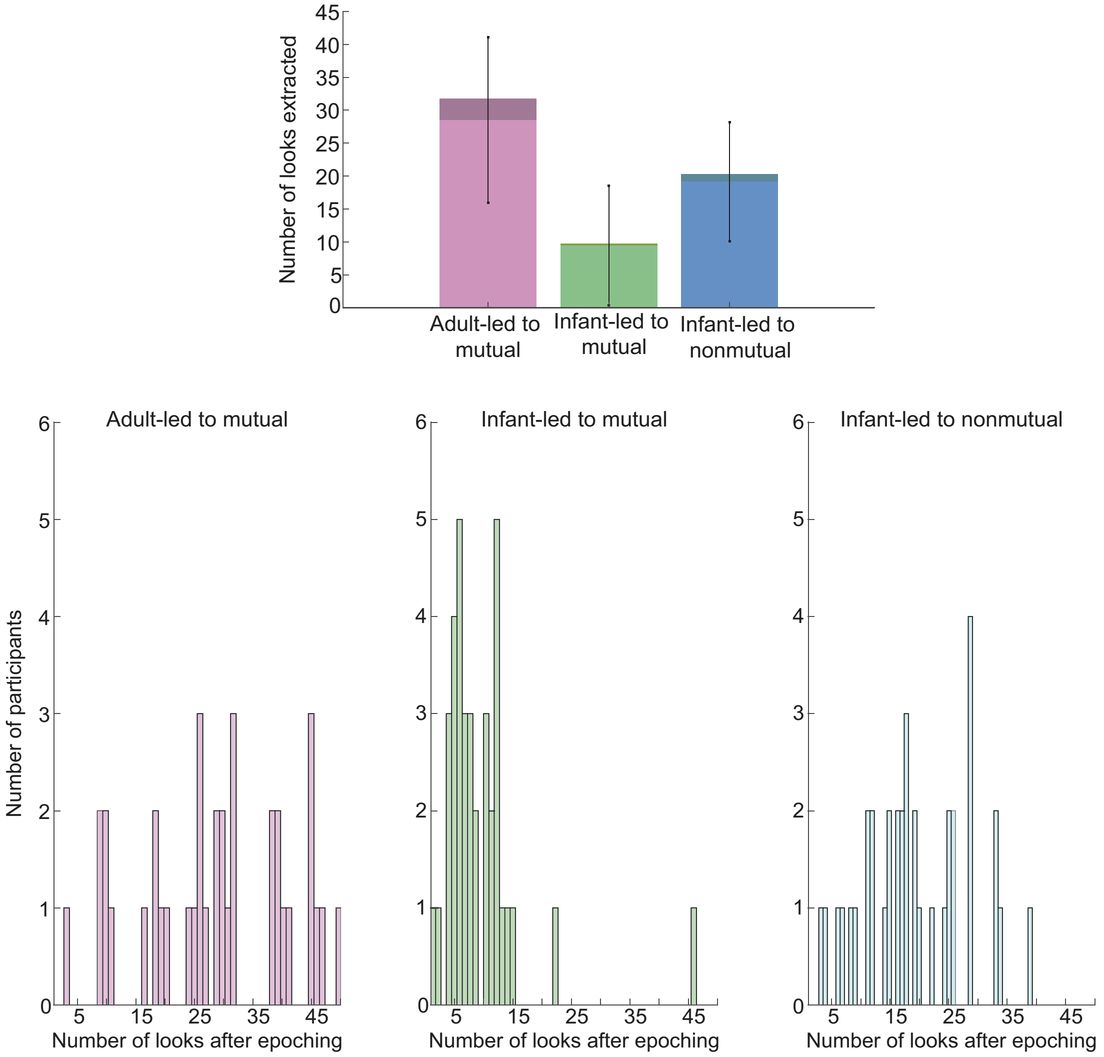

Figure S1 a) Bar plot shows the mean number of looks included in each category after EEG epoching. Error bars show the standard deviation (n=37). b) Histograms show the number of participants contributing a certain number of look epochs for each attention episode. One participant contributed fewer than five trials for adult-led looks to mutual attention; 5 for infant-led looks to mutual attention, and 2 for infant-led looks to nonmutual attention. Participants contributing fewer than 5 looks for one attention episode in each comparison (e.g., infant-led mutual attention compared to adult-led mutual attention) were excluded from the cluster-based permutation analysis.

1.2 Results

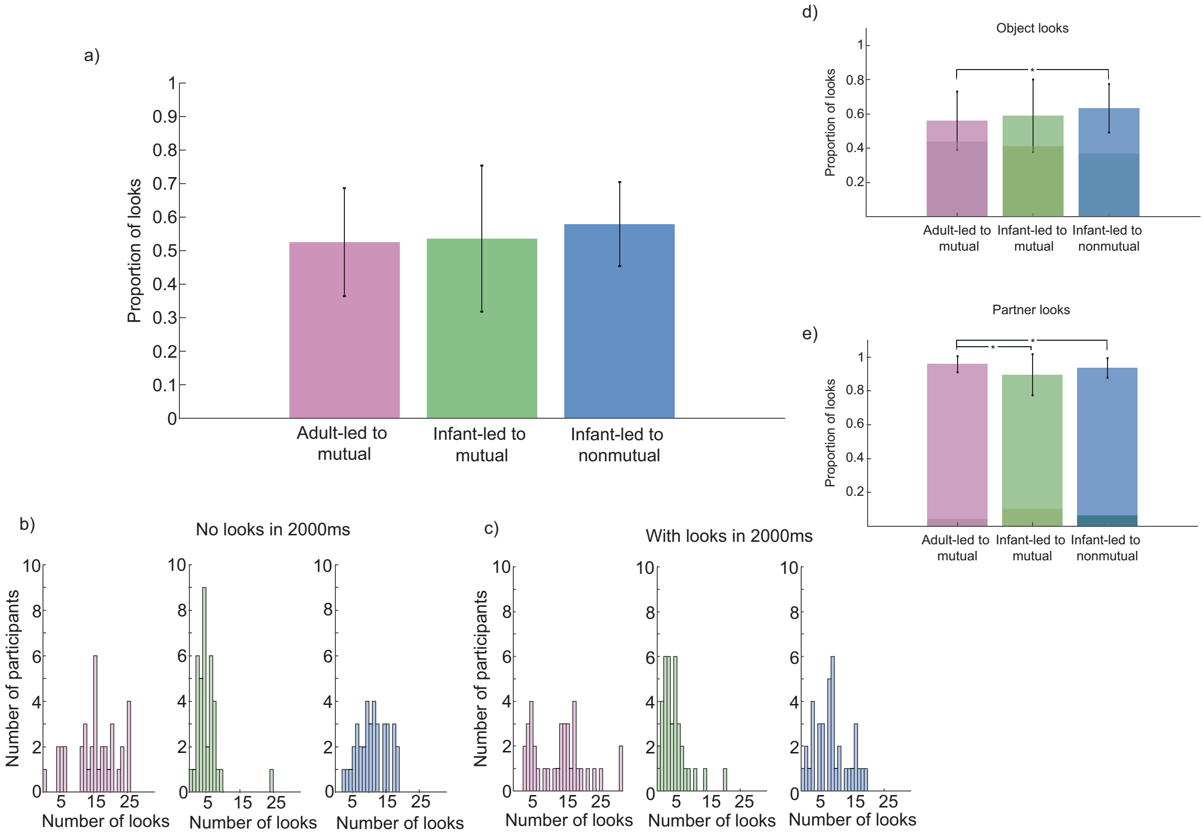

Figure S2. a) Bar plots show the proportion of looks without an object or partner look in the **2000ms preceding look onset** for each type of attention episode (error bars show the standard deviation (n=37)). Paired t-tests indicated no differences between attention episodes in the number of looks with an object or partner look occurring in the 2000ms before look onset (all $p>0.05$). Histograms show the number of trials that participants contributed without (b) and with (c) object or partner looks. d) and e) show the breakdown of proportions without object and partner looks, respectively; shaded areas show average proportions with looks for each attention episode (error bars show the standard deviation ($n=37$)). Paired t-tests showed that adult-led attention episodes have significantly fewer partner looks in the 2000ms before look onset compared to infant-led looks to mutual attention ($p<0.05$), and non-mutual attention ($p<0.05$). Adult-led looks also have fewer object looks compared to infant-led looks to nonmutual attention ($p<0.05$). All other comparisons ($p>0.05$).

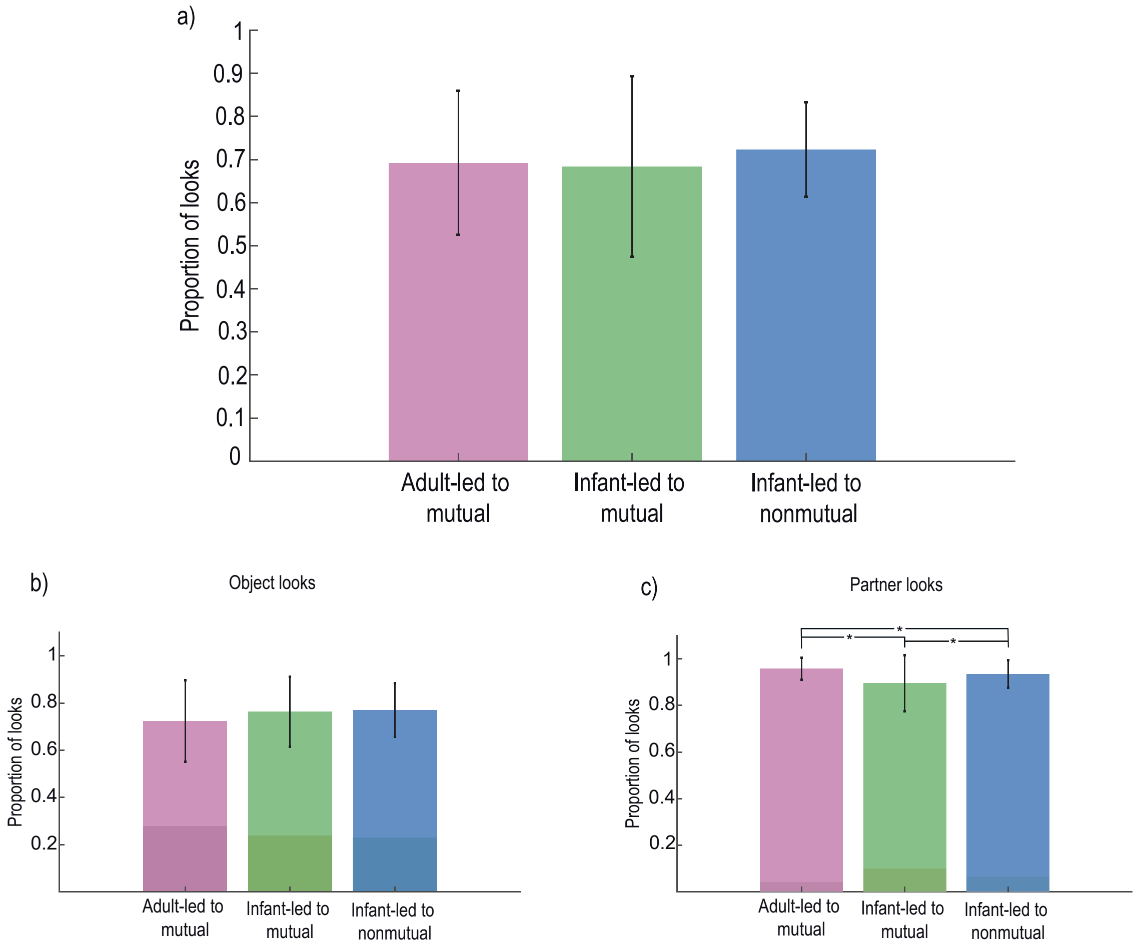

Figure S3 a) Proportion of looks without an object or partner look in the **2000-1000ms preceding look onset** for each type of attention episode (error bars show the standard deviation(n=37)). Paired t-tests indicated no differences between attention episodes (all p>0.05). b) and c) show the breakdown of proportions for object and partner looks, respectively, for each type of attention episode (error bars show the standard deviation(n=37)). Paired t-tests indicated no differences between attention episodes in the proportion of looks with object looks; partner looks differed between all three attention episodes (p<0.05).

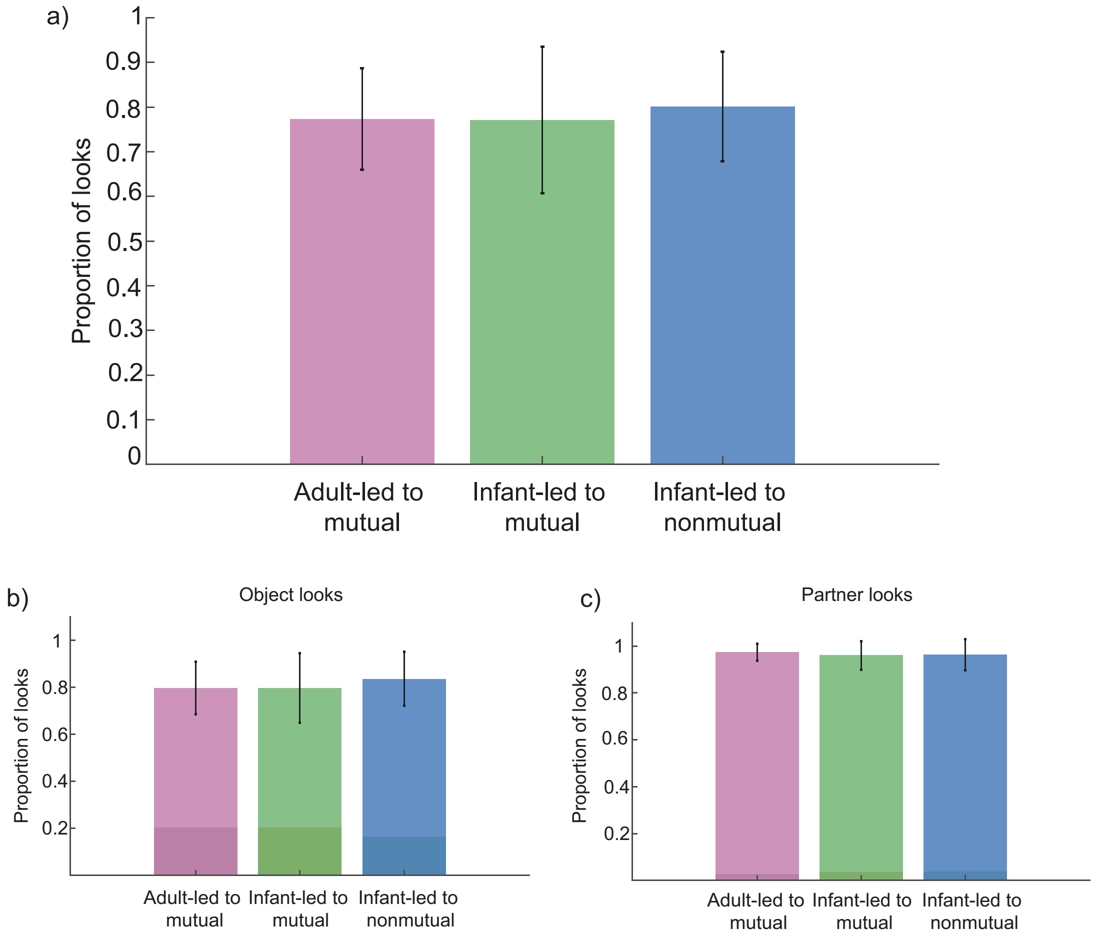

Figure S4 a) Proportion of looks without an object or partner look in the **1000-0ms preceding look onset** for each type of attention episode (error bars show the standard deviation(n=37)) .Paired t-tests indicated no differences in proportions between attention episodes (all p>0.05). b) and c) show the breakdown of proportions for object and partner looks, for each type of look (error bars show the standard deviation(n=37)). Paired t-tests again indicated that there were no differences between attention episodes in the number of object or partner looks (all p<0.05).

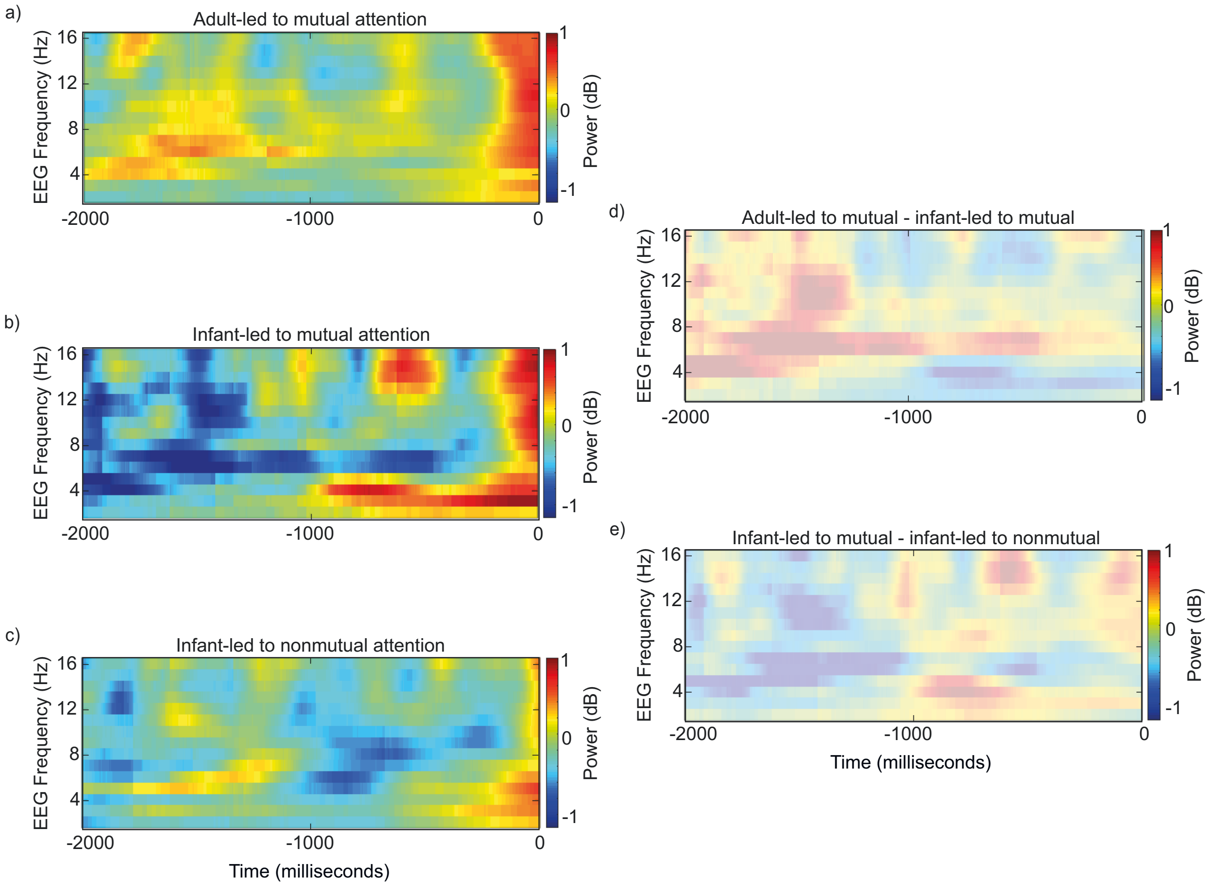

Figure S5 Secondary analysis excluding EEG activity in each look, where the infant is not focussed continuously on one object in the 2000ms before look onset. Time-frequency plots show infant EEG activity (2-16Hz) occurring 2000ms before look onset, for (a) adult-led looks to mutual attention (b) infant-led looks to mutual attention and (c) infant-led looks to nonmutual attention, over fronto-central electrodes (AF3, AF4, FC1, FC2, F3, F4, Fz). d) Difference in EEG activity between infant- and adult-led looks to mutual attention (adult-led – infant-led). Cluster-based permutation analyses showed no significant clusters of time*frequency points of the difference between attention episodes. e) Difference in EEG activity between infant-led looks to mutual and nonmutual attention (infant-led to mutual - infant-led to nonmutual). Cluster-based permutation analyses showed no significant clusters of time*frequency points of the difference between attention episodes.

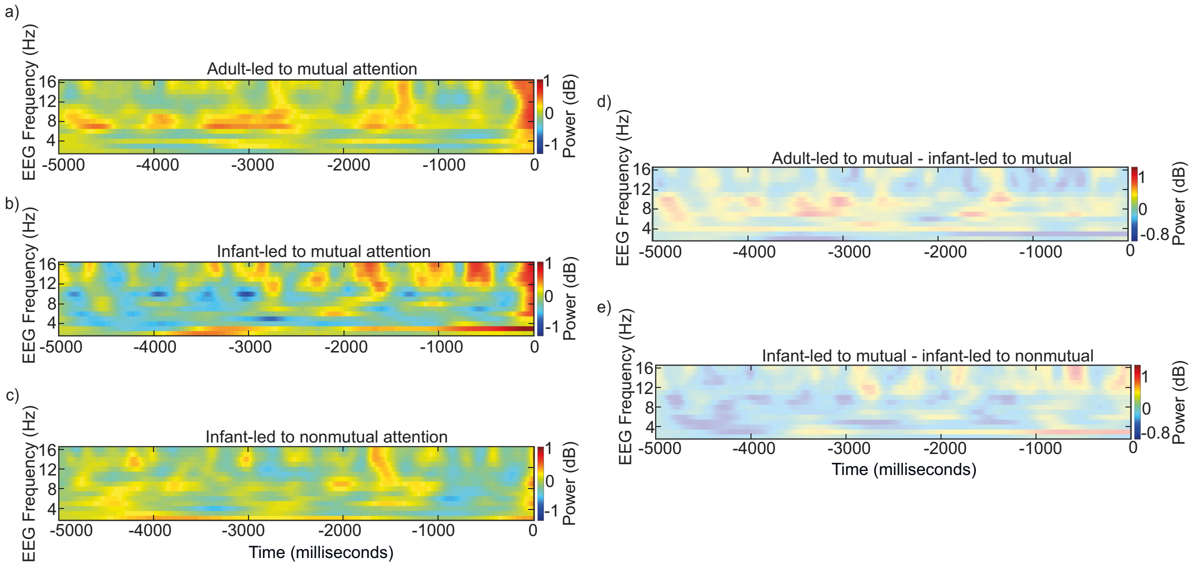

Figure S6 Time-frequency plots show infant EEG activity (2-16Hz) occurring 5000ms before look onset, for (a) adult-led looks to mutual attention (b) infant-led looks to mutual attention and (c) infant-led looks to nonmutual attention, over fronto-central electrodes (AF3, AF4, FC1, FC2, F3, F4, Fz). d) Difference in EEG activity between infant- and adult-led looks to mutual attention (adult-led – infant-led). Cluster-based permutation analyses showed no significant clusters of time*frequency points of the difference between attention episodes. e) Difference in EEG activity between infant-led looks to mutual and nonmutual attention (infant-led to mutual - infant-led to nonmutual). Cluster-based permutation analyses showed no significant clusters of time*frequency points of the difference between attention episodes.

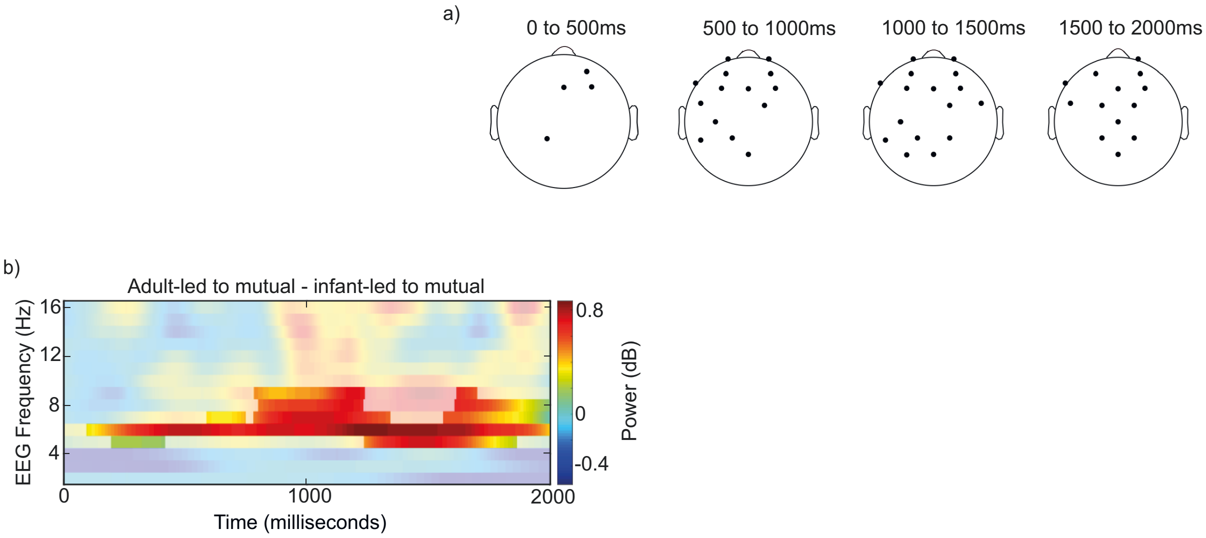

Figure S7 a) Time-course of the trend-level time*frequency*electrode positive cluster identified by the 3-dimensional cluster-based permutation analysis, in time*electrode space, in 500ms time-windows; channels included in the cluster at each time-point are highlighted in black. b) Time-frequency plot shows the difference between looks in time*frequency space averaged over electrodes included in the cluster (adult-led mutual – infant-led mutual). Highlighted area shows time-course of the cluster (p=0.099).

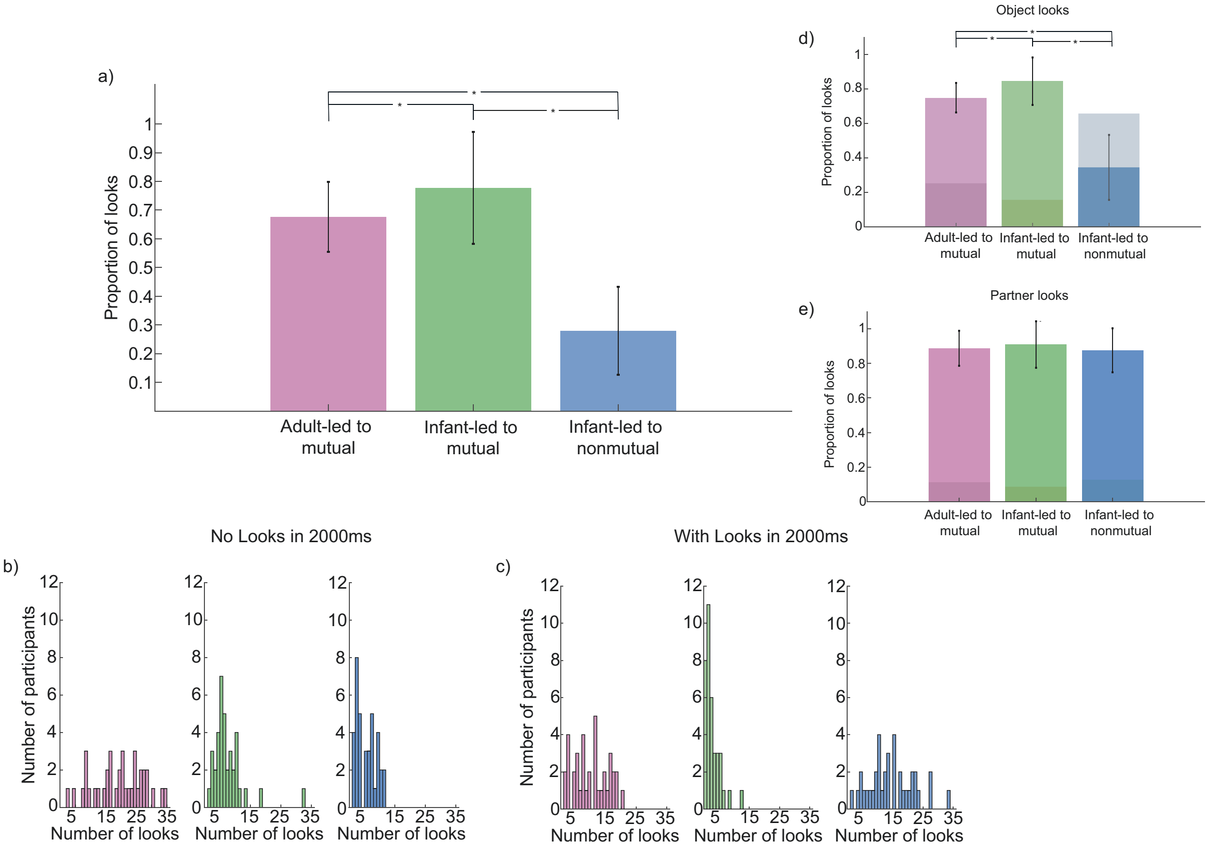

Figure S8 a) Bar plots show the proportion of looks without an object or partner look in the **2000ms post look onset** for each type of attention episode (error bars show the standard deviation (n=37)). Paired t-tests indicated significant differences between all attention episodes in the proportion of looks with object and partner looks combined occurring in the 2000ms after look onset (all $p<0.05$). Histograms show the number of trials that participants contributed without (b) and with (c) object or partner looks. d) and e) show the breakdown of proportions without object and partner looks, respectively; shaded areas show average proportions with looks for each attention episode (error bars show the standard deviation (n=37)). Paired t-tests showed that infant-led looks to nonmutual attention had significantly more looks with object looks in the 2000ms post look onset compared to adult-led looks and infant-led looks to mutual attention ($p<0.05$). Comparisons also showed that infant-led looks to mutual attention involved significantly fewer object looks, compared to adult-led looks ($p<0.05$). No differences in the number of partner looks were observed (all $p>0.05$).

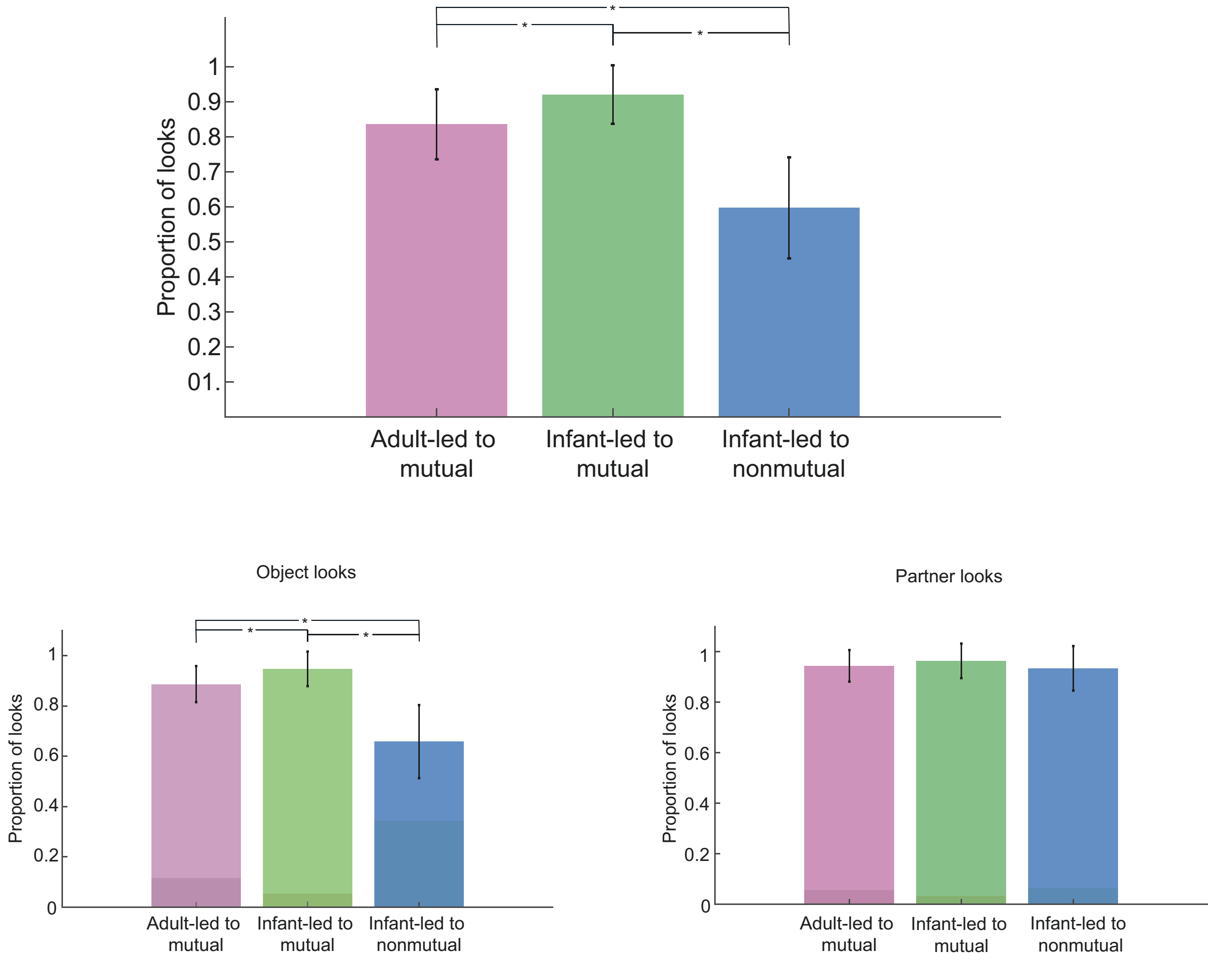

Figure S9 Proportion of looks without an object or partner look in the **0-1000ms post look onset** for each attention episode (error bars show the standard deviation (n=37)). Paired t-tests indicated significant differences between all attention episodes in the proportion of looks with object and partner looks combined occurring in the 1000ms after look onset (all p<0.05). b) and c) show the breakdown of proportions for object and partner looks for each attention episode. Paired t-tests showed that all looks have significantly different proportions of looks with object looks in the time-period 1000ms after look onset (all p<0.05). There were no differences in the proportion of looks involving partner looks (all p>0.05).

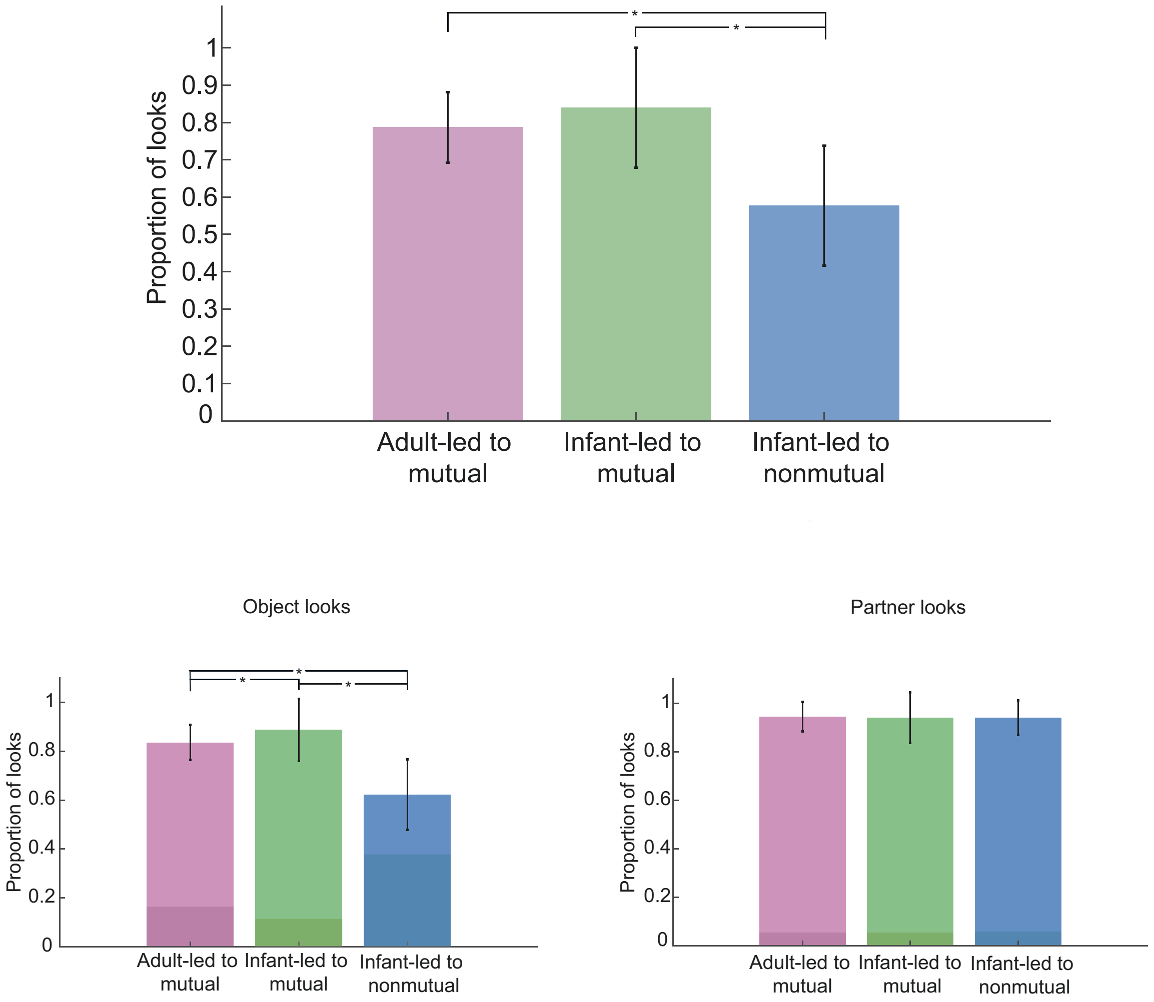

Figure S10 Proportion of looks without an object or partner look in the **1000-2000ms post look onset** for each attention episode. Paired t-tests showed that infant-led looks to nonmutal attention included significantly more looks with object and partner looks compared to infant-led looks to mutual attention, and adult-led looks ($p<0.05$). b) and c) show the breakdown of proportions for object and partner looks for each type of look. All looks have significantly different proportions of looks including object looks in the time-period 1000-2000ms after look onset (all $p<0.05$). There were no differences in the proportion of partner looks, $p>0.05$).

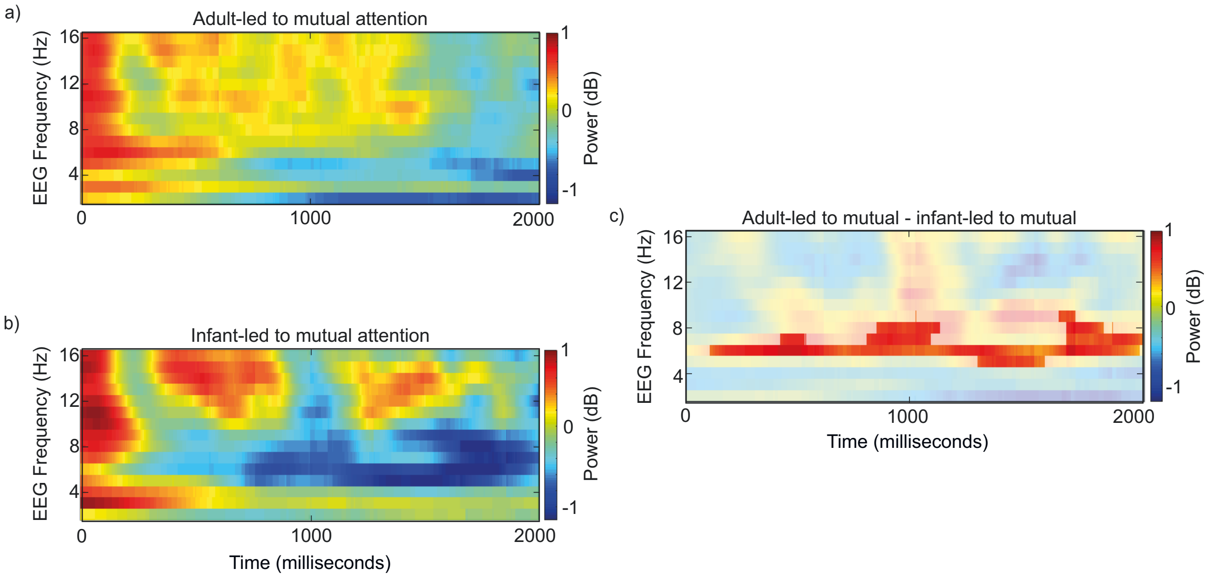

Figure S11 Secondary analysis excluding EEG activity in each look, where the infant shifts attention from the target object to another object/ the partner in the 2000ms after look onset. Time-frequency plots show infant EEG activity (2-16Hz) occurring 2000ms after look onset, for (a) adult-led looks to mutual attention and (b) infant-led looks to mutual attention, over fronto-central electrodes (AF3, AF4, FC1, FC2, F3, F4, Fz). c) Difference in EEG activity between infant- and adult-led looks to mutual attention (adult-led – infant-led); highlighted area shows the significant positive cluster identified by the cluster-based permutation analysis (p=0.01). The cluster ranges from 5-9Hz, from 104-1994ms post look-onset.

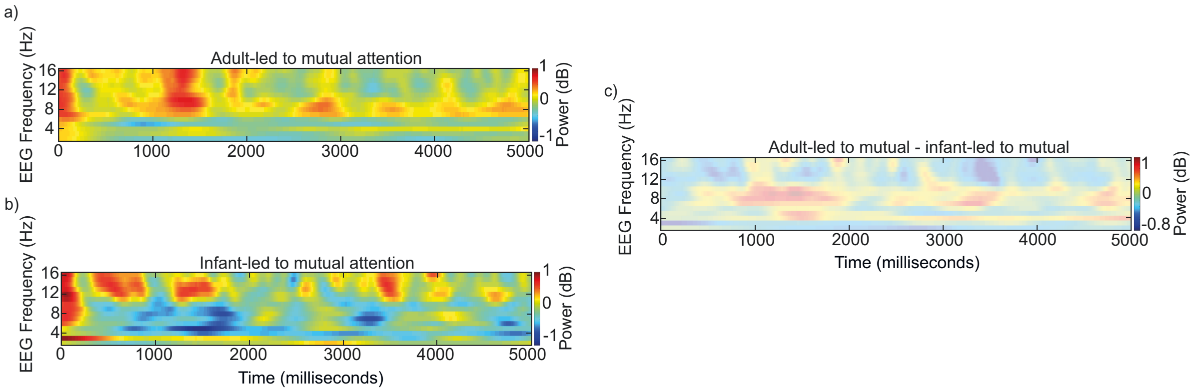

Figure S12 Time-frequency plots show infant EEG activity (2-16Hz) occurring 5000ms after look onset, for (a) adult-led looks to mutual attention and (b) infant-led looks to mutual attention, over fronto-central electrodes (AF3, AF4, FC1, FC2, F3, F4, Fz). d) Difference in EEG activity between infant- and adult-led looks to mutual attention (adult-led – infant-led). Cluster-based permutation analyses showed no significant clusters of time*frequency points of the difference between attention episodes.

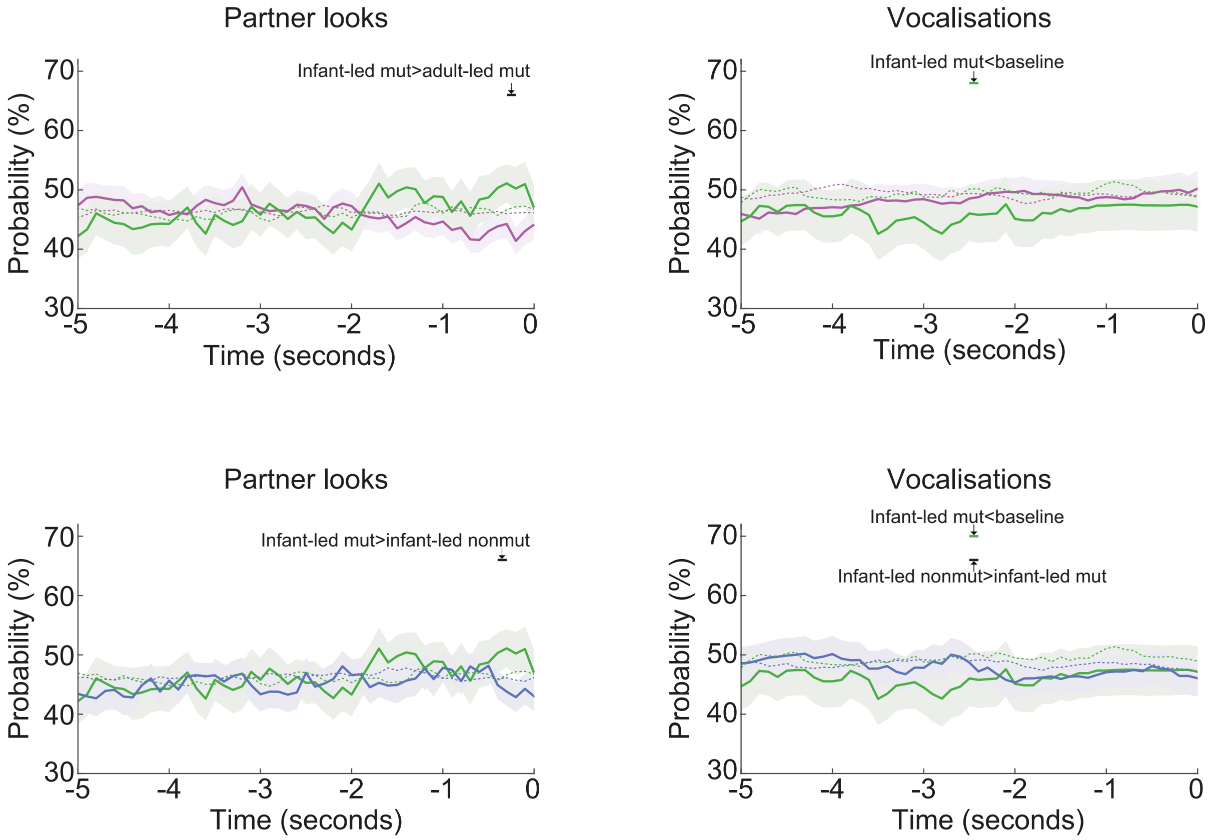

Figure S13 Probability of caregiver ostensive signals (partner looks and vocalisations) occurring in the time-period before attention episodes. a) Partner looks after infant-led mutual attention vs adult-led mutual attention; b) Vocalisations after infant-led mutual attention vs adult-led mutual attention. c) Partner looks after infant-led mutual attention vs infant-led nonmutual attention. d) Vocalisations after infant-led mutual attention vs infant-led nonmutual attention. In each case shaded areas show the SEM and horizontal black lines show the areas of significant difference, between attention episodes, identified by the cluster-based permutation analysis (Monte-Carlo p value<0.05). Dotted lines show the baseline time-series, plotted for each attention episode. Horizontal coloured lines show the areas of significant difference between each look type and baseline, identified by the cluster-based permutation analysis (Monte-Carlo p value<0.05).

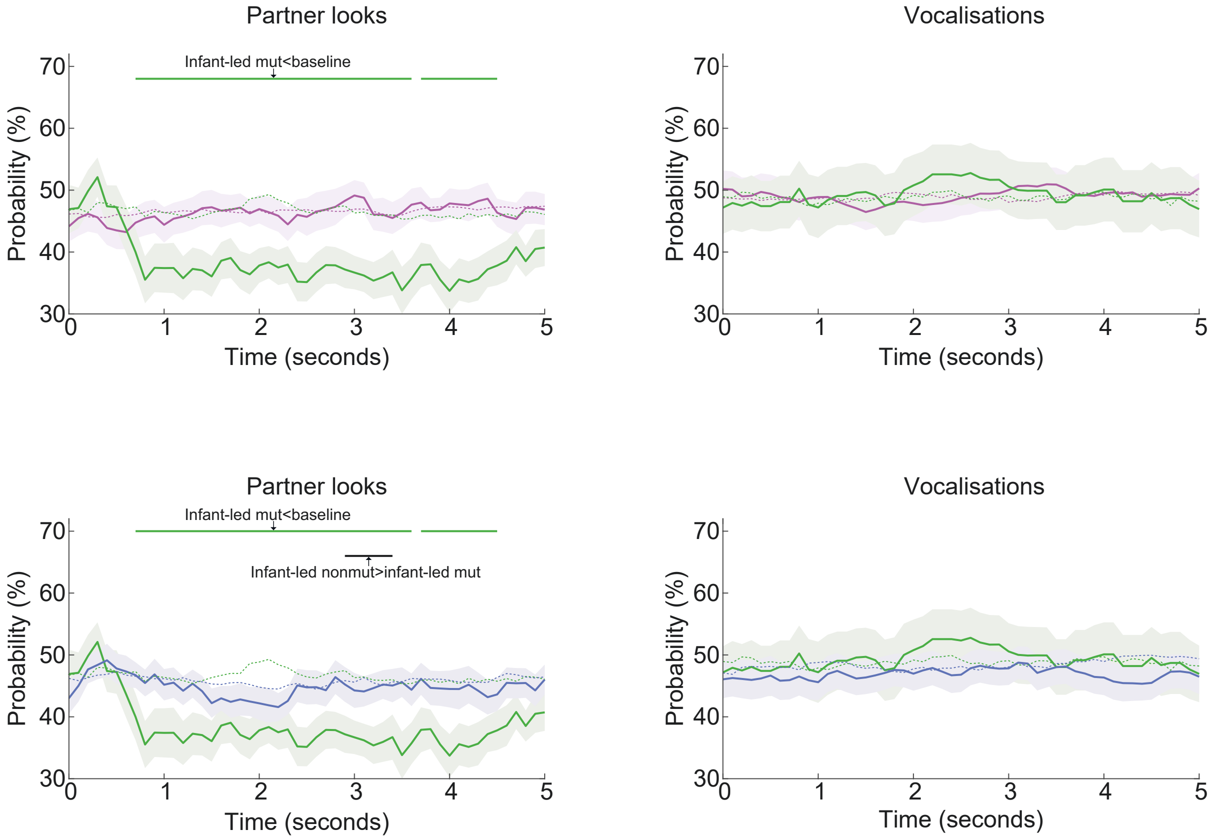

Figure S14 Probability of caregiver ostensive cues (partner looks and vocalisations) occurring in the time-period after attention episodes. a) partner looks after infant-led mutual attention vs adult-led mutual attention; b) vocalisations after infant-led mutual attention vs adult-led mutual attention. c) partner looks after infant-led mutual attention vs infant-led nonmutual attention. d) vocalisations after infant-led mutual attention vs infant-led nonmutual attention. In each case shaded areas show the SEM and horizontal black lines show the areas of significant difference, between attention episodes, identified by the cluster-based permutation analysis (Monte-Carlo p value<0.05). Dotted lines show the baseline time-series, plotted for each attention episode. Horizontal coloured lines show the areas of significant difference between each look type and baseline, identified by the cluster-based permutation analysis (Monte-Carlo p value<0.05).

APPENDIX B– Supplementary Materials for: Endogenous oscillatory rhythms and interactive contingencies jointly influence infant attention during early infant-caregiver interaction

1.1 Methods

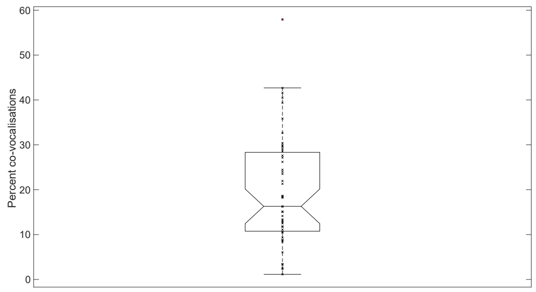

Figure S1 Percentage of caregiver vocalisations that were co-vocalisations. Box plot showing the percentage of caregiver vocalisations that were vocalisations across participants.

1.1.1 Clipping identification algorithm

The clipping algorithm was based on that outlined by Hansen et al. (2021). First, points in the speech signal reaching the maximum or minimum amplitude were identified. Next, to identify whether each max/min value was the beginning of a clipping event, the algorithm detected whether the value next to this point was 99.5% +/- of the max/min. A clipping event was considered to have ended where 3 consecutive values below/above the 99.5% threshold occurred. All vocalisations involving any clipping were excluded from analyses.

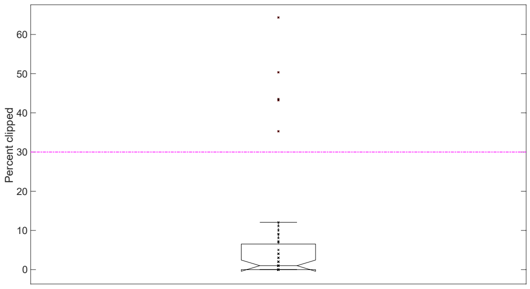

Figure S2 Percentage of clipped vocalisations. The bar plot shows the percentage of clipped vocalisations, for each participant. The pink horizontal line indicates the threshold at which participants were excluded from analyses (30%).

1.2 Results

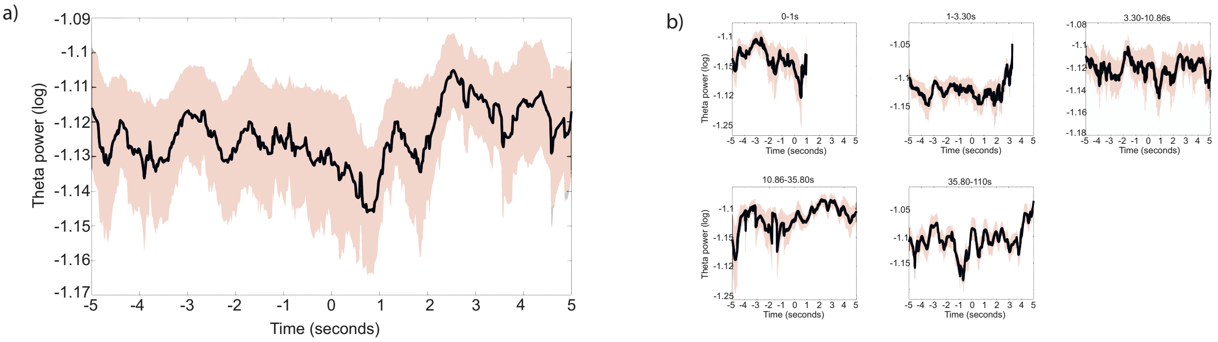

Figure S3 Event-related analysis for relative infant theta activity. a) Event-related analysis showing change in infant theta activity around infant attention onsets to objects: black line shows average relative theta activity (log); shaded areas indicate the SEM. Black horizontal line shows areas of significance revealed by the cluster-based permutation analysis (p <

0.05). Cluster-based permutation analysis revealed no significant clusters of time points (all p > 0.05). b) event-related analysis split by infant object attention-duration time bins. Black lines shows average infant relative theta activity (log); shaded areas indicate the SEM. Black horizontal lines shows areas of significance revealed by the cluster-based permutation analysis (p < 0.05). Permutation analysis again revealed no significant clusters of time points (all p >0.05).

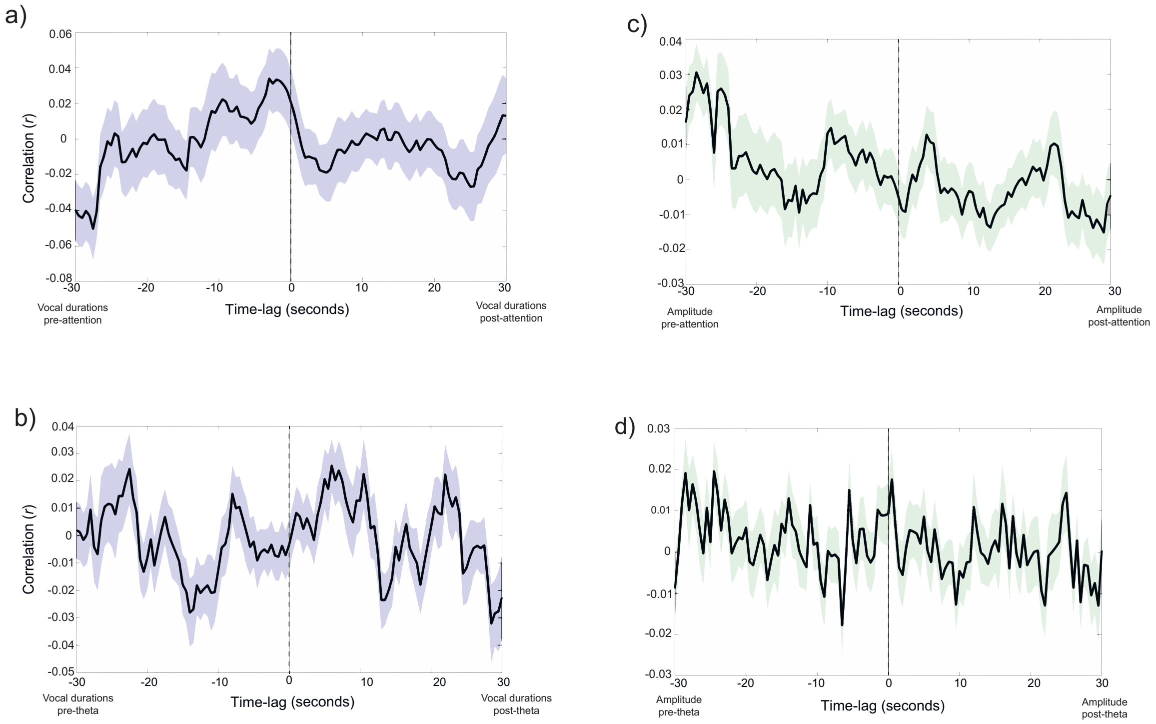

Figure S4 Assessing forwards-predictive associations between caregiver vocal behaviour, and infant attention, and endogenous oscillatory activity. Black lines show the Pearson cross-correlation between two variables; shaded areas indicate the SEM. Black horizontal lines show significant clusters of time lags (*p<0.05). First column shows the association between caregiver vocal durations and a) infant attention durations, and b) infant theta activity. Second column shows the association between caregiver amplitude modulations and c) infant attention durations and d) infant theta activity.

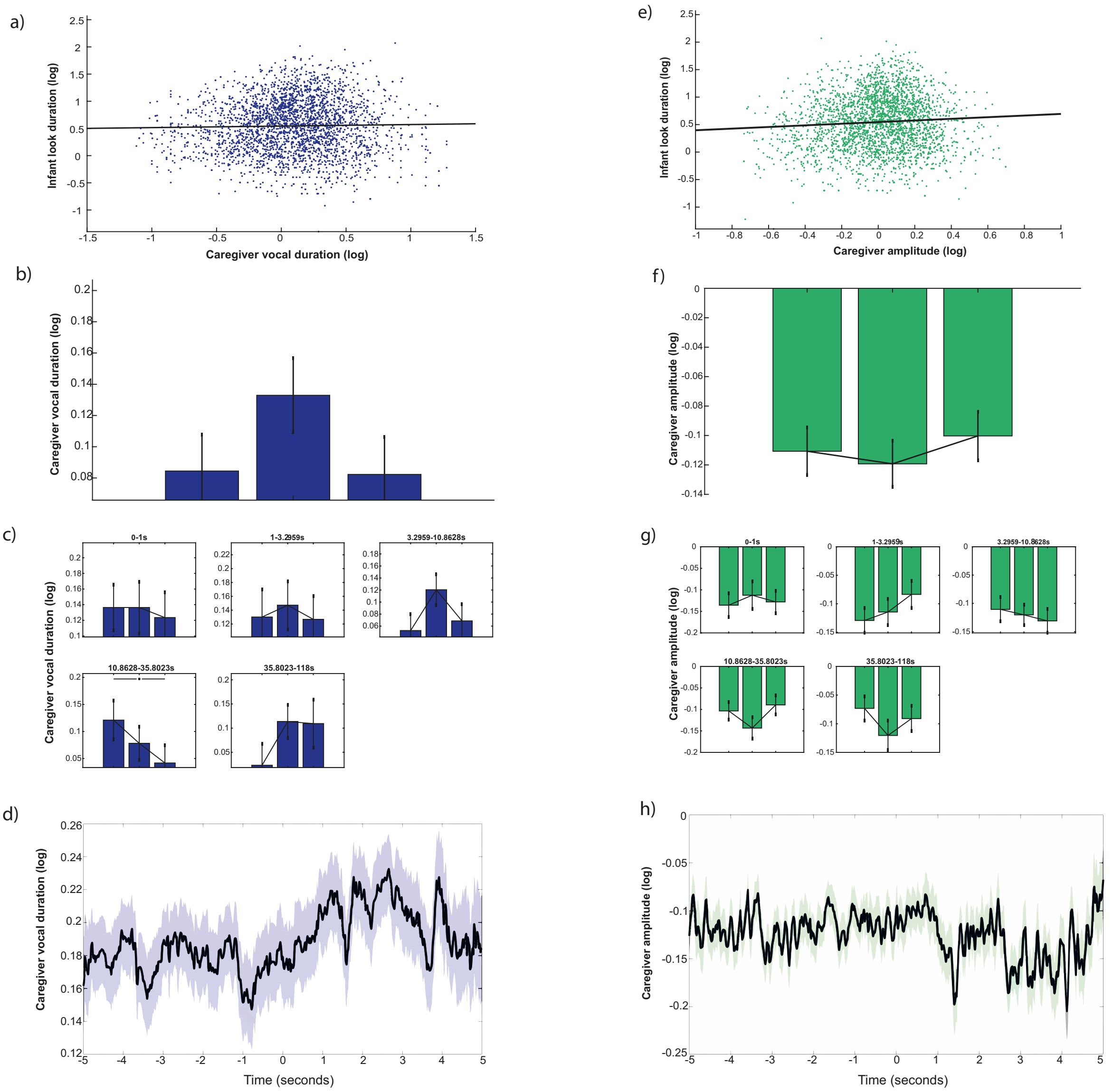

Figure S5 Reactive change in caregiver vocal behaviour relative to infant attention onsets. First row: scatter plots of the association between infant attention duration and a) caregiver vocal durations, e) caregiver amplitude modulations. Coloured dots show each infant object look; black line indicates linear line of best fit. Second row: attention chunk analysis for each infant object look for b) caregiver vocal durations, f) caregiver amplitude modulations. Each bar shows the median for each chunk across participants; errors bars show the SEM. Wilcoxon signed ranks tests explored significant differences between attention chunks (*$p < 0.05$). Third row: same as second row, binned by infant attention durations, for c) caregiver vocal durations, g) caregiver amplitude modulations. Bottom row: event related analysis to all infant object looks for d) caregiver vocal durations, h) caregiver amplitude modulations. Black line indicates the average across participants; shaded area indicates SEM.

APPENDIX C – Supplementary Materials for: Using continuous methods of analysis to examine speech-brain tracking during naturalistic caregiver-infant object play

1.1 Results

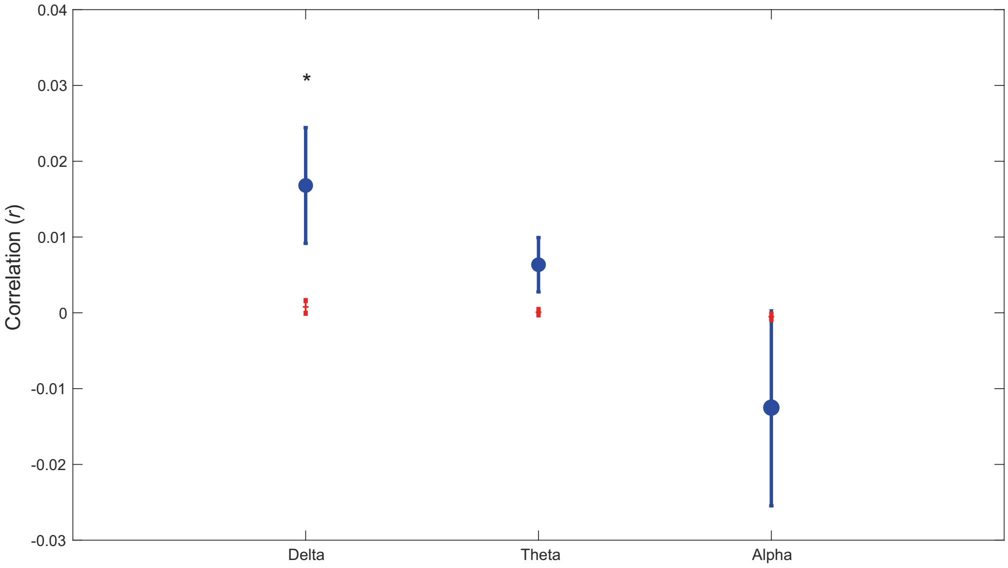

Figure S1 Grand average predicative accuracy values for individual continuous models before outlier removal, for each frequency band (Pearson's r). Blue circles show the mean across participants for each frequency band. Blue lines indicate the SEM of the averaged predictive accuracy values. Red horizontal lines show the averaged permutation values, averaged across permutations and participants. Red vertical lines indicate the SEM. Wilcoxon Signed Rank tests compared the observed r values to chance (*p<0.05).

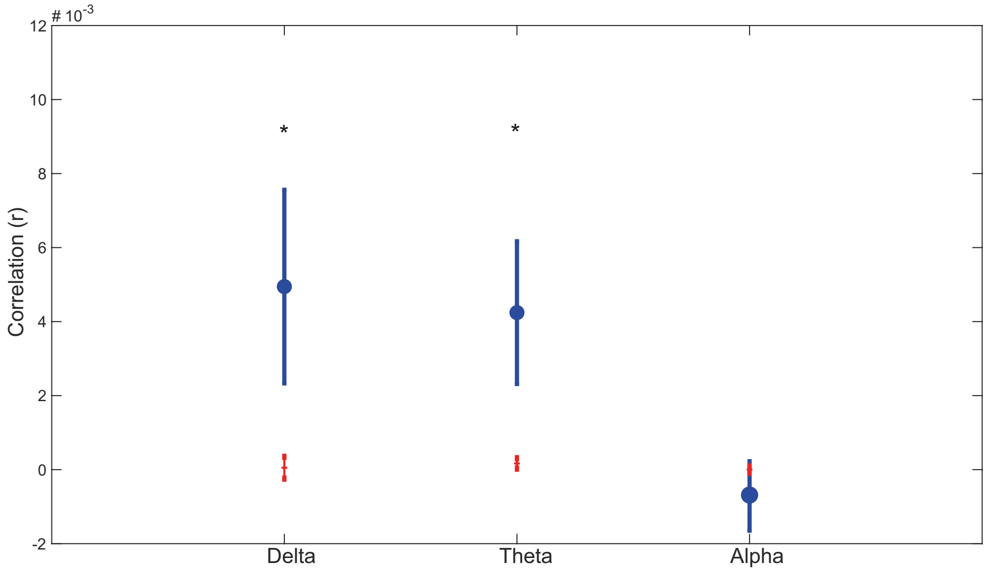

Figure S2 Grand average predictive accuracy values for generic continuous models before outlier removal, for each frequency band (Pearson's r). Blue circles show the mean across participants for each frequency band. Blue lines indicate the SEM of the averaged predictive accuracy values. Red horizontal lines show the averaged permutation values, averaged across permutations and participants. Red vertical lines indicate the SEM. Wilcoxon Signed Rank tests compared the observed r values to chance (*$p<0.05$).

APPENDIX D – Ethics

1.1 Ethical approval letter

The ethical approval letter received for our project is copied below. Note that our project was based on a joint application between myself, Megan Whitehorn and Ira Marriott-Haresign.

Ethics ETH1819-0141: Ms Megan Whitehorn (High risk)

Date	10 Apr 2019
Researcher	Ms Megan Whitehorn
Student ID	1920537
Project	The effects of Attention Training on Neural and Behavioural Responsivity in Infants During Parental Dyadic Play
School	Psychology

Ethics application

Checklist for research projects conducted during a pandemic

Project details

1.1 Is your research project taking place during a pandemic?

1. Project details

1.1. Title of proposed research or consultancy project
New insights into how the infant brain subserves dynamic social interactions

1.2. UEL Researchers or Consultancy lead'
Ms Megan Whitehorn

Ms Emily Phillips

Mr Ira Marriott Haresign

Prof Samuel Wass

1.3. Start date of project for which ethical approval is being sought
24 Jun 2019

1.4. Anticipated end date of project for which ethical approval is being sought
21 Jan 2022

1.5. If this project is part of wider research or consultancy work, please provide the UREC, EISC, URES, RRDE, SREC, CREB or NHS research ethics approval number
n/a

1.6. If this project is part of a wider research study or consultancy work please state the start and end dates
n/a

1.7. Specify where the research or consultancy project will take place
UEL, Stratford Campus

2. Aims and methodology

2.1. Aims and objectives of the project

Most early cognitive learning takes place in the presence of an adult social partner. Behavioural and psychophysiological research has suggested that as infants and parents communicate they adapt to each-other on a moment-by-moment basis- through, for example, cycles of vocalizing and pausing, of looking towards and away from each other, and of matching each-others' positive, negative and neutral affect (Feldman, 2007). Less well understood is whether similar mechanisms of mutual attunement also operate at the neural level, and if/how shared patterns of oscillatory activity between individuals sub-serve interpersonal influences on attention.

Using dual-EEG recordings, recent work has found that when interacting with an adult partner, increased neural synchronisation is observed during moments of direct gaze (Leong et al., 2017). A further study conducted with mothers and infants has also shown that when parents engage in joint play with their infant, parents' neural activity tracks and responds to infant attention, with greater responsivity associated with increased infant attentiveness.

The primary aim of the project is to further our understanding of interpersonal mechanisms that give rise to parent-infant neural "phase locking" (a phenomenon whereby two oscillatory signals become temporally aligned). Our secondary aim is to investigate the effects of child-focussed vs. parent-focussed intervention on parent-child behavioural synchrony, and how these effects are substantiated at the neural and physiological level. The child-focussed intervention group will undergo a computerised attention training intervention, previously shown to lead to increased voluntary attentional control and responsiveness to social cues (Wass et al.; Wass & Forssman, 2017). The parent-focussed intervention group will receive a parent-child book-sharing training programme. This intervention has been shown to increase parent responsivity to infant cues, leading to gains in child attentional control, as well as receptive and expressive language abilities (Valley et al., 2016). By comparing parent-child interactions in the two groups before and after intervention, we aim to identify casual mechanisms to behavioural, physiological (ECG) and neural (EEG) synchrony.

2.2. Methodology, data analysis and recruitment for the project

To assess neural phase locking, we shall record simultaneous Electrocardiography (ECG) and dual Electroencephalography (EEG) from N=210 typically developing 10-12-month-old infants and adults before and after intervention.

Participation in the study will involve attending 8 sessions, which will be scheduled weekly. The first 2 visits will be pre-intervention sessions; these will be identical for all participating infants. Then, the subsequent 5 visits will be intervention visits. The exact schedule for these will vary depending on which group the participant is in (see further details below). The last visit will be a post-intervention session. These will, again, be identical for all participants.

Allocation of participants to intervention group will be fully randomised and performed prior to the participants' first testing session.

Infants will be recruited from the participant database at UEL, as well as various local 'baby-groups', following the same procedures as used in multiple previous studies. Participants will also be recruiting from our pre-existing participant database. All participants will be aged between 270 and 330 days at the date of the first visit. Exclusion criteria will be: complex medical conditions, skin allergies, heart conditions, parents below 18 years of age, and parents receiving care from a mental health organisation or professional.

Pre-intervention assessment (visits 1 and 2)

The infant's parent will be with them at all times. During these pre-intervention visits, both parents and their parents will have their electrical brain activity measured using Electroencephalography (EEG) and Electrocardiography (ECG). EEG will be measured using a Biosemi 32-channel system optimised for dual EEG recording. This is a gel-based system that is already in use in multiple sites across the world for infant testing. ECG will be measured using stick-on electrodes placed in a modified lead II position.

Participants will complete three conditions, spread across two visits:

In Condition 1 (Video), infants will view a pre-recorded adult experimenter continually reciting nursery rhymes. Adult's pre-recorded EEG will be compared with the infant's live EEG during viewing. Behavioural data will be videoed and coded post-hoc.

In Condition 2 (Live), we shall record dual EEG and ECG from an infant and an unfamiliar adult (researcher) simultaneously, while they engage in table-top play. Behavioural data will be videoed and coded post hoc.

In Condition 3, the same procedure will be repeated while the infant interacts with their parent, we shall record dual EEG and ECG from an infant and and parent simultaneously, while they engage in table-top play. Behavioural data will be videoed and coded post hoc.

Participants will also be administered a battery of cartoon-based cognitive and language outcome measures using eyetracking. The eyetracker used will be a Tobii TX300, that has been used in numerous studies around the world. Parent questionnaires will also be administered, namely the: MacArthur Communicative Development Inventory; the Infant Behaviour Questionnaire (short version); the GAD-7 and PHQ-9 (clinical assessments of parental anxiety and depression symptoms); the Penn State Worry Questionnaire, and a demographics questionnaire. All questionnaires proposed for inclusion are given in the appendix. Sessions will last a total of 2 hours. This is a standard length of time for lab visits with infants of this age; although assessments will last around 30-60 minutes, we like to allow enough time for the infant to settle in, and have breaks (for naps, snacks etc).

Intervention 1- Child focused attention training intervention

This intervention will be delivered once a week over a five-week period, with each session lasting 30 (+/- 10) minutes. Each session will involve the child being positioned on their parent's lap in front of an eye-tracker, while training stimuli are presented on a computer monitor. Training stimuli consist of 6 tasks presented consecutively in the same order for each session; different events take place in the stimuli contingent on where the infant is looking. The intervention has already been published (Wass et al., 2011, 2018) and been used in several studies without any problems or unintended negative outcomes being reported.

Intervention 2 – Parent focussed book sharing intervention

This intervention trains parents in 'dialogic' book-sharing techniques; active and evocative behaviours that engage children in reciprocal communication during book-sharing. The intervention will be delivered once a week over five weeks by one of the PhD students on a one-to-one basis, with each session lasting 60 +/- 10 minutes. Each week will involve a specific theme, e.g. 'elaborating and linking', as well as a book-of-the-week for parents to take home and practice the book-sharing techniques with their children. The intervention has already been published (Vally et al., 2015) and been used in several studies without any problems or unintended negative outcomes being reported.

Intervention 3 – Control

Infant in the control group will make similar lab visits to the intervention groups; they will watch some computerised animations, and the researcher will discuss ways to incorporate foods into the child's diet that may be particularly helpful to supporting their early development.

Post Intervention

Following training, all participants will attend two post-assessments. These assessments will follow exactly the same format as the pre-assessment sessions. Participants will receive the same battery of pre-assessment outcome measures.

Through analysis we aim to explore interpersonal neural and physiological correlates of social interaction. We will compare parent-child behavioural, physiological and neural synchrony before and after intervention, in order to investigate the effects of receiving each intervention, and in comparison to no intervention at all.

2.3. Is the data accessed, collected or generated of a sensitive nature?
Yes

2.3.1. If yes, please provide details. Please ensure that all data of a sensitive nature is handled carefully and stored appropriately.
We will be collecting heart rate and EEG data. These will be stored securely on a password-protected drive kept under lock and key in the university and accessible only by project researchers.

3. About your project

3.1. Is the research/consultancy project funded?
Yes

3.2. Does the project involve external collaborators?
No

3.3. Does the project involve human participants?
Yes

3.4. Does the project involve non-human animals?
No

3.4.1. If yes, where is the research project taking place?

3.5. Does your project involve access to, or use of, material (including internet use) covered by the Terrorism Act (2006) and / or Counter-Terrorism and Border Security Act (2019) or which could be classified as security sensitive?
No

3.6. Does the project involve secondary research, secondary data or analysing an existing data set?
No

3.7. Does the project raise ethical issues that may impact on the natural environment over and above that of normal daily activity?
No

3.8 Does the research/consultancy project involve data collected online via social media, advertising the project online or via social media or include a questionnaire/survey?
No

If yes, please provide details.

3.9. Will the research/consultancy project take place overseas?
No

3.10. Will the researcher or research team be responsible for the security of all data collected in connection with the research/consultancy project?
Yes

3.11. Does your research/consultancy project require third-party permission?
No

If yes, please provide details.

3.12. Does your research/consultancy project involve any circumstances where the professional judgement of you and/or the team is likely to be influenced by personal, institutional, financial or commercial interests?

If yes, please provide details.

3.13. Does the project involve consultancy or contract research?

If yes, please provide details.

4. Funding

4.1. Funder(s)
Leverhulme Trust

4.2. Grant type
Research Council

If you selected other, please provide further details.

4.3. Value of grant
£ 327093

4.4. Please upload a letter advising of the award of the grant.

6. Recruitment

6.1. Are the research participants able to give informed consent (in written or verbal form)?
No

6.1.1. If no, is this because they are perceived to lack mental capacity or because they are vulnerable?
Vulnerable

6.1.2. If the participants are perceived to lack mental capacity, please provide the reason(s).

6.1.3. Further details

6.1.4. If the participants are perceived to be vulnerable, please provide details of the vulnerability.
Research will involve children aged 10-12 months at time of training and testing. In line with previous studies of this type undertaken by the university, parents will be asked to consent to each test proposed in this study. Written information about each test will be emailed to parents at least a week before their visit to the lab, and they will be encouraged to ask for any clarification they feel they need. All participants are invited to ask questions and voice concerns about our consent and information documents, so that we can respond to or expand on any part of the process that is not clear.

6.1.5. Does the research/consultancy project involve children or young people under the age of 16?
Yes

6.1.6 If yes, are the children or young people able to give informed assent?
No

6.1.7. If no, is this because they are perceived to lack mental capacity or because they are vulnerable?
Vulnerable

6.1.8. If the participants are perceived to lack mental capacity, please provide the reason(s).

6.1.9. Further details

6.1.10. If the participants are perceived to be vulnerable, please provide details of the vulnerability.
The infant participants will be aged between 10 and 12 months at the time of training and testing. At these ages, the babies will be too young to give informed consent before starting, therefore parents

will consent on behalf of their infants/children, as standard in this age cohort. In addition, they will be constantly monitored for signs of distress, and any procedure that is deemed upsetting to either parent, researcher or the baby will be halted. All participants will be informed that they have the right to halt proceedings at any stage. Babies and children will be constantly monitored, and any test will be discontinued should they become distressed or show signs of discomfort. This includes both training and testing sessions.

6.2. How will participants be recruited?

Infants will be recruited from the participant database at UEL, as well as from various recruitment drives at local baby-groups and Children's Centres. Fully informed consent will be obtained before testing commences.

All of the parents of babies in this study will have received an information sheet and been given an explanation of the aims of this study before providing their contact details. Different information sheets will be given depending on which group the participants have been pre-allocated to. When they are initially contacted, they will be given more detailed information about what participation in the study involves via an information sheet, which will be emailed to them. At each interaction, an opportunity will be given to ask any questions or gain clarification. Before any data is collected, the parent/carer will be asked to sign a consent form. All participants are invited to ask questions and voice concerns about our consent and information documents, so that we can respond to or expand on any part of the process that is not clear.

6.3. Please upload recruitment documents.

6.4. How many participants are being sought for the project?

210

6.5. How long will participants be required for the project?

The intervention blocks will take place over 9 weeks, with testing sessions once each week. The pre- and post- assessments will each last for two hours. Intervention sessions will last 30 +/- 10 minutes for the attention training and 60 +/- 10 minutes for the book-sharing.

6.6. Will the participants be remunerated for their contribution?

Yes

6.6.1. If yes, please specify monetary value of cash or giftcard / vouchers.

£50 in shopping vouchers for those who receive the interventions and £30 in vouchers for those who do not.

7. DBS

7.1. Do you require Disclosure Barring Service clearance (DBS) to conduct the research/consultancy project??

Yes

7.2. Is your DBS clearance valid for the duration of the research/consultancy project?

Yes

7.2.1. If you have current DBS clearance, please provide your DBS certificate number.
001659510034

8. Medical

8.1. Is your project a clinical trial and / or involves the administration of drugs, substances or agents, placebos or medical devices?
No

8.1.1. If you answered yes, please explain why you have chosen to use this application form instead of the NHS/HRA ethics application form. If you have selected yes, your project requires approval by the NHS/HRA, as it is falls under the classification of Medicines for Human Use (Clinical Trials) Regulations (2004) or Medical Devices Regulations (2002) or any subsequent amendments to the regulations.

8.2. Does your project involve the collecting, testing or storing of human tissue / DNA including organs, plasma, serum, saliva, urine, hair, nails or any other associated material?
No

8.2.1. If you answered yes, please explain why you have chosen to use this application form instead of the NHS/HRA ethics application form. If you have selected yes your project requires approval by the NHS/HRA, as it is falls under the classification of the Human Tissue Act (2004).

9. Risk

9.1. Does the project have the potential to cause physical or psychological harm or offence to participants and / or researchers?
No

9.1.1. If yes, please provide details of the risk or harm explaining how this will be minimised.

9.1.2. Please complete and upload a research risk assessment form
9.2. Does the project involve potential hazards and/or emotional distress?
Yes

9.2.1. If yes, provide an outline of support, feedback or debriefing protocol.
Infants sometime express mild distress on the application of the EEG equipment. However, we have received extensive training in how to minimise this. We do not obtain any usable EEG recordings if infants are at all distressed, so we take every possible step to minimise this when it occurs. The process of recording EEG data from infants is in place in numerous other research labs across the world, and the process we follow will be identical to those used in other labs.

9.3. Provide an outline of any measures you have in place in the event of an adverse event or reaction or unexpected outcome, the potential impact on the project and, if applicable, the participants.

We do not expect adverse events arising from the study procedures. Under circumstances of an unexpected adverse event, the participants’ health and safety will be the highest priority. If health and safety is in any way compromised, the participant will be withdrawn from the study, with clear reasoning given.

10. Anonymisation

10.1. Will the participants be anonymised at source?
No

10.1.1. If yes, please provide details of how the data will be anonymised.

10.2. Are participants' responses anonymised or are an anonymised sample?
No

10.2.1. If yes, please provide details of how the data will be anonymised.

10.3. Are the samples and data de-identified?
Yes

10.3.1. If yes, please provide details of how the data will be anonymised/pseudonymised.
Direct and indirect identifiers will be removed from data and participants will be assigned a participant code. This will be entered along with the date and time of testing on the files containing raw data, and will be used to record all other data collected during pre-post assessments. Participant names and ID numbers will be stored in a separate password protected database. This is so that participants can withdraw their data up to the point at which it is included in the final analysis. Consent forms will be kept securely and separately from the raw data in a locked cabinet. Only members of the research team will have access to both the raw data and consent forms. Data will be retained in a secure place at the end of the project as, should funding allow, we might wish to follow up the sample over a longer time period. Video recording of the behavioural paradigms will only be viewed by members of the research team.

10.4. Please provide details of data transcription.
All data will be analysed using MATLAB software. Behavioural paradigms will be coded according to their manuals. External research assistants may be recruited in order to code videos; assistants will be fully briefed on ethics related to the study and will not be allowed access to consent forms, code keys, or any other data that would allow them to identify the participant. Research assistants will only be permitted to code data on university computers (that are disconnected from the internet) and will not be allowed to copy or move video files from the university. Research assistants will be asked to complete a confidentiality agreement before undertaking any work on the project.

10.4.1 Will the data be transcribed by person(s) outside of the project team?

If yes, please upload a blank copy of the confidentiality agreement.

10.5. If applicable, will all members of the project team know how the code links the data to the individual participant?
Yes

10.5.1. If no, in the event of a researcher's absence please specify the process should access to the research data be required.

10.6. Will participants be anonymised/pseudonymised in publications that arise from the research/consultancy project?
Yes

10.6.1. If no, please provide details.

10.7. Will participants have the option of being identified in the study and dissemination of research findings and / or publication?
No

10.7.1. If yes, please provide details.

16. Data security

16.1. Will the researcher or the project team be responsible for the security of all data collected in connection with the proposed research/ consultancy project?
Yes

16.1.1. If no, please provide details.

16.2. Will the research/consultancy data be stored safely on a password protected computer?
Yes

16.2.1. If no, please provide details.

16.3. Will the research/consultancy project data be stored on a UEL data managed device?
Yes

16.3.1. If no, please specify where the electronic data will be stored and how the data will be kept secure.

16.4. Will you keep research/consultancy project data, codes and identifying information in a separate location?
Yes

16.4.1. If yes, please explain how you will store the research data.
Consent forms will be kept in hard copies in a locked cupboard on university premises. Only the research team will have access to this cupboard. The cupboard will be locked when not in use. Data and audio/video recordings will be stored separately from the consent forms. Hard copies of personal

data will be stored in locked cupboards in the lab. Electronic versions of personal data will be stored on a secure computer network (to which only the research team have access) on password protected computers at the University of East London. Participant names and IDs will be stored on a separate password-protected database.

16.5. Will the raw data be shared with individuals outside of the project team?
No

16.5.1. If yes, please specify the names, positions and their relationship to the research/ consultancy project

---

Name

Position

Relationship to research

16.6. Will participants be audio and/or video recorded?
Yes

16.6.1. If yes, please explain how you will transfer, store and, where relevant, dispose of audio and/or video recordings.
Digital audio-video recordings will be transferred onto the UEL secure computer network that only the research team will have access. Video files may need to be stored externally; in this case, recordings will be stored and transferred on a password-encrypted hard drive with access limited to members of the research team.

16.7. If audio and/or video recordings will be retained, please provide details and state how long the recordings will be kept.
Recording will be kept for up to a year after the data has been analysed unless further funding can be obtained and the data is used again in further analyses, for which ethics approval of analyses of secondary data will be sought.

16.8. Will you retain hard copies of the data?
Yes

16.8.1. If yes, please provide details of how the data will be transported safely and, where relevant, undergo secure disposal.
Data will be collected and stored in the BabyLab at UEL. Hard copies of any data will be kept in locked cabinets in the BabyLab offices. Only members of the research team will have keys to those cabinets. The data will not leave the premises. These will be shredded by a team member when appropriate.

16.9. Will the research/consultancy project data be encrypted and transferred inside of the UK?

Yes

16.10. How long will the research data that details personal identifiers be stored?
Any personal identifiers will only be retained until the research has been published and it is no longer possible for participants to request to withdraw their data

16.11. Please upload a copy of your Data Management Plan.

18. Dissemination

18.1. Will the results be disseminated?
Yes

18.1.1. If yes, how will the results of the research/consultancy project be reported and disseminated?
Dissertation / Thesis
Peer reviewed journal
Internal report
Conference presentation

18.1.2. If you selected other, please provide further details.

18.1.3. If the results of the research/consultancy /project will not be reported and disseminated, please provide a reason

20. Attachments

You can generate a Participant Information Sheet and Consent Form using the answers provided in your ethics application form. The Word files generated can be edited. You should upload the final version(s) before submitting your application form.

20.1. Upload any additional files to support your application which have not already been uploaded within your application.