Narrative: The Temporal Architecture of Adult-Infant Interaction

Timothy McGowan

A thesis submitted in partial fulfilment for the degree of

Doctor of Philosophy

Institute of Education

Faculty of Humanities and Social Sciences

University of Strathclyde, Glasgow

April, 2024

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Chapters 1 and 2 from this thesis contain already published material that was majority written by the author with intellectual guidance and editing provided by Prof. Jonathan Delafield-Butt.

Signed: Timothy McGowan

Date: 1st April 2024

Abstract

This thesis examines the temporal architecture of adult-infant interactions (i.e. the manner in which an interaction unfolds over time). In particular, we explore the existence and nature of a common narrative temporal framework underpinning adult-infant interaction, consisting of phases of arousal and intensity split into four distinct states: introduction, development, climax and resolution. This framework is considered a fundamental structure of human cognition, and to be central in human communication and learning. We began by exploring the current state adult-infant interaction research, and identified that recent work largely fails to consider an underlying narrative element. We then sought to address this ‘narrative gap’ through theoretical, empirical and methodological contributions. We first applied narrative theory to the neonatal imitation paradigm, viewing the imitative exchange between experimenter and infant as being inherently dialogical in nature. On this basis we proposed that underlying successful displays of imitation by neonates was a narrative framework. We then explored the development of narrative through infancy by conducting a longitudinal examination of mother-infant interactions when infants were aged 4 months, 7 months and 10 months. We hypothesised that the duration of infant positive affect would be a function of the narrative phase reached in an interaction, with older infants engaging in longer interactions that reached more advanced narrative phases. Our results supported these predictions (except for interaction durations that decreased with infant age). Finally, we outlined a methodological pipeline to automate the identification of narrative phases in adult-infant engagements. We first described the training and evaluation of a deep learning based markerless motion tracking model specifically tailored for the tracking of adult and infant movement during dyadic engagements. We then proposed a machine learning analysis pipeline for the clustering of this movement data according to narrative phases, thus removing the potential for human bias in the identification of narrative.

Table of Contents

Tables and Figures

General Introduction

Dyadic interactions during infancy are a child's first experience of social engagement. Within these shared experiences are laid the foundations of social and emotional development, with the ebb and flow of intensity and arousal allowing for the creation of affective meaning making (Delafield-Butt & Trevarthen, 2015; Malloch, 1999; Malloch & Trevarthen, 2009). This thesis explores these interactions through infancy, making theoretical, empirical and methodological contributions to the current state of research in the field.

In particular we are interested in the 'temporal architecture' of early dyadic engagements. We define temporal architecture as the manner in which an interaction is organised through time, or how levels of intensity and arousal vary and develop across the duration of an engagement. This is conveyed in the multi-modal (including movement, vocalisation and expression etc) reciprocity of human communication. More specifically, we aim to identify and explore the presence of narrative. Narrative is the underpinning theory of this thesis, and is described and explored extensively in both Part 1 and Part 2. We all have a foundational appreciation of the power of narrative through the art, literature, and culture we have been exposed to since early childhood. However, narrative is also fundamental to human experience and existence. It has given rise to core theories regarding human intelligence, development, and evolution (e.g. Cobley, 2013; Dautenhahn, 2001; Delafield-Butt & Trevarthen, 2015; Delafield-Butt & Trevarthen, 2013; Gallagher, 2014; Hutto, 2007;; Trevarthen & Delafield-Butt, 2013; Zahavi, 2007).

Very briefly, narratives consist of recurring patterns involving interest, arousal, affect, and intention, resulting in a distinctive four-part framework comprising an introduction, development, climax, and resolution. When shared with others, these narratives become collective acts of meaning creation, punctuated by discrete beginnings and endings that delineate a consistent pattern, thereby framing social engagements and solitary projects.

1.0 The Thesis

The contribution of this thesis will be three-fold. In Part 1 of the thesis, we advance our theoretical contribution. Part 1, Chapter 1 will provide a review of the current state of the literature with regard to adult-infant interaction, and we demonstrate the clear gap that exists with regard to narrative theory. Part 1, Chapter 2 begins to address this gap, by making a theoretical application of narrative theory to the highly debated phenomenon of neonatal imitation. In Part 2, Chapter 3, we present our empirical contribution to the field with a longitudinal mapping (using human rater coding) of mother-infant interactions at 3 time points in infancy. Finally, in Part 3, Chapters 4 and 5, we make a methodological contribution to the field, outlining an approach for the markerless tracking of movement in video footage of adults and infants and a novel machine learning analysis for the automatic identification and clustering of narrative phases.

1.1 Part 1

1.1.1 Chapter 1

Part 1, Chapter 1 explores the influence of non-verbal, embodied narratives during infancy, highlighting their role in shaping shared experiences, emotional regulation, and early learning before the acquisition of language. By examining these narratives within both individual projects (including simple movement orientated projects such as reach to grasp, and more complex projects such as building a sand castle) and social contexts, our objective was to enhance our comprehension of narrative as a foundational temporal framework for early child development.

However, this chapter also highlights the gap that currently exists in infant interaction research, and child development research more widely - this being that many current studies into co-regulation and adult-infant interaction fail to acknowledge an underlying narrative dimension. Fortunately, recent scholarship has begun to recognize the infant's prospective agency and their pivotal role in shaping interactions with their environment and caregivers. This shift in perspective is particularly evident in research exploring adult-infant synchrony, entrainment, and co-regulation, encompassing both physiological and psychological dimensions. By elucidating the functional and temporal dynamics of these fundamental communication processes, we argue it is possible to gain valuable insights into broader human development and contribute novel perspectives to ongoing scholarly discourse.

1.1.2 Chapter 2

In Part 2, Chapter 2 we begin to address the gap identified in Chapter 1 relating to the absence of narrative from chid development research areas and paradigms. To this end, we produced a theory paper applying narrative theory to the earliest experimental example of adult-infant interaction - neonatal imitation. Despite the number of publications associated with neonatal imitation, there has been a notable neglect of the temporal framework within which this phenomenon unfolds. To address this, we apply narrative theory and posit that the efficacy of neonatal imitation is contingent upon an inherent narrative structure predominantly expressed and comprehended through bodily movements (as well as other modalities of communication). We propose that the variety of conflicting results in empirical investigations of neonatal imitation is the result of the successful or unsuccessful development of narrative architecture within the imitation experimental paradigm. By building on the work of Kugiumutzakis and Trevarthen (2015), Nagy (2006) and Nagy and Molnar (2004), and viewing neonatal imitations as an early form of dialogue between adult and infant, founded upon the same narrative underpinnings as other initial infant interactions, we perform a re-examination of prominent theories concerning neonatal imitation in light of narrative theory.

1.2 Part 2

1.2.1 Chapter 3

In Part 2, Chapter 3 we present the core empirical work of the thesis. This original research represents the first-time narrative has been mapped in adult-infant interactions across the first year of life. To this end we utilise longitudinal recordings of mothers and infants engaging in naturalistic interactions, and perform a microanalysis coding for the presence of an infant's visual attention to their mother, expressions of positive affect, duration of engagement, and discernible phases of narrative structure. This work unveils a notable trend wherein interactions progressively shorten with advancing infant age whilst more often reaching the later phases of the narrative cycle (the climax and resolution phases) (thus indicating older infants could achieve the full complexity of a complete narrative engagement in a shorter space of time). Additionally, we identify that engagements reaching a climatic peak or resolution phase resulted in longer durations of positive affect. This chapter adds to existing evidence that infants possess cognitive and psychological sensitivity and capability for the creation of complete narratives with climactic and resolved experiences as early as 4 months. We provide the first evidence of how these narrative engagements change and evolve across infancy.

1.3 Part 3

1.3.1 Chapter 4

In Part 3 of this thesis, we make our contribution towards the advancement of the methodological approaches for the identification of narrative. Through the development of our approach, we sought to address the common criticism of human identification of narrative that, as inherently narrative beings, humans will be biased to finding narrative even if no narrative exists. To do this, in Chapter 4 we focus on the tracking and measurement of movement within adult-infant interaction and explore the application of markerless computer vision tracking. Utilising videos from the same archive that formed the basis of our analysis in Chapter 3, we train a deep learning model (using open-source software, DeepLabCut) to create pose estimates for both adults and infants during interactions. We describe in detail the process for this, and report the accuracy of the final model.

1.3.2 Chapter 5

Using the pose estimation model we trained in Chapter 4, we analyse the videos that formed the basis of our analysis in Chapter 3. This produced a dataset made up of pose estimates for each frame of video footage relating to chosen body parts of infants and adults. In Part 3, Chapter 5 we outline a proposed machine learning approach for the identification and clustering of narrative phases utilising this dataset. We describe in detail an unsupervised analysis pipeline utilising a k-means clustering algorithm. This represents an entirely novel methodology for the identification of narrative phases, and opens the possibility for a completely automatic approach that removes the need of human coded input.

2.0 Contributions for the Future

Through this thesis we aim to advance general understanding of the form, effect and development of narratives during infancy. This work also leaves fertile ground for future research and development. This is discussed in detail in the general discussion (as well as throughout the thesis more generally), but we hope this work forms part of a broadening area of narrative research. We believe this can encompass fields as diverse as early years education, developmental psychology, philosophy of mind, machine learning, comparative psychology and primatology; thus demonstrating the extensive impact and broad importance of narrative theory to the human experience.

References

Cobley, P. (2013). *Narrative* (2nd ed..). New York : Routledge.

Dautenhahn, K. (2001). The narrative intelligence hypothesis: In search of the transactional format of narratives in humans and other social animals. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, *2117*, 248–266. https://doi.org/10.1007/3-540-44617-6_25

Delafield-Butt, J. T., & Trevarthen, C. (2015). The ontogenesis of narrative: from moving to meaning. *Frontiers in Psychology*, *6*(September), 1–16. https://doi.org/10.3389/fpsyg.2015.01157

Delafield-Butt, J., & Trevarthen, C. (2013). Theories of the development of human communication. In P. Cobley & P. J. Schulz (Eds.), *Theories and models of communication* (pp. 199–222). Mouton de Gruyter.

Gallagher, S. (2014). Narrative Competence and the Massive Hermeneutical Background. *Education, Dialogue and Hermeneutics*. https://doi.org/10.5040/9781472541123.ch-002

Hutto, D. D. (2007). Narrative and Understanding Persons. *Royal Institute of Philosophy Supplement*, *60*(August), 1–16. https://doi.org/10.1017/s135824610700001x

Kugiumutzakis, G., & Trevarthen, C. (2015). Neonatal Imitation. In *International Encyclopedia of the Social & Behavioral Sciences: Second Edition* (pp. 481–488). https://doi.org/10.1016/B978-0-08-097086-8.23160-7

Malloch, S. N. (1999). Mothers and infants and communicative musicality. *Musicae Scientiae*, *3*(1_suppl), 29–57. https://doi.org/10.1177/10298649000030s104

Malloch, S., & Trevarthen, C. (2009). Musicality: Communicating the vitality and interests of life. In *Communicative musicality: Exploring the basis of human companionship.* (pp. 1–11). Oxford University Press.

Nagy, E. (2006). From imitation to conversation: The first dialogues with human neonates.

Infant and Child Development, *15*, 223–232. https://doi.org/10.1002/icd.460

Nagy, E., & Molnar, P. (2004). Homo imitants or homo provocans? Human imprinting model of neonatal imitations. *Infant Behavior and Development*, *27*(1), 54–63. https://doi.org/10.1016/j.infbeh.2003.06.004

Trevarthen, C. (n.d.). Movement: Its Awareness of the World, and Its Responsibilities in Community. *Journal of Cognitive Semiotics, IV*, *1*, 6–56. http://www.cognitivesemiotics.com.

Trevarthen, C., & Delafield-Butt, J. (2013). Biology of shared experience and language development: Regulations for the intersubjective life of narratives. In *The infant mind: Origins of the social brain.* (pp. 167–199). The Guilford Press.

Zahavi, D. (2007). Self and Other: The Limits of Narrative Understanding. *Royal Institute of Philosophy Supplement*, *60*, 179–202. https://doi.org/DOI: 10.1017/S1358246107000094

Part 1

Prologue

The first steps of my doctoral research walked a very different path to that described in the forthcoming pages. In a pre-Covid world, I had planned to explore the movement of infants in an attempt to identify a motion signature in neonates at increased risk of developing autism. This project was to be conducted through the maternity and neonatal units in the Central Hospital de Funchal, Madeira. Despite significant planning and development, the onset of the Covid 19 pandemic created restrictions to in-person data collection and travel that made this project impossible to undertake. As such, in collaboration with my supervisor, I began to plan and develop new research ideas that could form the basis of my doctoral work. Building on a mutual interest in narrative theory (which I developed working and teaching alongside Prof. Delafield-Butt), we began to build a new doctoral research plan that would explore the role of narrative as the temporal architecture of adult-infant interaction during infancy. Chapter 1 represents our review of the literature in this regard.

Initially, we hoped to experimentally explore (using secondary data that formed the basis of past studies) the temporal architecture of neonatal imitation and reframe the debate on this phenomenon so as to view it through a dialogical lens. To this end, a literature search was performed and all experimental studies into neonatal imitation (in both human and non-human primates) reviewed. The living authors of these studies were then contacted to request copies of their recordings. Despite some initially optimistic responses, authors were unable to access and share recordings due to a combination of Covid 19 related restrictions and ethical permissions. As such, we instead authored a theory paper (Chapter 2) which proposed a reframing of the neonatal imitation debate and applied narrative theory to our current understanding of the phenomenon.

Chapter 1 - Narrative as Co-regulation: A Review of Preverbal Embodied Narrative in Infant Development

This chapter includes sections of previously published work (McGowan, T., & Delafield-Butt, J. (2022). Narrative as co-regulation: A review of embodied narrative in infant development. *Infant Behavior and Development,* 68, 101747. https://doi.org/10.1016/j.infbeh.2022.101747) for which I have been responsible. I was the first author of this work and took the lead in the writing of the manuscript under the guidance of my supervisor.

Abstract

We review evidence of non-verbal, embodied narratives in human infancy to better understand their form and function as generators of common experience, regulation, and learning. We examine their development prior to the onset of language, with a view to improve understanding of narrative as regular motifs or schemas of early experience in both solitary and social engagement. Embodied narratives are composed of regular patterns of interest, arousal, affect, and intention that yield a characteristic four-part structure of (i) introduction, (ii) development, (iii) climax, and (iv) resolution. Made with others these form co-created shared acts of meaning, and are parsed in time with discreet beginnings and endings that allow a regular pattern to frame and give predictive understanding for prospective regulation (especially important within social contexts) that safely returns to baseline again. This characteristic pattern, co-created between infant and adult from the beginning of life, allows the infant to contribute to, and learn, the patterns of its culture. We conclude with a view on commonalities and differences of co-created narrative in non-human primates, and discuss implications of disruption to narrative co-creation for developmental psychopathology.

1.0 Narrative

Bruner (1986, 1990) identified two modes of cognition. On the one hand, there exists a logico-scientific mode of cognition able to form concepts with knowledge of their lawful properties. This is a timeless, abstract knowledge that enables a technical intelligence with its rich knowledge of facts that can be approached from multiple imagined perspectives to work with the generation of plans for imagined futures, and solutions to present problems. This static, timeless mode of cognition is the focus of educational systems, is cherished by technological societies, and its capability is what we normally identify as intelligence. However, on the other hand, Bruner identified another form of cognition - narrative. Narrative intelligence places the lawful properties of objects and persons into an animated temporal order, making sense of their events and processes as they relate one to another through lived time. The logico-scientific knowledge of objects and persons become contextualised into this order. Their relations, psychological motives, and the feelings that power their interactions become known and understood within the context of lived, embodied experience. The social value of objects, persons, behaviours, and events become meaningful through one’s lived experience, be that in their creation by oneself, observed in others, or recalled in their re-telling.

Narrative informs and structures logico-scientific knowledge because experience generates stories that descibes the lawful relations between parts, enabling an abstract understanding of those parts. But as lived experience necessarily progresses through time, all logico-scientific intelligence must therefore be expressed back through the animated temporal order of narrative. The two intelligences work hand-in-hand, informing and structuring each other.[1]

---

[1] Think, for example, of the animated inspiration of a good mathematics or physics lecture. It is in the affective telling of the story of technical knowledge that inspires, and is learned.

In this paper, we explore the origins of narrative intelligence in early infant development and review its major theoretical claim as a fundamental structure of animated events, presenting ‘narrative units’ of meaning-making. A growing body of work proposes an early narrative intelligence operative in the cycles of activity of young infants, co-created in pre-verbal dialogues of expressive action with attentive adult care-givers, and structured psychologically and temporally by innate psychobiological rhythms, both autonomic and voluntary. We review these claims together with closely related evidence of early adult-infant interaction to explore how narrative might offer improved explanatory resolution to developmental science of intelligence and meaning-making, and provide insight into other experimental frameworks and analysed interactions, such as in neonatal imitation. In doing so, we identify a self- and other-regulating ability of narrative units, and review evidence in the literature of their presence in early human development, with attention from birth to eighteen months of age.

By reviewing and consolidating the literature, we present an account of narrative as an essential structure that governs and gives shape to infant-adult interactions, and their regulations. Finally we draw this together with a comment on disruption to narrative in autism, the place of narrative in primatology, and the potential for narrative explanations for non-human primate intelligence.

2.0 Narrative as a Structure of Lived, Embodied Experience

Human experience appears continuous and regular as a steady ‘stream of consciousness’ (James, 1890), but it is made up of discreet events that unfold in time, with beginnings, middles and ends. These discreet events are made in active engagement with the world by a physical body expressive of the interests, knowledge, and intentions of the Self-as-Agent (MacMuray, 1957). Generative participation is created by action, itself an energetic and organised activity of the neuromusculature that conveys the feelings of the agent in their vital forms of intent (Stern, 2010). As embodied agents, these participatory experiences become structured by the spatiotemporal properties of the body, and critical to this is its purposeful, future-oriented nature that seeks meaning in interaction with the world (Reed, 1996). Each action, each self-generated movement of the human body, is necessarily organised prospectively from the beginning of life, with an eye to the future (Delafield-Butt et al., 2018; Hofsten, 1993; 2007), and primed to affectively evaluate the consequences of one's self-generated actions, or those of an observed other (Brâten, 1998). The finite and future-oriented nature of human motor control asserts a powerful prospective structure to knowledge as anticipatory and structured by the human body, driven by agent's intentions and interest, and excited by the affects that motivates them.

Prospective, narrative organisation, with its origins in human motor control, delivers a necessary unit that opens, then gives closure to activities and engagements to draw experience to a purposeful conclusion, and to make sense of them. In this way, human agency seeks to parcel what seems on the surface to be continuous experience, into the embodied experiences of purposeful events.

2.1 The Serial Organisation of Action into Projects, an Embodied Foundation of Narrative

> "Not only speech, but all skilled acts seem to involve the same problems of serial ordering, even down to the temporal coordination of muscular contractions in such movement as reaching and grasping. Analysis of the nervous mechanisms underlying order in the more primitive acts may contribute ultimately to the solution even of the physiology of logic."
>
> (Lashley, 1951, p. 121)

Human movement generates personal, and shares social, meaning (Sheets-Johnson, 2011; Stewart, Gapenne, Di Paolo, 2011). Each action we make, with its future-oriented vision, creates consequential responses from both objects and social others in the environment that can be evaluated and learned (Delafield-Butt & Trevarthen, 2020). This generative action-response psychology recognises human experience is created by self-generated agent action, whilst its prospective evaluative nature creates a fundamental psycho-motor structure (Delafield-Butt & Gangopadhyay, 2013; Delafield-Butt, 2014). This stands in contrast to a stimulus-response psychology predicated on experimental paradigms that treat the mind as passive and reactive. Rather, attention to the prospective nature of human movement presents an agent-centred, ecological view of the mind as giving embodied structure to its content, rather than as receiving a stimulus passively.

Made in engagement with the world, each consequence of sensorimtor action can be learned, then chained to a subsequent action. A 'reach-to-touch' develops into a 'reach-to-grasp'. A 'reach-to-grasp' develops into a 'reach-to-grasp-to-drink' (von Hofsten, 1993; 2007; Pezzulo & Castelfranchi, 2009). Over early infancy, simple actions become chained to present complex, serially organised actions that develop stories with greater spatiotemporal reach to achieve aims and intentions further into the future (Delafield-Butt, 2018). Their prospective nature forming the embodied foundation of joint action (Fantasia & Delafield-Butt, 2023).

Where an early 'reach-to-touch' enables acquisition of a goal of about one second into the future, a 'reach-to-grasp-to-place' enables a goal of about three seconds into the future. Several 'reach-to-grasp-to-place' action sequences generates a project many tens of seconds or minutes into the future, from stacking blocks to cooking dinner. In infant development, this is transition from a proximal awareness to one that expands with experience, and the cogntive tools of memory and planning to enable projects with greater and greater distal reach (Pezzulo & Castelfranchi, 2009; Delafield-Butt & Gangopadhyay, 2013; Trevarthen & Delafield-Butt, 2017). These projects are rooted in movement and are goal orientated, generative, and expressive creations of the infant. Their underlying motor development is driven through the interaction and interplay of multiple systems (e.g. muscual, postural, and neural) which both feed into and are directly influenced by the infant propectivity (Thelan & Smith, 1994). Such projects can also be co-created in dyadic partnership with an attentive other (Figure 1.1).

Made in social engagement with an attentive and engaged other, a simple expressive action such as a single gesture of the arm or utterance of the voice can become the first in a sequence of expressions (Delafield-Butt & Trevarthen, 2016). In face-to-face dialogue, these expressions structure the social dynamic and inform the content of social engagement in what Trevarthen identified as ‘primary intersubjectivity’ - an attention purely to the feelings, interests, arousal, and intentions of the other without additional reference to an external object (Trevarthen, 1979; 2012). Patterned over time, these exchanges between infants and adults last from a few seconds, to fifteen or twenty seconds with consistent engagement, ultimately forming narrative units of meaning-making (Trevarthen & Delafield-Butt, 2013).

Evidence we review below shows that states of increasing arousal and interest in the social other are shared in regular patterns of expressive movement of the body and voice in serial order with shared time. Regulation of the autonomic system on both sides becomes tightly coupled to the social engagement system expressive of affect, arousal and interest (Porges and Furman, 2011). These physiological systems of metabolic and autonomic regulation support the psychological interests and intentions of the infant, structuring their capacity for activity and regulated to serve its needs over time. These coupled psycho-biological processes are at the heart of narrative, and learning the expected regulations of a family and culture (Gratier, 2003; Gratier & Apter-Danon, 2009).

Knowledge of the world is built on this ontogenetically primary and psychophysiologically basic narrative regulation to include attention to objects, symbolic displays by persons, and complex events experienced, then held in memory (Vandekerckhove & Panksepp, 2009). Their experience – parcelled in narrative units – becomes an experience of the regulation of autonomic physiology, and intentional body action, but also of its emerging higher conceptual context and knowledge of the relations between actors and objects. These narrative units become learned in the classical sense, through conceptual organisation, symbolism, and memory. In this way, technical learning is tightly coupled to the more basic understanding of physiological regulation – a feature especially important in early years education and care. Both are organised within the same narrative experiences, which gives them their tight relationship. It is through this narrative organisation we can understand language as tightly coupled to feelings, and its movement in music or poetry (Trevarthen, 1995; 2005). Scottish Englightenment philosopher Thomas Reid highlighted the importance of this affective aspect of language, which he called 'natural signs', as more powerful than the technical meaning of 'artificials signs' of the words placed on top - "Artificial signs signify, but they do not express; they speak to the understanding, as algebraical characters may do, but the passions, the affections, and the will, hear them not: these continue dormant and inactive, till we speak to them in the language of nature, to which they are all attention and obedience." (Reid, 1764, pp. 106-108).

2.2 Narratives of Communicative Musicality

Narrative patterns in early infant social engagement were first identified in the reciprocity of affects and interests of shared activities in play between adults and infants, with their characteristic rise and fall of arousal and excitement structured by repeated cycles of reciprocal action in what seminal child psychologist Daniel Stern called 'protonarrative envelopes' (Stern, 1985). It is these early narratives, and patterns of intermodal sharing of affective states (Stern et al., 1985), that generate what Trevarthen (1979) similarly identified as 'protoconversations' in the face-to-face vocal expressive sharing of internal states. These experiences later absorb language as well as representations of intentions, objects, events and emotions, creating the conversations of adulthood imbued with meanings in increasingly complex narratives (Delafield-Butt & Trevarthen, 2013).

Narratives can be seen "inherent in the praxis of social interaction before it achieves linguistic expression" (Bruner, 1990, p. 77). Indeed, embodied, non-verbal narratives can form complete acts of meaning expressed in the affective vitality of movements of the body through gesture and postural shifts, changing facial expressions, and non-verbal vocalisations (Stern, 2010). Even between linguistic adults, expressive movement can convey narrative meaning, with or without language, within a vital musicality or shared timing and quality. Consider the dancer, expressing her story. Words are not necessary to experience a story, nor to re-tell it (even if language if often interwoven and intwined with movement to achieve performative and communicative outcomes [Keevallik, 2015]), with mutliple modalities rooted in movement being core to the temporality of human bodily interaction (Deppermann & Streeck, 2018). Read and Miller (1995) consider narratives to be "universally basic to conversation and meaning making" (p. 143) irrespective of its modality.

Using a sensitive accoustic analysis, Malloch (1999) identified formal features of music within the vocal structure of mother-infant protoconversations. These features included (a) a shared sense of time between partners, and an alignment of expressions within this temporal order, or rhythm, (b) a reciprocity of the quality of expression of the voice or body, where one's expression informs the quality of expression in the other, and importantly for this paper, (c) a musical narrative form created by shared time and recriprocal qualities of expression across a four-part structure of introduction, development, climax, and resolution.

The importance of this discovery is essential. By identifying narrative within the musicality of protoconversation, Malloch and Trevarthen (2009) overturned conventional wisdom that narrative was necessarily predicated on words made with reference to remembered content. Rather, they discovered an ontogenetic primary narrative form that structured sharing experiences of affect, arousal, and interest, conveying and forming the experience of feelings and emotions in mutual, reciprocal interaction patterned regularly in lived, shared time. This fundamental structure of sharing experiences is the earliest expression of a growing narrative intelligence operative in the cycles of activity of young infants, and in the co-created pre-verbal dialogue with attentive adult care givers (Malloch, 1999; Stern, 1985, 2010; Trevarthen, 1999; Trevarthen & Gratier, 2008).

2.3 Narratives Structure Vital Meaning

Narrative structure underpins our earliest interactions, and remains invariant across the human lifespan, although its musicality can be lost to favour the technical symbolic arrangements of words. Nevertheless, narratives remain universally basic to human meaning-making and are ubiquitous across cultures (Bruner, 1990). They are enriched and elaborated to a high degree of skill to amplify and enhance feelings of affective interest and arousal in all the time-based arts of drama, dance, music, and poetry (Stern, 2010). They structure the arts and entertainment we enjoy daily, or co-create with others. Narratives underpin our knowledge, understanding, and experience of the world at a basic, foundational level. They allow us to create meaning with others, to understand how others behave in a given situation, and even what another individual's underlying thoughts and motivations might be for acting in a certain manner. They give us interpretive insights into another human being, and provide a general structure which we can use to frame our understanding of, and interactions with, other individuals (Gallagher, 2011). Bruner (1990) viewed narrative as an organising life principle of human cognition, pervading our actions and communication, forming a temporal structure of our individual goal-orientated movements and our interactions with friends, parents, colleagues and children.

The complexity of the narratives one creates and encounters vary, but their core features remain invariant. They connect the past and present to an unfolding future, linking our historical experiences with immanent possible future realities (Bruner, 1986). Narratives are experienced and learned, then form a template for understanding future similar events. They are first experienced in the vitality dynamics of action, interaction, and play in infancy, building and fading in a predictably patterned dynamic of phases of arousal and intensity (Delafield-Butt & Trevarthen, 2013; Delafield-Butt & Trevarthen, 2015; Stern, 2010). Although the complexity of narratives expands during human development, the narrative form is as present in the interactions of mothers and babies as teachers and students, or the complex technical explanations in a court of law, or a quantum physics laboratory (Bruner, 1990; Delafield-Butt & Adie, 2016; Delafield-Butt & Trevarthen, 2013).

The typical pattern of arousal and intensity develops over four distinct phases: (i) the introduction, (ii) development, (iii) climax, and (iv) resolution (Malloch, 1999; Malloch and Trevarthen, 2009; Trevarthen & Delafield-Butt, 2013; Figure 1.2). Each phase has characteristics which define them and give regular structure to joint meaning-making (Delafield-Butt & Trevarthen, 2015). First, the introduction phase generates shared attention between involved parties to established and begin the narrative. Their interaction then builds in the development phase, typically with rhythmic reciprocity between participants that can resemble a musical exchange composed of expressive acts of the body or voice with shared timing. The quality and form of each expression becomes contingent on that of the other, reflecting and adapting its form in creative contribution like two jazz musicians riffing off each other’s expressive motif (Gratier, 2008; Malloch, 1999). The energy and excitement builds, sometimes only subtly and sometimes to a great extent, until a moment of maximal tension is reached, a climax. This is a moment of peak, held energy, which is typically followed by a release in the resolution phase, where the energy can recede to baseline again. At this point, the participants of the interaction may feel free to break from one another, as they process the memory of their shared experience, and reflect on the meaning that was created.

Whilst this structure may accurately describe a single narrative encounter, many of the narratives adults engage and create will be formed of multiple complex cycles, with many derivations of form based on the same principle of its four-part structure.

3.0 Narrative as Learned Regulations

The presence of narrative within non-verbal expression allows us to consider their importance prior to the onset of language. From the earliest time in postpartum infant life, individual movement orientated projects can become interwoven with those of a caregiver, creating intersubjective events that contain many of the characteristics of fully formed conversations. The importance of narrative in solo projects and communication also makes it integral to the development of our understanding of the world, understanding of others, and understanding of ourselves (Bruner, 1990; Zahavi, 2007). The patterned nature of arousal and energy, inherent to narrative architecture, gives structure to the process of emotional regulation of all forms of movement and meaning-making in intersubjective states (Damasio, 1999; Trevarthen, 2005). Its predicable, regular patterns enable the coordination of sympathy between individuals in the shared time of vocal and motor expressions of affect, interest, and intention (Delafield-Butt & Trevarthen, 2015).

3.1 Co-regulation of Psychobiological States

Co-regulation derived through narrative impacts basic biological and physiological systems. The phases of arousal and intensity in the vitality dynamics of infancy not only form an outward demonstration of a person's inner states, but also allows those internal states of autonomic activity to be influenced by another human being (Porges & Furman, 2011; Stern, 2010). This influence of one individual upon another leads to the explicit co-regulation of physiological states based on autonomic and neurophysiological systems, with the requirement of sensorimotor engagement (Porges, 2011). Indeed, many basic self-regulated biological capacities are supported socially in infancy. Infants prompt support from adults through voluntary communicative signals, which can be very specific and constructive. For example, Vallotton (2009) found infants pro-actively elicit responses from their caregivers during their interactions in nursery, provoking different qualities of care across multiple caregivers to suit their needs. This selective elicitation of social care demonstrates the infant's power of agency, even in social interactions.

Examples of infants actively eliciting support to aid in the regulation of physiological states can be seen through the use of shared body heat with caring adults, or the elicitation of milk from a caregiver (Delafield-Butt, 2018; Tronick, 2005). In these examples, basic needs serve to prompt an engagement with their caregiver, leading the two parties to coordinate their movements in joint projects with shared intention that result in psycho-physical needs being met on both sides – infant and parent. Co-operation allows infants to overcome their individual limitations and at the same time fulfills the needs of the adult to love and nurture, generating satisfaction on both sides of the partnership. The individual acts of behaviour – a vocal cry, facial grimmace, movement of the arms, etc. – are coupled to both immediate physiological needs, warmth and food in these examples, as well as to deeper needs and desires, such as developing the attachments, care, and nurture of companionship (Trevarthen et al., 2006; Lewis and Amini, 2001).

The patterns of behaviour generated and learned in these engagements form remembered patterns of action and interaction, and their consequential effects, that serve as a template with which to plan and organise future action and interaction in a new instance. These learned narrative ‘schemas’ underpin the successful repetition of projects in future encounters.

3.2 The Prospective Nature of Narrative

Narratives formed in basic movements extend the infant’s world through time to structure early self-generated experiences. Purposeful and goal directed acts also follow a basic narrative structure (Delafield-Butt & Trevarthen, 2015) that must "(i) initiate toward that future, (ii) develop in its progression over time and through space with sensory feedback and adaptive anticipatory response charged with memories, and (iii) reach its target before (iv) resolving into a quiet state again" (p. 4). These purposeful and goal directed behaviours are organised in a nested hierarchy of action, where an act, or action chain, is made up of smaller action units, which must be continuously modulated to achieve its desired ‘goal’ or state (Delafield-Butt & Gangopadhyay, 2013) through a continuous interplay of bodily, enviromental and psychological feedback and regulation (Thelan & Smith, 1994).

The nature of a dyadic regulation allows infants to achieve these future orientated goals more readily than if they act independently (Tronick, 2005). For example, infants who attempt to implement an action chain/project, but fail due to physiological limitations, will enter a state of distress. However, in a system of dyadic regulation, adults are able to sense the purpose of the infant’s efforts, and provide additional support that enables them to complete the sequence of actions to achieve that goal. Narrative structuring of solitary actions quickly expands into the social realm of shared projects, where a narrative interplay aids parents and infants in learning the intricacies of one another's patterns and expectations, allowing for the completion of critical endeavours such as feeding. Narratives may also underpin the first dialogues of neonatal imitation and the proto-conversations of early infancy (Nagy, 2006; Nagy & Molnar, 2004).

In these early engagements, bodily movements, vocalisations, and facial expressions operate in a patterned rhythmicity to create a narratively structured state of intersubjectivity. Neonates have a behavioural repertoire of basic abilities which makes them well adapted to become attuned (both temporally and affectively) to another human being, enabling the formation of intersubjective states. These abilities includes a variety of facial expressions of emotions, shaping of the mouth, and lip and tongue movements (Trevarthen, 1979), and are enhanced as infants develop other capacities, such as the capacity to follow another individual's eyes, and perceive their movements as being meaningful and goal directed. As a package, these behaviours allow infants to engage in primary intersubjective interactions, and become tuned to the vocalisations and gestures of adults (Gopnik & Meltzoff, 1997; Stern, 1971).

For primary intersubjectivity to be created, and therefore co-regulation between adults and infants, affective coordination between the movements and expressions of both parties is required. As children grow and develop, so the narratives they are involved in develop in complexity and modality to structure secondary intersubjective engagements involving shared attention to objects, and become imbued with language that allows for the complex story making characteristic of human culture.

For narrative co-regulation of physiological states or meaning making to take place, participants must be aware of and attentive to each other's expressive vocalisations, movements, and ready to respond sympathetically, in time and in tune with the other.

Sustained attention with a sympathetic reciprocity of feeling and expressive action will naturally take on its particular narrative character.

3.3 A Narrative Hierarchy of Embodied Regulations

Self- and other-regulations are achieved within the expectations the narrative template provides, enabling autonomic energy systems to be prospectively organised to maintain delivery of vital resources to the neuromusculature during an activity. This is the first, basic level of learning self-regulation, made in primary intersubjective face-to-face engagements. Secondary to this, essential needs or desires can be met, such as heat retention for thermoregulation, food during sucking for sustenance, or curious care and attention for love and companionship. And once these needs and desires are met, then the infant can develop shared attention to objects where learning technical mastery is accomplished in concert with a social other.

4.0 Development of Narrative Co-regulation in Adult-Infant Interaction

Despite the importance of narrative as a psychological theory of human meaning-making, a large section of the current literature on co-regulation and adult-infant interaction does not address this underlying narrative aspect. However, recent decades have seen a greater appreciation of the agency and generative role of the infant in interacting with the world and other individuals within it. This is especially true of studies investigating adult-infant synchrony and entrainment, as well as physiological and psychological co-regulation. Understanding the functional and temporal architecture that underlies these core communicative processes can further aid understanding of wider human development, and offer new perspectives on ongoing debates. In this section we explore studies examining adult-infant interaction and co-regulation across early human life (the neonatal period, 4 weeks to 8 months, and 9 months to 15 months) to see how co-regulation is important and present in these early interactions, and how it develops the nature of narrative interactions.

4.1 The Neonatal Period

New-born infants display sensitivity and awareness of purposeful movements and vocalisations of adults in relation to their own person (Condon & Sander, 1974; van der Meer & van der Weel, 2011; Nagy, 2011). In the hours after birth human infants have been shown to demonstrate an ability to take part in imitative exchanges with adults, conversations rooted in movement (Kugiumutzakis & Trevarthen, 2015, Meltzoff & Moore, 1983, 1989). These imitative exchanges are considered to be a foundation for social communication, and as an example of an infant-generated communicative act seeking responsive company (Meltzoff, 2007; Nagy & Molnar, 2004). They can also be considered a first step in the gradual development through infancy to a more abstract and reflective social intelligence that allows for a theory of mind (Meltzoff, 2007). Indeed, these imitative exchanges represent some of the first examples of a co-operative, co-creation of interest and affect between infants and adults, an initial expression of primary intersubjectivity (Trevarthen, 2011). The infant's ability to engage in these social exchanges is built of multiple skills that allow social engagement and joint meaning-making, including the ability to identify their mother's voice and face, and a preference for humanoid faces (Burnham, 1993; Decasper & Fifer, 1980; Valenza et al., 1996; van der Meer & van der Weel, 2011). These contribute to what some consider an innate disposition for intersubjective engagement (Nagy & Molnar, 2004; Stern, 2000; Trevarthen et al., 2006).

Further evidence for an innate readiness or predisposition within neonates towards social interaction has been found through other means. For example, Farroni et al. (2002) demonstrated larger event related potential (ERP) responses when infants (even in the neonatal period) were presented with images of adults with direct as opposed to averted gaze, and Murray and colleagues (2016) desscribed a social architecture that allowed both adults and infants to respond in a regular and expected manner when interacting in a dyadic engagement. In this study, infants were recorded (in the first nine weeks of life) in interactions with their mothers, and social facial behaviours (e.g. smiling), biological events (e.g. sneezing), non-social mouth movement and expressions of negative affect were coded and mothers monitored for contingent responses. The prominence of contingent maternal responses relative to non-contingent behaviours were examined, and the form of maternal response analysed with respect to their infants' behaviours and ages. The specificity of associations between behaviours in infants and adults were estimated. Their findings demonstrated the infant social expressiveness was structured on specific forms of parental response: a functional, bi-directional social architecture where infant expressivity was contingency of maternal responsiveness, and vice versa.

These early exchanges support the continued development of cooperative engagements, which grow to become more complex and prospective in their nature (Nagy, 2011; Delafield-Butt, 2018). Each act (imitative or otherwise) can be considered part of a larger narrative, which creates meaning over multi-second encounters in a "mutually sustained expectation of a rhythmic project of communication, engaging two persons in a nonverbal 'narrative'" (Kugiumutzakis & Trevarthen, 2015, p. 487).

It is only in the last sixty years, following Meltzoff and Moore's (1977) seminal work, that neonatal imitation has become a focus of academic attention. From this paper followed a host of studies that supported and clarified the nature of neonatal imitation (e.g., Field et al., 1982; Kugiumutzakis, 1985; Meltzoff & Moore, 1983) as well as others that attempted to replicate, but failed to find supporting evidence (e.g., Fontaine, 1984; Hayes & Watson, 1981; Koepke et al., 1983). Nagy and Molnar (2004), however, went further than to simply replicate the test for neonatal imitation in humans. Their study measured an infant's ability not only to imitate a social other, but to initiate social interaction. In this study, 45 neonatal infants, aged 2-54 hours confirmed imitation of tongue protrusion, but they also found their infants initiated voluntary exchanges with an adult by presenting a spontaneous tongue protrusion of their own. The researchers utilised behavioural and psychophysiological measures to differentiate initiation from imitation. Heart rate determined levels of arousal (increased heart rate was considered reflective of increased levels of arousal) and of orientation, learning and expectance (decreased heart rate was taken to be indicative of these during imitation).

Following several cycles of imitation, it was found that the neonates not only imitated the modelled action, but also initiated an exchange and waited for the adult to respond, demonstrating a motivational desire to take part in the dialogue. Such findings help to illustrate the social motivation that appears to underlie neonatal imitation, as the infants actively sought to continue exchanges with their adult partners. Indeed, during the study it is reported that there were several examples of prolonged exchanges between neonate and experimenter, which were initiated by the participant and developed into what the researchers describe as reciprocal ‘conversations’ (Nagy, 2006). Nagy has designated these exchanges as experimental examples of the first dialogue (Nagy, 2006, 2011; Nagy & Molnar, 2004). It is these findings, together with the wealth of data on imitation and the social orientation of neonates, that allow us to consider neonatal imitation one of the earliest expressions of primary intersubjectivity in an experimental setting. These early intersubjective exchanges can go on to develop into rich protoconversations involving the serial organisation of expressive actions of the hands, voice, face and body movement later in infancy that structure the pre-verbal narratives of communicative musicality first described by Malloch and Trevarthen (Malloch 1999; Trevarthen, 1999; Malloch & Trevarthen, 2009).

The structure of adult-neonate interactions includes findings that extend beyond neonatal imitation. For example, Condon and Sander (1974) showed infant sensitivity to the motor impulses of adult speech through the use of microkinesics. This method involved a frame-by-frame analysis of the type of movement performed by infants mapped to the acoustic features of the adult's voice. Infants involved in the study were found to move in synchrony to the adult's syllables, words, phrases, and broader rhythm of speech. That is, the changes in the movement of the body parts of the neonates were coordinated with shared sub-second timing to the sound patterns of adult speech. Condon and Sander described this as an interactive dance that preludes the development of speech. Furthermore, they suggested that infants have an ability to discriminate human speech from other sounds and noises, and arrange their motor responses accordingly.

The neonatal period also plays an important role in establishing patterns of co-regulation that have significant impact through development. For example, skin-to-skin contact in premature neonates has been shown to have long term effects on physiologic organisation and behaviour control, whilst physiologic organisation in the neonatal period has also been shown to impact an infant's ability to successfully engage in rhythmic adult-infant dialogues as they grow (Feldman, 2006; Feldman et al., 2014). These findings are supported by evidence suggesting that maternal-neonate separation dramatically increases autonomic activity and greatly decreases quiet sleep duration - quiet sleep being closely related with cortical organisation and neuronal reorganisation processes (Eiselt et al., 2001; Milde et al., 2011; Morgan et al., 2011). Skin-to-skin contact is also important for neonatal temperature regulation, with it having a significant impact on both core and peripheral temperature, not only during the state of skin-to-skin contact but also during subsequence separation (Bystrova et al., 2003; Fransson et al., 2005).

4.2 From 4 weeks to 8 months

Moving beyond the neonatal period, considerable attention has been given to the structure and composition of the interactions and co-regulation that exists between adults and infants. Early studies by the likes of Stern and Beebe highlighted the importance of movement, as well as vocalisation, in understanding the nature of adult-infant interaction. For example, Stern and colleagues (1977) examined the temporal structure of adult-infant interaction and highlighted the importance of a multimodal analytical approach and the need to consider movement as independent units of communication. Beebe and colleagues (1985) also investigated both vocal and kinaesthetic behaviours in the temporal organisation of adult-infant dyadic interactions. Using 15 dyads of mothers and infants (infants aged 4 months), they found that the nature of the temporal patterning varied within each of the behavioural forms investigated. For example, when one member of the dyad increased their levels of vocalisation, this trend was mirrored by their partner. Conversely, with regard to movement, mothers and infants altered their activity level as a function of their partner's level of movement in such a way that as one member of the dyad increased movement activity the other decreased theirs. It is within this ebb and flow of activity that the narrative temporal organisation of an engagement can be found, and it is within this narrative framework that mothers and infants regluate one another's arousal, interest, affect and attention.

However, understanding the temporal nature and organisation of these interactions as synchronous or alternating has been the subject of debate. Cordon and Sander (1974) found evidence of a synchronous patterning to the interaction, whereas Beebe and colleagues (1979) found evidence for both simultaneous and alternating patterns in the temporal organisation of vocal and movement behaviour in mother-infant interaction. Meanwhile, Fogel (1977) showed synchronous and alternating temporal patterns could be present at the same time within an interaction. We now understand the synchronous and alternating organisation of expressive movement yields ‘shared time’ integral to the matching of psychomotor states, and its co-regulatory processes (Osborne, 2009; Malloch & Trevarthen, 2018). In any dyadic interaction between two human beings, expressive movement is integral to dyadic regulation and meaning making (Fogel, 1993). When such interactions are successful and with shared regulation, they become greater than the sum of their parts and there emerges a more complex and coherent understanding of the world (Tronick, 2005).

Underpinning all these engagements is a mutual sensitivity in both adults and infants to the intricacies and details of their social interactions. For example, Grossmann and colleagues (2008) found that infant brains are highly sensitive to social cues. Across two experiments the researchers investigated the effect of mutual and averted gaze followed by a raised eyebrow, by an animated photorealistic adult face, upon an observing infant's haemodynamic responses and oscillatory brain activity. They found that observing facial communication signals activated regions of infant’s temporal and prefrontal cortex that are associated with the same processes in adults. Urakawa and colleagues (2015) also measured hemodynamic responses, as well as eye movements, whilst infants aged 7 months engaged with a partner during a game of peek-a-boo. Their results found that direct gaze during social play prompted infant attention to be drawn to their adult partner's eyes, and there was a specific activation of the dorsomedial prefrontal cortex during both direct and averted gaze. The impact of gaze on neural activity was also investigated by Leong and colleagues (2017) during the singing of nursery rhymes. Across two experiments adults recited nursery rhymes, either displayed on a screen (experiment one) or during a live interaction (experiment two), and alternated between direct and indirect gaze (with a control condition in experiment one only in which adults turned their heads as if averting gaze, but remained looking at the infant). Adult and infant neural activity was measured through electroencephalography and the researchers found that adults had a significant causal influence on infant neural activity in both experiments. This influence was stronger during the directed gaze conditions (regardless of head position) than in the indirect gaze condition. In the second experiment of the study, adults were also influenced more by their infant partners through direct gaze than indirect gaze. Additionally, infants were found to vocalise more during the second experiment, and infants who vocalised more had a more pronounced neural influence on their adult partners. These results again highlight the impact of gaze in bidirectional neural connectivity, but also serve as evidence for the role of ostensive signals within intersubjective interactions in creating a mutual temporal alignment, a basis of shared timing.

Such sensitivity to ostensive signal timing is likely underpinnned by shared psychobiological rhythms between adults and infants. Trevarthen identified a common temporal structure in adult and infant limb movements made in communicative gesture, as well as in inter-saccade intervals of eye gaze, despite their differences in size (Trevarthen, 1986). Such common timing principles of the human body, including autonomic and central nervous regulation, enable coupling between individuals along multiple levels of arousal, interest, affect and attention regulation important for human developmental health (Trevarthen et al., 2006).

In addition to sensitivity to social cues, infants are also able to co-ordinate interpersonal timing within a dyadic interaction. That is, they can alter the timing of their own behaviour in line with the timing of another individual with whom they are in engagement. Crown and colleagues (2002) recorded the interactions of infants (aged 6 weeks) with their mother and a stranger, and then coded for mother/stranger vocal behaviour and infant gaze behaviour. A time series regression was then conducted to ascertain the relationship between infant gaze and adult vocal behaviour. The results showed that infants, even as young as 6 weeks, demonstrated a sensitivity to the temporal structure of dyadic interactions – infants were able to extract the temporal properties of vocalisations from both their mothers and strangers (characterised by the timings of vocalisations and pauses in relation to infant gaze), and express it through gaze behaviour. Additionally, adult vocal behaviour during these interactions was impacted by the temporal organisation of infant gaze behaviour. These results highlight the dialogic nature of the temporal structure of the dyadic interaction.

Key to the maintenance of this temporal structure (and the resulting dyadic co-regulation) is not only infant attunement to the actions and vocalisations of an adult caregiver, but also an adults ability to stimulate infant involvement in interactions. Stern (1974) focused on adult behaviour during dyadic interactions between 24 mother-infant dyads. He found that the actions of adults were organised into discernible units and believed the adult's aim was to maintain a level of arousal in order to promote positive social behaviours. Beebe and Gerstman (1980) also show the importance of considering adult behaviour in adult-infant dialogues in a case study which demonstrates that maternal multimodal acts in the dyadic interaction are communicative in nature. In this study, the relationship between maternal behaviours and infant engagement was measured and a significant relationship found with packages of maternal behaviour (consisting of kinesic hand rhythms and facial display variations) varying systematically with levels of engagement from the participating infant. From these results the researchers inferred the communicative quality of maternal acts and behaviour, which underlies their role in the co-regulatory process.

A final element to consider in relation to co-regulation in this age group concerns the link between pysiological and affective co-regulation. Feldman and colleagues (2011) measured the cardiac responses of mothers and infants (aged 3 months) during face to face interactions whilst also performing a micro-analysis of vocal synchrony, gaze and affect. In their analysis the researchers found the heart rhythms of both infants and adults were closely coordinated, but also that there was an increase in concordance between the biological rhythms of participants when affective and vocal synchrony existed. Furthermore, Ham and Tronick (2009) found a close link between the physiological and psychological regulations that take place within the adult-infant dyad. Ham and Tronick explored emotional regulation through measuring respiratory sinus arrhythmia and skin conductance concordance whilst participants took part in the still face paradigm. They found that whilst infants were presented with still face, skin conductance concordance in the dyad correlated with infant negative engagement, but following the reestablishment of the interaction it correlated with behavioural synchrony. Maternal respiratory sinus arrhythmia was found to be negatively correlated with infant negative engagement during normal interaction, but was significantly postively correlated to it during the rengagement episode (possibly due to mothers calming themselves to aid soothing their infants). These results not only demonstrate the close link between physiological and psychological regulation within the dyad, but also the impact behaviour can have not only on a dyadic partner but also on oneself.

As infancy proceeds, an important transition from primary to secondary intersubjectivity develops as infants engage with shared interest to objects and events in their environment (Trevarthen & Hubley, 1978; Hubley & Trevarthen, 1979). Shared narratives transition from exclusive reference to the bodily experiences, affects, interests, and arousal, to include the shared attention to objects and persons in space. An interesting point along this transition is found in tickling play, where attention to a specific, discreet activity of stimulation is made in addition to the primary intersubjective sharing of psychological states, and expectations. In a study by Ishijima and Negayama (2017), tickling was identified as a precursor to secondary intersubjectivity, with shared attention to body parts, and its peculiar tickling sensations made in anticipatory play. What was striking in their analysis was the narrative form in which tickling play took part, with a clear four-part stucture of introduction, development over repetitive cycles, before the final climax of contact, then resolution to quiet again. Similarly, Rossmanith and colleagues (2014) analysed book reading as another example of adult-infant triadic shared attention that developed before the onset of full-blown secondary intersubjectivity at 9 months. Again, these adult-infant interactions to a common project were structured "in a highly co-ordinated way, with caregivers carving out interaction units and shaping actions" into what they termed 'action arcs', or non-verbal narrative arcs of shared interest, arousal, and intention. Finally, Reddy and colleagues (2013) studied the anticipatory nature of infant cooperation in a simple task of being picked up. They found infants as young as 3 months prepared their body posture for the forces involved in being picked up, and this anticipatory awareness of their involvement in a shared project developed with specificity over the next months. In a closely related set of studies comparing Japanese and Scottish narrative projects (see Appendix A for definition of terms) of infant pick-up and spoon feeding, at 6 and 9 months, Japanese mothers were found to be more sensitive to the subtle cues of interest in their infants, and responded to follow their infants' initiative, whereas Scottish mothers were more driven and led their infants to follow their adult initiative (Negayama et al., 2015; 2021). But in both countries, the infants learned the narrative sequence of the project, and actively anticipated its steps to co-create joint action in shared time. Altogether, these studies demonstrate a growing awareness of infant involvement in the co-creation of joint projects that extend through time, and an anticipation of how those experiences will be patterned across time. By sharing attention jointly, even young infants demonstated a prospective awareness of the story, and organised their actions in the present moment, contingent on an expected future. Tickling plays on this anticipatory awareness, with exceptionally elatory effects, a feature that can only be created by a trusted, intimate social partner.[2]

---

[2] Interestingly, tickling is not permitted, or does not work when performed by a stranger. It's intimate nature and counterplay between care and aggression is a potential risk, and so it's co-created experience is exclusive to trusted others (Negayama, 2011; Negayama, 2022, p.160).

4.3 From 9 months to 15 months

As infants grow and develop the nature of their interactions becomes more complex, and the narratives they weave and create with others become more intricate. One example is through the onset of secondary intersubjectivity in which infants and adults share attention to elements of the external world. Striano and colleagues (2006) investigated the underlying neural activity in joint attention using a live action paradigm which involved an adult gazing at an infant's (aged 9 months) face and then at an object displayed on a screen, or alternatively only gazing at the object presented on the screen. Event related potentials were monitored in infants and found to be greater in amplitude in the joint attention condition. Neural activity during joint attention was also investigated by Hoehl and colleagues (2014), who examined infant's oscillatory neural activity during a joint attention-based interaction with an adult. The researchers monitored activation of the alpha frequency range and found desynchronisation during a joint attention condition but not in a non-social condition, suggesting that the activation of specific neural pathways is triggered by adult eye contact prior to visual object investigation. Rayson and colleagues (2019) investigated infant (aged 6.5 and 9.5 months) neural responses (using electroencephalography) in reaction to static images of adults either following infant gaze to a specific object (congruent actor), or looking at an alternative object (incongruent actor). Both infant age groups demonstrated greater neural reactivity to the congruent actor, although this was stronger in infants aged 9.5 months.

Neural activity has also been explored in adult-infant communication across everyday, naturalistic interactions. Piazza and colleagues (2020) utilised functional near-infrared spectroscopy to monitor neural activity (through changes in bood oxygenation) in both infants (aged 9 - 15 months) and adults during activities such as playing, singing and reading. The researchers found that adult and infant brains are linked (to differing degrees) during interactions with social cues such as joint attention, smiling, gaze and speech prosody. These findings give an indication into the role of social cues in creating and maintaining intersubjective states, and how the infant brain processes input from adults during interactions, and how adults interpret infant feedback.

Importantly, as discussed, the co-regulation that takes place between adults and infants cannot be limited to social or physiological experiences, but also inherently includes the co-regulation of affective states. Waters and colleagues (2014) examined how affective states are shared and regulated between mothers and infants (aged 12-14 months). In the study, mothers completed a modified Trier Social Stress Test (Kirschbaum, Pirke & Hellhammer, 1993) in which they were asked to give a 5-minute speech about their strengths and weaknesses to two evaluators before a 5 minute question and answer session. Within this task mothers were randomly assigned to one of three conditions. In the first condition mothers received social evaluation with positive feedback from their evaluators, in the second they received social evaluation and negative feedback, and in the third control condition they received no feedback and the evaluators were not present. Cardiovascular measurements were taken of mothers and infants, the positive and negative affect schedule was completed by mothers, and post-stress interviews were conducted during which infant behaviour was coded. The researchers found that infants embodied the mothers stress when they were reunited following completion of the task, and that infants showed greater avoidance of strangers if their mother had experienced social evaluation during their task than infants whose mothers had completed the control condition. Additionally, the dyads showed greater physiological covariation if mothers had experienced social evaluation with negative feedback, as opposed to positive feedback or the control condition. These results highlight how dyadic co-regulation can also create states of negative regulation; i.e. when mothers experience stressful situations, their affective state is transferable to infants, and members of the dyad are capable of destabilising one another.

The loops created by such feedback were explored by Wass and colleagues (2019). For example, if during an interaction a mother is influencing the affective state of her infant, how does this reciprocally influence the mother? To investigate this, Wass and colleagues recorded vocalisations, heart rate, heart rate variability and movement in naturalistic interactions in home settings. They found that in scenarios where the mother's initial arousal level was low but the child's level of arousal was increasing, mothers would correspondingly increase their level of arousal. However, when a high arousal level already existed in mothers, and infant arousal increased, mothers responded by lowering their own arousal level. These results highlight the active role of parents in regulating the arousal state of infants, and how the context of the interaction plays an important role in dictating parental responses. Ultimately, parents connected and disconnected their level of arousal from infants in order to maintain allostasis within the adult-infant dyad (Wass et al., 2019).

Moving beyond the physiological measures utilised by Wass and colleagues, Santamaria and colleagues (2020) looked at the impact of positive and negative affective states in adults upon neural entrainment. In social interactions with their infants (mean age 10.3 months), mothers were asked to model either positive or negative emotions towards pairs of objects. During the interaction neural activity of both mothers and infants was measured using electroencephalography. The researchers found that during trials in which mothers modelled positive emotion towards an object, both members of the dyad demonstrated stronger integration of neural processes, and mother to infant directional influences were stronger. Such results underline the importance of affective states, not only in social learning, but more broadly in the overall entrainment between adults and infants in dyadic interactions. The researchers suggest that the parent-infant inter-brain network, is in fact, modulated by the affect and tone of a social engagement.

5.0 Brainstem as the Primary Regulator of Experience

Underpinning co-regulation of autonomic and voluntary behaviour are core brainstem regions responsible for integration of perceptual information and coordination of action (Panksepp, 2005; Porges & Furman, 2011). The brainstem is the primary regulator of the autonomic nervous system with responsibility over sleep, breathing, heart rate, alertness and feeding (Ngeles Fernández-Gil et al., 2010). It also has important functionality with regard to sensory perception, including the uptake of proprioceptive information, touch, hearing and taste all of which are important in the self-regulation of movement for expressive communication (c.f. Dadalko & Travers, 2018), and is critical for action responses in rhythmic social engagement (Trevarthen, Delafield-Butt, & Schögler, 2011; Delafield-Butt & Trevarthen, 2017). Further, the brainstem plays a role in some higher cognitive functions including the regulation of social attention, and the modulation of emotions (Geva et al., 2017; Venkatraman et al., 2017).

The brainstem's role in regulating behaviour and emotion, whilst also being the primary regulator of the autonomic nervous system, has allowed it to evolve into a powerful social engagement system that impacts intonation in our voices, facial expressions, and hand movements (Porges & Furman, 2011). This link between expressive communication and autonomic physiological functions provides human beings the ability to socially share aspects of our wellbeing, vitality, and needs. Such an ability is core to dyadic regulation. Furthermore, it also allows the brainstem to generate the basic impulse for movement, which forms the core of an individual’s sense of “self” (Panksepp, 1998; Panksepp & Northoff, 2009; Delafield-Butt & Trevarthen, 2013). The actions that form these basic movements require a knowledge of the external world, rooted in the brainstem's role in sensory perception and visceroception of internal organs. The brainstem serves an integrative and generative regulatory function of the agent (Sherrington, 1906). It is through this function that the brainstem creates the foundation of an ontogenetically primary conscious experience and sense of self (Ciaunica, Safron, & Delafield-Butt, 2022).

Dyadic regulation plays a key role in the creation of this first conscious sense of self . The integration of sensory information is influenced through dyadic regulation. It is not a passive process, but is actively generated by the interests and intentions of the infant that becomes regulated through and within joint interaction. Perception of the environment is structured by the self-in-relation to its objects and other persons, which becomes structured in time through social interaction and active, embodied engagement. This activity – self-generated and co-regulated – creates in human experience what Jaak Panksepp (2005) refers to as the Primary 'SELF'.

This brainstem-based consciousness is an adaptable form of mental agency, generating purpose and meaning, as well as anticipating and perceiving the external social world, and its objects (Delafield-Butt & Trevarthen, 2015). It plays a key role in the co-regulations that take place between adults and infants as they actively seek to create and share meaning – meaning that we are coming to recognise as narratively organised action and interaction. Kugiumutzakis and Trevarthen (2015) contextualise a newborn infant's ability to imitate is an example of its desire for narratively structured dialogical interaction, and demonstrates the infant's ability to co-operate in temporal sequencing.

Co-ordination and co-regulation in the temporal organisation and sequencing of dyadic interactions is vital for the successful creation of meaning, as well as the individual achievement of goal-directed actions. Each action unit, let alone action chain, requires a host of muscles within the body to work in near perfect collaborative coordination to initiate and manage the forces involved in the achievement of a goal-directed movement (Bernstein, 1967; von Hofsten, 2007). The ability to co-ordinate and regulate these movements in tandem and synchrony with another is key to wider co-operation in social groups through sympathetic harmonization, and forms a key element of an innate, embodied narrative (see Appendix A for definition of terms) organisation of shared psychomotor time in communicative musicality (Malloch, 1999; Trevarthen, Delafield-Butt & Schögler, 2011; Malloch & Trevarthen, 2009, 2018).

The importance of the brainstem discussed here is not intended to undermine the role of the cortex, but rather highlight an often-undervalued aspect of neuroanatomy (Winn, 2012). The complex relationship between the brainstem its wider neuroanatomy underpins the wider role played by this core region (both in terms of location and function) in core aspects of human experience (Buzsáki, 2006; Damásio, 2010; Panksepp & Biven, 2012).

6.0 Evolution of Narratives Across the Primate Order

Shared social projects create narratives between individuals. They make stories that can be held in memory for later recollection and use in planning and decision-making, enabling imagination to extrapolate and chain potential sequences of events into the future. This process informs choice in the present moment, enabling predictions of possible futures, possible outcomes of events based on choices in the present moment. Knowledge of narrative structure allow predictive time travel, and enacting that knowledge yields new instances of its realisation.

Scientific knowledge of the evolution and development of human narrative intelligence has much to gain from comparative study of narrative within non-human primates. As a social order, all primates share projects together, and coordinated interaction between infants and parents remains a necessity for survival, development, and learning, with human child-rearing extending this process significantly and allowing for uniquely human evolution of the development of intelligence (Hrdy, 2009). Yet, despite significant attention across primate species on the origins of language, social cognition, and shared understanding, the concept of narrative – a cornerstone of human meaning-making – has not yet influenced the field of comparative psychology, nor primatology, except in a few, brief cases. At the same time, there is considerable primate research on the nature of shared projects in primate evolution and development. Much of this work has been motivated by a desire to shed light on the evolution of human social development, particularly around the productive and technical capabilities of shared intentionality and sustained coordinated action to solve tasks, and inform language. But fundamental to all of this is how these projects – these stories – are affectively shared in rhythms of the body in primary intersubjectivity, and then later recalled (Tomasello, 2008; Tomasello & Carpenter, 2007; Tomasello, Carpenter, Call, Behne, & Moll, 2005; Voinov, Call, Knoblich, Oshkina, & Allritz, 2020).

6.1 Imitation in Non-Human Primates

The imitative capacities of neonatal non-human primates, the manner in which rhythms can be shared and understood, and the underlying cognitive capacities give some indication of the capacities of non-human primate infants and juveniles to share and coordinate their intentions and actions in social projects (Bard, 2007; Ferrari et al., 2009; Hattori & Tomonaga, 2019; Hattori, Tomonaga, & Matsuzawa, 2013; Martin, Biro, & Matsuzawa, 2017; Matsuzawa, 2007; Myowa-Yamakoshi et al., 2004). Additionally, new observation shows a precocious capacity for imaginative, pretend play in chimpanzee juveniles (Matsuzawa, 2020) that suggests creative capacities for exploring and ‘mucking about’ are intact in juvenile chimps (c.f. Reddy, 2008). But more work is required to better understand the composition, character, and species limitations and capacities of embodied narrative meaning-making. As a social order, all primate species share projects together.

6.2 Narrative Intelligence Hypotheisis

The Narrative Intelligence Hypothesis (NIH) provides a potential foundation upon which future comparative work can be conducted. It places the structure of animal projects and interactions within a narrative framework (Dautenhahn, 2001) and outlines an interwoven "relationship between the evolution of narrative and the evolution of social complexity in primate societies" (p. 249), leading to the formation of linguistic communication whilst also acting as a binding agent in the fabric of human social relations. NIH proposes an evolutionary trajectory of intelligence that tracks the growing complexity of narrative transactions.

NIH is in agreement with the narrative projects in human infancy first identified by Stern, Trevarthen, and others. In non-human primates, it is seen in projects of physical contact, such as one-to-one grooming, but it is also evident in non-verbal stories enacted between individuals that begin to allow for communication about social matters. Finally, NIH identifies language-imbued interactions and stories common to human society and culture.

Beyond the impact of narratives on social interactions, NIH proposes they aid in the extension of the temporal horizon, allowing individuals to view time and reality in abstract forms that might not otherwise be possible, and potentially therefore aid in the creation of an autobiographical self (Dautenhahn, 2001; Nelson, 1993). Within these abstract notions of time and self, solo projects and sequences of actions can be built (although within the framework of NIH these represent a form of "pre-narrative" that lacks a social or transactional nature, and so fails to meet the requirement for fully fledged narratives).

An important feature of this view of narrative is they are not the exclusive creation of language, but are also found in the transactions of non-verbal and pre-verbal interactions. This is significant due to the importance of non-verbal and physical interactions in human intersubjectivity, as well as within non-human primate social interactions. For example, Bard and colleagues (2019) reported in young chimpanzees a flexibility and diversity in their form and use of touch when used communicatively across various types of affiliative contexts, as well as simple dominance/submission scenarios.

NIH proposes that narratives evolved because their structure is particularly suited to communicate about the social world, or for conveying meaningful and socially relevant information including content regarding emotions and intentions (Dautenhahn, 2002). They allow individuals, and the groups they belong to, the ability to deal with increasing complexity in the social field (with increasing complexity and group size also argued as underlying human neocortex size, which helps manage the additional processing associated with human social grouping (Dunbar, 1992, 1993, 1998)). This is not an exclusively human ability, and in animals (such as chimpanzees) can be found in interactions where the content is about the social field. Dautenhahn (2002) predicts "a correlation between the complexity of the narrative format and an increasing complexity of the primate social field" (p. 116), which appears to align with Bruner’s notion of the role of narrative in human society (Bruner, 1986; 1990). Important to note, the ‘social field’ is not limited to group size but could also includes other elements of social complexity, whilst the idea of format relates to the exact structure of the transaction. The nature of the narrative format in non-human primates requires more detailed investigation, but NIH proposes it follows the four stages described by Bruner and Feldman (1993): (i) a canonical steady state followed by (ii) a precipitating event which then reaches (iii) a restoration, and is finally followed by (iv) a coda, marking the end of the interaction.

6.3 Evidence of Narrative as a Communicative Structure in Chimpanzees

The pant hoot is one particular example of narrative structure important in the communicative patterns of chimpanzees. The pant hoot allows direct communication, and offers some flexibility for adapted social co-creation, a phylogenetically early form of vocal narrative co-creation. Pant hooting is common in adult males and serves multiple purposes, most well-known as a long-distance call for the recruitment and maintenance of social relationships (Mitani & Nishida, 1993). Male chimpanzees also pant hoot with nearby individuals in 'choruses' in which two or more chimpanzees will pant hoot together, often in an overlapping fashion. The acoustic features of a typical pant hoot has a temporal pattern that follows the same four-part structure identified in human mother-infant narrative interactions, but these are described under slightly different labels within primatology (Figure 1.3). Chimpanzees manipulate the temporal fabric of the pant hoot to allow for social interaction, for example by lengthening the 'build-up' phase to facilitate other chimpanzees in joining a chorus (Fedurek et al., 2013). Recent acoustic analyses further define manipulation of its acoustic qualities to communicate, for example social status and context (Fedurek et al., 2016).

The social function of pant hooting is further exemplified by other underlying features of the behaviour. For example, chorusing is more common among preferred conspecifics than neutral ones, it is a good signal of short-term social bonds, and males who associate more often with one another are more likely to produce similar pant hoot calls (Fedurek et al., 2013; Mitani & Gros-Louis, 1998; Mitani & Brandt, 1994). Although it is not clear if the matching of pant hooting between individuals is linked with shared affective states, in pygmy marmosets it has been shown that individuals modified their calls in terms maximum frequency and frequency range in line with fellow group members, and these increases are likely linked with increased arousal (Elowson & Snowdon, 1994; Mitani & Gros-Louis, 1998; Scherer, 1986). Further investigation is required into the co-regulatory function of alterations in pant hoot calls, but that pant hoot chorusing is narratively structured and serves a social bonding purpose highlights the importance of narrative across species, and gives evidence of its early phylogenetic origins (Mitani & Brandt, 1994).

7.0 New and Developing Approaches

Advancements in computational methodologies, machine learning and computer vision movement tracking offer new avenues to explore adult-infant interactions, and infant development more generally. Such approaches can allow for the integration of the multi-modal nature of human interaction, as well as offer new persepectives on older experimental paradigms (Hammal et al., 2015; Vinciarelli et al., 2009). For example, Hammal and colleagues (2015) automatically tracked and modelled adult and infant head movements during the still face paradigm. The researchers found that mother and infant head movements, and their coordination, varied systematically across the three phases of the paradigm. Head movement has also been automatically tracked with motion capture in infants in dyadic face-to-face engagement (Væver et al., 2015) and in children using computer vision whilst they watched social and non-social stimuli (Martin et al., 2018). Facial expressions (smile strength, eye constriction and mouth opening) have also been modelled using computer vision software in a pilot study involving two adult-infant dyads (Messenger et al., 2009).

Moving beyond the head and face, computational motion tracking of body parts has also been implemented in the study of adult-infant dyadic interaction. For example, to differentiate between situations of emotional neglect and typical development, Leclère and colleagues (2016) utilised 2D and 3D video sensors to monitor free play sessions between mothers and infants. The researchers also used machine learning techniques to classify dyads on the basis of the captured motion data, successfully classifying 100% of dyads as members of the control group or as dyads with mothers showing neglect.

Machine learning approaches have also been applied to investigate the predictability of mother and infant behaviour during dyadic engagements (Messinger et al., 2010), but despite their growing and varied use in exploring adult-infant interaction, the application of this approach to examine narrative architecture is yet to be implemented. Its ability to integrate multi-modal sources, produce highly accurate markerless motion tracking, and classify data of a temporal nature, make machine learning a significant approach for future research into the role of narrative in dyadic interactions (Leclere et al., 2016; Pou et al., 2019; Vinciarelli et al., 2009).

8.0 Conclusions

Despite the long-held belief that narratives rely on a form of abstract, reflective intelligence and language (Hutto, 2007), the pre-linguistic exchanges of infant-adult proto-conversations have been demonstrated to utilise both vocalisations and movement to co-create non-verbal narratives that share affective and social meaning (Delafield-Butt & Trevarthen, 2015; Malloch, 1999; Malloch & Trevarthen, 2009). These narratives form a bedrock of adult-infant co-regulation in social engagement that serves to support meaning-making, as well as to co-regulate affective and physiological states. These pre-verbal narratives, and the sensorimotor systems that underlie them, are integrated and organised by a brainstem-mediated primary conscious awareness. It is within this subcortical core self that the origins of narrative co-regulation are to be found, with their psychomotor nature structured by the temporal frame of reference and serial organisation required of coordinated, prospective animal action. Narratively structured co-regulation is rooted in bodily movement: the generative, goal-directed foundation of human consciousness (Merker, 2005).

The infant's role in the creation of narratives and corresponding meaning-making creates a foundation for learning patterns and their temporal organisation, ultimately allowing the learning and development of complex embodied practices with their schemas held in memory (Cobley, 2013; Delafield-Butt & Adie, 2016; Delafield-Butt & Trevarthen, 2015; Gratier & Trevarthen, 2008; Negayama et al., 2015). These narratively co-regulated projects, completed in infancy, lay the bedrock for all manner of later learning. The brainstem processes underpinning them regulate the goal-directed movements that represent the initial stage in the conceptual development of patterns of motor behaviours that form the first schemas. Before higher forms of linguistic and conceptual learning can take place, it is necessary that affective and embodied patterns be embedded. As such, where narrative co-regulation is thwarted, the downstream consequences can be profound.

For example, Bruner and Feldman (1993) in their narrative deficit hypothesis argue that autism spectrum disorder (ASD) develops as a result of infant inability to create and participate in joint, non-verbal narrative meaning-making. They propose that ASD symptomatology could be the result of a deficit in narrative communication skills, and an inability to organise and structure one's own experiences into a narrative framework. This deficit also impacts the ability to understand narratives used by others, and engage in co-regulation through a narrative framework. ASD can be linked with intersubjective narratives through a dependency on accurate prospective timing and rhythm, which are integral to the effective integration of motor intentions (Delafield-Butt et al., 2020; Daniel, 2019; Daniel et al., 2022). The affective state created as a result of thwarted motor intentions results in distress and isolation leading to emotional and social compensations (Trevarthen & Aitken, 2001; Trevarthen & Delafield-Butt, 2013). This disruption likely has its roots in the brainstem (Dadalko & Travers, 2018), as the successful execution of goal-directed movement requires subsecond accuracy of timing to allow for the coordination of muscle groups shared in preverbal narrative (Delafield-Butt & Trevarthen, 2017), which can lead to disruption in the formation of Panksepp's (2005) core self, and wider ASD symptomology (Delafield-Butt, Dunbar, and Trevarthen, 2022).

Evidence of narrative structuring in the pant hooting of chimpanzees, and the importance of this behaviour in social bonding points to the potential importance of narratives and co-regulation in non-human primates. This is an area of narrative research ripe for further exploration that would help develop our understanding not only of narratives across species, but also their role in our evolution. The continued advancement of machine learning techniques also holds considerable promise in developing our understanding of the role of narrative in co-regulation in both chimpanzees and human beings.

Currently much of our understanding of narratives is limited to human beings, where narratives and the synrhythmic co-regulation they facilitate plays an integral role in the learning and connection that characterises much of early childhood, and indeed later life. For learning and social participation to take place there must be arousal, interest, organisation of action, affect, and the expressive energy of movement, all of which must be regulated for successful and unthwarted co-created meaning. The dyadic nature of intersubjectivity and meaning-making continues to be integral to human experience throughout life, as connection itself "is the regulation and co-creation of the age-possible meanings individuals make of the world and their place in it. The making of meanings is dyadic and continuous" (Tronick, 2005 p21). Were this process not organised and structured in accordance with narrative, human experience would be profoundly altered.

References

Bard, K. A., Maguire-Herring, V., Tomonaga, M., & Matsuzawa, T. (2019). The gesture 'Touch': Does meaning-making develop in chimpanzees' use of a very flexible gesture? *Animal Cognition*, *22*(4), 535–550. https://doi.org/10.1007/s10071-017-1136-0

Bernstein, N. A. (1967). *The co-ordination and regulation of movements*. Oxford: Pergamon Press.

Beebe, B., & Gerstman, L. J. (1980). The "packaging" of maternal stimulation in relation to infant facial-visual engagement: a case study at four months. *Merrill-Palmer Quarterly of Behavior and Development*, *26*(4), 321–339. http://www.jstor.org/stable/23084041

Beebe, B., Stern, D., & Jaffe, J. (1979). The kinesic rhythm of mother-infant interactions. In A. W. Siegman & S. Feldstein (Eds.), *Of speech and time: Temporal patterns in interpersonal contexts* (pp. 23-34). Hillsdale, NJ: Lawrence Erlbaum.

Beebe, B., Jaffe, J., Feldstein, S., Mays, K., & Alson, D. (1985). Interpersonal timing: The application of an adult dialogue model to mother-infant vocal and kinesic interactions. In T. Field & N. Fox (Eds.), *Social perception in infants* (pp. 217-247). Norwood, NJ: Ablex.

Bråten, S. (Ed.). (1998). *Intersubjective communication and emotion in early ontogeny*. London: Cambridge University Press.

Bruner, J. S. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

Bruner, J. S. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Bruner, J. S. (2004). Life as Narrative. *Social Research, 71*(3), 691-710.

Bruner, J. S. & Feldman, C. (1993). Theories of mind and the problem of autism. In S. Baron-

Cohen, H. Tager-Flusberg, D. J. Cohen (Eds.), *Understanding other minds: Perspectives from autism*. Oxford: Oxford University Press.

Burnham, D. (1993). Visual recognition of mother by young infants: Facilitation by speech. *Perception*, *22*(10), 1133–1153. https://doi.org/10.1068/p221133

Buzsáki, G. (2006). *Rhythms of the brain*. Oxford: Oxford University Press.

Bystrova, K., Widström, A.M., Matthiesen, A.S., Ransjö-Arvidson, A.B., Welles-Nyström, B., Wassberg, C., Vorontsov, I., & Uvnäs-Moberg, K. (2003). Skin-to-skin contact may reduce negative consequences of "the stress of being born": A study on temperature in newborn infants, subjected to different ward routines in St. Petersburg. *Acta Paediatrica, International Journal of Paediatrics*, *92*(3), 320–326. https://doi.org/10.1080/08035250310009248

Ciaunica, A., Safron, A., & Delafield-Butt, J. (2021). Back to square one: the bodily roots of conscious experiences in early life. *Neuroscience of Consciousness, 2021*(2). https://doi.org/10.1093/nc/niab037

Cobley, P. (2013). *Narrative* (2nd ed.). New York : Routledge.

Condon, W. S., & Sander, L. W. (1974). Synchrony demonstrated between movements of the neonate and adult speech. *Child Development*, *45*(2), 456–462. https://doi.org/10.2307/1127968

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. *Journal of Psycholinguistic Research*, *31*(1), 1–23. https://doi.org/10.1023/A:1014301303616

Dadalko, O. I., & Travers, B. G. (2018). Evidence for brainstem contributions to autism spectrum disorders. *Frontiers in Integrative Neuroscience*, *12*, 47. https://doi.org/10.3389/fnint.2018.00047

Damásio, A. (1999). The feeling of what happens: Body and emotion in the making of consciousness. London: Vintage.

Damásio, A. (2010). *Self comes to mind: Constructing the conscious brain.* New York: Pantheon Books.

Daniel, S. (2019). Loops and jazz gaps: engaging the feedforward qualities of communicative musicality in play therapy with children with autism. *The Arts in Psychotherapy, 65*, 101595. https://doi.org/https://doi.org/10.1016/j.aip.2019.101595

Daniel, S., Wimpory, D., Delafield-Butt, J., Malloch, S., Holck, U., Geretsegger, M., Tortora, S., Osborne, N. et al. (2022). Rhythmic relating: bidirectional support for social timing in autism therapies. *Frontiers in Psychology*. *13*. https://doi.org/10.3389/fpsyg.2022.793258.

Dautenhahn, K. (2001). The narrative intelligence hypothesis: In search of the transactional format of narratives in humans and other social animals. In M. Beynon, C. L. Nehaniv, & K. Dautenhahn (Eds.), *Cognitive technology: Instruments of mind* (pp. 248–266). Berlin Heidelberg: Springer.

Dautenhahn, K. (2002). The origins of narrative: In search of the transactional format of narratives in humans and other social animals. *International Journal of Cognition and Technology*, *1*(1), 97–123. https://doi.org/10.1075/ijct.1.1.07dau

Decasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928

Delafield-Butt, J. (2018). The emotional and embodied nature of human understanding:

Sharing narratives of meaning. In C. Trevarthen, J. Delafield-Butt & A.W. Dunlop (Eds.), *The child's curriculum: Working with the natural voices of young children*. Oxford: Oxford University Press.

Delafield-Butt, J. T. (2014). Process and action: Whitehead's ontological units and perceptuomotor control units. In S. Koutroufinis (Ed.), *Life and Process* (pp. 133-156). Berlin, Boston: De Gruyter.

Delafield-Butt, J. T., & Adie, J. (2016). The embodied narrative nature of learning: Nurture in school. *Mind, Brain and Education*, *10*(2), 117–131. https://doi.org/10.1111/mbe.12120

Delafield-Butt, J. T., Freer, Y., Perkins, J., Skulina, D., Schögler, B., & Lee, D. N. (2018). Prospective organization of neonatal arm movements: A motor foundation of embodied agency, disrupted in premature birth. *Developmental Science, 21*(6), e12693. https://doi.org/doi:10.1111/desc.12693

Delafield-Butt, J., Dunbar, P., & Trevarthen, C. (2022). Disruption to the core self in autism, and its care. *Psychoanalytic Inquiry, 42*(1), 53-75. https://doi.org/10.1080/07351690.2022.2007031

Delafield-Butt, J. T., & Trevarthen, C. (2013). Theories of the development of human communication. In P. Cobley & P. Schultz (Eds.), *Theories and models of communication* (pp. 199-222). Berlin/Boston: De Gruyter Mouton.

Delafield-Butt, J. T., & Trevarthen, C. (2015). The ontogenesis of narrative: from moving to meaning. *Frontiers in Psychology*, *6,* 1–16. https://doi.org/10.3389/fpsyg.2015.01157

Delafield-Butt, J. T., & Trevarthen, C. (2017). On the brainstem origin of autism: disruption to movements of the primary self. In: E. Torres, and C. Whyatt eds. *Autism: The movement sensing perspective* (pp. 119-138) Boca Raton, FL: Taylor & Francis, CRC

Press

Delafield-Butt, J., & Trevarthen, C. (2020). Infant intentions: Learning with others. In M. Peter (Ed.), *Encyclopedia of teacher education*. Singapore: Springer Nature. https://doi.org/10.1007/978-981-13-1179-6

Delafield-Butt, J. T., Zeedyk, M. S., Harder, S., Vaever, M. S., & Caldwell, P. (2020). Making meaning together: Embodied narratives in a case of severe autism. *Psychopathology, 53*(2), 60-73. https://doi.org/10.1159/000506648

Deppermann, A., & Streeck, J. (2018). The body in interaction: Its multiple modalities and temporalities. In A. Deppermann & J. Streeck (Eds.), *Time in Embodied Interaction: Synchronicity and Sequentiality of Multimodal Resources* (pp. 1-25). John Benjamins. https://doi.org/10.1075/pbns.293.01dep

Dunbar, R. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, *22*(6), 469–493. https://doi.org/https://doi.org/10.1016/0047-2484(92)90081-J

Dunbar, R. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, *16*(4), 681–694. https://doi.org/DOI: 10.1017/S0140525X00032325

Dunbar, R. (1998). The social brain hypothesis. *Evolutionary Anthropology*, *6*(5), 178–190. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8

Eiselt, M., Schindler, J., Arnold, M., Witte, H., Zwiener, U., & Frenzel, J. (2001). Functional interactions within the newborn brain investigated by adaptive coherence analysis of EEG. *Neurophysiologie Clinique*, *31*(2), 104–113. https://doi.org/10.1016/S0987-7053(01)00251-9

Elowson, A. M., & Snowdon, C. T. (1994). Pygmy marmosets, cebuella pygmaea, modify vocal structure in response to changed social environment. *Animal Behaviour*, *47*(6), 1267–1277. https://doi.org/10.1006/anbe.1994.1175

Fantasia, V., & Delafield-Butt, J. (2023). Time and sequence as key developmental dimensions of joint action. *Developmental Review*, 69, 101091.

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

Fedurek, P., Schel, A. M., & Slocombe, K. E. (2013). The acoustic structure of chimpanzee pant-hooting facilitates chorusing. *Behavioral Ecology and Sociobiology*, *67*(11), 1781–1789. https://doi.org/10.1007/s00265-013-1585-7

Fedurek, P., Zuberbühler, K., & Dahl, C. D. (2016). Sequential information in a great ape utterance. *Scientific Reports*, *6*(December), 1–11. https://doi.org/10.1038/srep38226

Feldman, R. (2006). From biological rhythms to social rhythms: Physiological precursors of mother-infant synchrony. *Developmental Psychology*, *42*(1), 175–188. https://doi.org/10.1037/0012-1649.42.1.175

Feldman, R., Magori-Cohen, R., Galili, G., Singer, M., & Louzoun, Y. (2011). Mother and infant coordinate heart rhythms through episodes of interaction synchrony. *Infant Behavior and Development*, *34*(4), 569–577. https://doi.org/10.1016/j.infbeh.2011.06.008

Feldman, R., Rosenthal, Z., & Eidelman, A. I. (2014). Maternal-preterm skin-to-skin contact enhances child physiologic organization and cognitive control across the first 10 years of life. *Biological Psychiatry*, *75*(1), 56–64. https://doi.org/10.1016/j.biopsych.2013.08.012

Field, T. M., Woodson, R., Greenberg, R., & Cohen, D. (1982). Discrimination and imitation of facial expressions by neonates. *Science*, *218*(4568), 179–181. https://doi.org/10.1126/science.7123230

Fogel, A. (1977). Temporal organization in mother-infant face-to-face interaction. In H. R. Schaffer (Ed.), *Studies in mother-infant interaction: Proceedings of the Loch Lomond symposium* (pp. 119-152). London, UK: Academic Press.

Fogel, A. (1993). Two principles of communication: co-regulation and framing. In J. Nadel & L. Camaioni (Eds.), *New perspectives in early communicative development*, (pp. 9–21). London: Routledge and Kegan Paul.

Fontaine, R. (1984). Imitative skills between birth and six months. *Infant Behavior and Development*, *7*(3), 323–333. https://doi.org/10.1016/S0163-6383(84)80047-8

Fransson, A. L., Karlsson, H., & Nilsson, K. (2005). Temperature variation in newborn babies: Importance of physical contact with the mother. *Archives of Disease in Childhood: Fetal and Neonatal Edition*, *90*(6), 500–504. https://doi.org/10.1136/adc.2004.066589

Gallagher, S. (2011). Narrative competency and the massive hermeneutical background. In Paul Fairfield (Ed.), *Hermeneutics in Education* (pp. 21-38). New York: Continuum.

Geva, R., Dital, A., Ramon, D., Yarmolovsky, J., Gidron, M., & Kuint, J. (2017). Brainstem as a developmental gateway to social attention. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *58*(12), 1351–1359. https://doi.org/10.1111/jcpp.12746

Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Gratier, M. (2003). Expressive timing and interactional synchrony between mothers and infants: cultural similarities, cultural differences, and the immigration experience. *Cognitive Development, 18*, 533-554.

Gratier, M., & Apter-Danon, G. (2009). The musicality of belonging: Repitition and variation in mother-infant vocal interaction. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality*. Oxford: Oxford University Press.

Gratier, M., & Trevarthen, C. (2008). Musical narrative and motives for culture in mother-infant vocal interaction. *Journal of Consciousness Studies*, *15*, 122–158.

Grossmann, T., Johnson, M. H., Lloyd-Fox, S., Blasi, A., Deligianni, F., Elwell, C., & Csibra, G. (2008). Early cortical specialization for face-to-face communication in human infants. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1653), 2803–2811. https://doi.org/10.1098/rspb.2008.0986

Ham, J., & Tronick, E. (2009). Relational psychophysiology: Lessons from mother-infant physiology research on dyadically expanded states of consciousness. *Psychotherapy Research*, *19*(6), 619–632. https://doi.org/10.1080/10503300802609672

Hammal, Z., Cohn, J.F., & Messinger, D.S. (2015) Head movement dynamics during play and perturbed mother-infant interaction. *IEEE Trans Affect Comput,* 6, 361–370.

Hattori, Y., & Tomonaga, M. (2019). Rhythmic swaying induced by sound in chimpanzees (Pan troglodytes). *Proceedings of the National Academy of Sciences, 117*(2), 936–942. doi:10.1073/pnas.1910318116

Hattori, Y., Tomonaga, M., & Matsuzawa, T. (2013). Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Scientific Reports*, 3(1), 1566. doi:10.1038/srep01566

Hayes, L. A., & Watson, J. S. (1981). Neonatal imitation: Fact or artifact? *Developmental*

Psychology, *17*(5), 655–660. https://doi.org/10.1037/0012-1649.17.5.655

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

Hrdy, S. B. (2009). *Mothers and others: The evolutionary orgins of mutual understanding*. Cambridge, MA: Belknap Pres

Hubley, P., & Trevarthen, C. (1979). Sharing a task in infancy. In I. C. Uzgiris (Ed.), *New directions for child development* (Vol. 4), *social interaction and communication during infancy,* (pp. 57–80). San Francisco: Jossey-Bass.

Hutto, D. D. (2007). Narrative and understanding persons. *Royal Institute of Philosophy Supplement*, *60*(August), 1–16. https://doi.org/10.1017/s135824610700001x

Ishijima, K., & Negayama, K. (2017). Development of mother–infant interaction in tickling play: The relationship between infants' ticklishness and social behaviors. *Infant Behavior and Development, 49*, 161-167. https://doi.org/https://doi.org/10.1016/j.infbeh.2017.08.007

James, W. (1890). *The principles of psychology*. New York: Dover.

Keevallik, L. (2015). Coordinating the temporalities of talk and dance. In A. Deppermann & S. Günthner (Eds.), *Temporality in Interaction* (pp. 309-336). John Benjamins.

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'trier social stress test'--a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*(1-2), 76–81. https://doi.org/10.1159/000119004

Koepke, J. E., Hamm, M., Legerstee, M., & Russell, M. (1983). Neonatal imitation: Two failures to replicate. *Infant Behavior and Development*, *6*(1), 97–102.

https://doi.org/10.1016/S0163-6383(83)80012-5

Kugiumutzakis, G. (1985). *Development of imitation during the first six months of life (Uppsala Psychological Reports No. 377)*. Uppsala, Sweden, Uppsala University.

Kugiumutzakis, G., & Trevarthen, C. (2015). Neonatal imitation. In J. Wright (Ed.) *International encyclopedia of the social & behavioral sciences: Second edition* (pp. 481–488). Oxford: Elsevier https://doi.org/10.1016/B978-0-08-097086-8.23160-7

Lashley K.S. (1951). The problem of serial order in behavior. *Cerebral Mechanisms in Behavior*, *7*, 112–136.

Leclère, C., Avril, M., Viaux-Savelon, S., Bodeau, N., Achard, C., Missonnier, S., Keren, M., Feldman, R., Chetouani, M., & Cohen, D. (2016). Interaction and behaviour imaging: A novel method to measure mother-infant interaction using video 3D reconstruction. *Translational Psychiatry*, 6(5). https://doi.org/10.1038/tp.2016.82

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 of the United States of America*, *114*(50), 13290–13295. https://doi.org/10.1073/pnas.1702493114

Lewis, T., & Amini, F. (2001). *A general theory of love* (Reprint edition). New York: Vintage Books.

Macmurray, J. (1957). *The self as agent*. London: Faber and Faber.

Malloch, S. N. (1999). Mothers and infants and communicative musicality. *Musicae Scientiae*, *3*(1_suppl), 29–57. https://doi.org/10.1177/10298649000030s104

Malloch, S., & Trevarthen, C. (2009). Musicality: Communicating the vitality and interests of life. In S. Malloch and C. Trevarthen (eds.) *Communicative musicality: Exploring the* *basis of human companionship.* (pp. 1–11). Oxford: Oxford University Press.

Malloch, S., & Trevarthen, C. (2018). The human nature of music. *Frontiers in Psychology*, 9, 1–21. https://doi.org/10.3389/fpsyg.2018.01680

Martin, K. B., Hammal, Z., Ren, G., Cohn, J. F., Cassell, J., Ogihara, M., Britton, J. C., Gutierrez, A., & Messinger, D. S. (2018). Objective measurement of head movement differences in children with and without autism spectrum disorder. *Molecular Autism*, *9*, 14. https://doi.org/10.1186/s13229-018-0198-4

Matsuzawa, T. (2007). Comparative cognitive development. *Dev Sci*, 10(1), 97-103. doi:10.1111/j.1467-7687.2007.00570.x

Matsuzawa, T. (2020). Pretense in chimpanzees. *Primates*, 61(4), 543-555. doi:10.1007/s10329-020-00836-z

Meltzoff, A. N. (2007). “Like me”: A foundation for social cognition. *Developmental Science*, *10*(1), 126–134. https://doi.org/10.1111/j.1467-7687.2007.00574.x

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*(4312), 75–78. https://doi.org/10.1126/science.198.4312.75

Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, *54*(3), 702–709.

Meltzoff, A. N., & Moore, M. K. (1989). Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology, 25*(6), 954–962. https://doi.org/10.1037/0012-1649.25.6.954

Merker, B. (2005). The liabilities of mobility: A selection pressure for the transition to consciousness in animal evolution. *Consciousness and Cognition, 14*(1), 89-114.

Messinger, D,S., Mahoor, M,H., Chow S,M., & Cohn, J.F. (2009). Automated measurement of facial expression in infant-mother interaction: a pilot study. *Infancy*, 14, 285–305.

Messinger, D. M., Ruvolo, P., Ekas, N. V, & Fogel, A. (2010). Applying machine learning to infant interaction: the development is in the details. *Neural Networks : The Official Journal of the International Neural Network Society*, *23*(8–9), 1004–1016. https://doi.org/10.1016/j.neunet.2010.08.008

Milde, T., Putsche, P., Schwab, K., Wacker, M., Eiselt, M., & Witte, H. (2011). Dynamics of directed interactions between brain regions during interburst-burst EEG patterns in quiet sleep of full-term neonates. *Neuroscience Letters*, *488*(2), 148–153. https://doi.org/10.1016/j.neulet.2010.11.018

Mitani, J C, & Brandt, K. L. (1994). Social factors influence the acoustic variability in the long-distance calls of male chimpanzees. *Ethology*, *96*(3), 233–252.

Mitani, J. C., & Gros-Louis, J. (1998). Chorusing and call convergence in chimpanzees: tests of three hypotheses. *Behaviour*, *135*(9–10), 1041–1064. https://doi.org/10.1163/156853998792913483

Mitani, J. C., & Nishida, T. (1993). Contexts and social correlates of long-distance calling by male chimpanzees. *Animal Behaviour*, *45*(4), 735–746. https://doi.org/https://doi.org/10.1006/anbe.1993.1088

Morgan, B. E., Horn, A. R., & Bergman, N. J. (2011). Should neonates sleep alone? *Biological Psychiatry*, *70*(9), 817–825. https://doi.org/10.1016/j.biopsych.2011.06.018

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*(November), 1–

9. https://doi.org/10.1038/srep39019

Nagy, E. (2006). From imitation to conversation: The first dialogues with human neonates. *Infant and Child Development*, *15*, 223–232. https://doi.org/10.1002/icd.460

Nagy, E. (2011). The newborn infant: A missing stage in developmental psychology. *Infant and Child Development*, *20*, 3–19. https://doi.org/10.1002/icd.683

Nagy, E., & Molnar, P. (2004). Homo imitants or homo provocans? Human imprinting model of neonatal imitations. *Infant Behavior and Development*, *27*(1), 54–63. https://doi.org/10.1016/j.infbeh.2003.06.004

Negayama, K. (2011). Kowakare: A New Perspective on the Development of Early Mother–Offspring Relationship. *Integr. Psych. Behav. 45*: 86-99

Negayama, K. (2022). *Parent-Infant Centrifugalism and Centripetalism*. Waseda University Press.

Negayama, K., Delafield-Butt, J. T., Momose, K., Ishijima, K., Kawahara, N., Lux, E. J., Murphy, A., & Kaliarntas, K. (2015). Embodied intersubjective engagement in mother-infant tactile communication: A cross-cultural study of Japanese and Scottish mother-infant behaviors during infant pick-up. *Frontiers in Psychology*, *6*(FEB), 1–13. https://doi.org/10.3389/fpsyg.2015.00066

Negayama, K., Delafield-Butt, J. T., Momose, K., Ishijima, K., & Kawahara, N. (2021). Comparison of Japanese and Scottish mother–infant intersubjectivity: Resonance of timing, anticipation, and empathy during feeding. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.724871

Nelson, K. (1993). The psychological and social origins of autobiographical memory. *Psychological Science*, *4*(1), 7–14. https://doi.org/10.1111/j.1467-9280.1993.tb00548.x

Ngeles Fernández-Gil, M., Palacios-Bote, R., Leo-Barahona, M., & Mora-Encinas, J. P. (2010). Anatomy of the brainstem: A gaze into the stem of life. *Seminars in Ultrasound, CT and MRI*, *31*(3), 196–219. https://doi.org/10.1053/j.sult.2010.03.006

Osborne, N. (2009). Towards a chronobiology of musical rhythm. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship.* (pp. 545-564). Oxford: Oxford University Press.

Panksepp, J. (1998). The periconscious substrates of consciousness: Affective states and the evolutionary origins of the self. *Journal of Consciousness Studies, 5*(5-6), 566–582.

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, *14*(1), 30–80. https://doi.org/10.1016/j.concog.2004.10.004

Panksepp, J., & Biven, L. (2012). *The archaeology of mind: Neuroevolutionary origins of human emotions*. New York: Norton.

Panksepp, J., & Northoff, G. (2009). The trans-species core self: The emergence of active cultural and neuro-ecological agents through self-related processing within subcortical-cortical midline networks. *Consciousness and Cognition, 18*, 193-215. https://doi.org/10.1016/j.concog.2008.03.002

Pezzulo, G., & Castelfranchi, C. (2009). Thinking as the control of imagination: a conceptual framework for goal-directed systems. *Psychological Research*, 73, 559-577.

Piazza, E. A., Hasenfratz, L., Hasson, U., & Lew-Williams, C. (2020). Infant and adult brains are coupled to the dynamics of natural communication. *Psychological Science*, *31*(1), 6–17. https://doi.org/10.1177/0956797619878698

Porges, S. W., & Furman, S. A. (2011). The early development of the autonomic nervous system provides a neural platform for social behaviour: A polyvagal perspective. *Infant and Child Development*, *20*(1), 106–118.

Pouw, W., Trujillo, J. P., & Dixon, J. A. (2019). The quantification of gesture–speech synchrony: A tutorial and validation of multimodal data acquisition using device-based and video-based motion tracking. *Behavior Research Methods*, 52(2), 723–740. https://doi.org/10.3758/s13428-019-01271-9

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*, *37*(July), 100631. https://doi.org/10.1016/j.dcn.2019.100631

Read, S. J., & Miller, L. C. (1995). Stories are fundamental to meaning and memory: For social creatures, could it be otherwise. In R. S. Wyer (Ed.), *Knowledge and memory: The real story* (pp. 139–152). Hillsdale, NJ: Lawrence Erlbaum Associates.

Reddy, V. (2008). *How infants know minds*. Cambridge, MA: Harvard University Press.

Reddy, V., Markova, G., & Wallot, S. (2013). Anticipatory adjustments to being picked up in infancy. *PloS One, 8*(6), e65289. https://doi.org/10.1371/journal.pone.0065289

Reed, E. S. (1996). *Encountering the world: Toward an ecological psychology*. Oxford: Oxford University Press.

Reid, T. (1764). *An inquiry into the human mind on the principles of common sense.* London: A. Millar, and Edinburgh: A. Kincaid and J. Bell. (Republished 1997, Edinburgh University Press). doi: 10.1037/11974-000

Rossmanith, N., Costall, A., Reichelt, A. F., López, B., & Reddy, V. (2014). Jointly structuring triadic spaces of meaning and action: book sharing from 3 months on.

Frontiers in Psychology, 5, 1390. https://doi.org/10.3389/fpsyg.2014.01390

Santamaria, L., Noreika, V., Georgieva, S., Clackson, K., Wass, S., & Leong, V. (2020). Emotional valence modulates the topology of the parent-infant inter-brain network. *NeuroImage*, *207*, 116341. https://doi.org/10.1016/j.neuroimage.2019.116341

Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. *Psychological Bulletin*, *99*(2), 143–165.

Sheets-Johnstone, M. (2011). *The primacy of movement* (2nd ed.). Amsterdam: John Benjamins.

Sherington, C. A. (1906). *The integrative actions of the nervous system.* New York, NY: Scribner's.

Solms, M., & Panksepp, J. (2012). The "id" knows more than the "ego" admits: Neuropsychoanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience. *Brain Science*, 2, 147-174.

Stern, D. N., Beebe, B., Jaffe, J., & Bennett, S. L. (1977). The infant's stimulus world during social interaction: A study of caregiver behaviors with particular reference to repetition and timing. In H. R. Schaffer (Ed.), *Studies in mother-infant interaction: The Loch Lomond symposium* (pp. 177–202). (London: Academic Press).

Stern, D. N. (1971). A micro-analysis of mother-infant interaction: behavior regulating social contact between a mother and her 3 1/2-month-old twins. *Journal of the American Academy of Child Psychiatry*, *10*(3), 501–517.

Stern, D. N. (1974). Mother and infant at play: The dyadic interaction involving facial, vocal, and gaze behaviors. In M. Lewis & L. A. Rosenblum (Eds.), *The effect of the infant on its caregiver* (pp. 187-214). New York, NY: Wiley.

Stern, D. (1985). *The interpersonal world of the infant.* New York: Basic Books.

Stern, D. N., Hofer, L., Haft, W., & Dore, J. (1985). Affect attunement: The sharing of feeling states between mother and infant by means of inter-modal fluency. In T. M. Field & N. A. Fox (Eds.), *Social perception in infants* (pp. 249-268). Norwood, NJ: Ablex.

Stern, D. N. (1999). Vitality contours: The temporal contour of feelings as a basic unit for constructing the infant's social experience. In P. Rochat (Ed.), *Early social cognition: Understanding others in the first months of life* (pp. 67–80). London: Lawrence Erlbaum.

Stern, D. N. (2000). *The interpersonal world of the infant : a view from psychoanalysis and developmental psychology.* (2nd ed.). Basic Books. New York.

Stern, D. N. (2010). *Forms of vitality: exploring dynamic experience in psychology, the arts, psychotherapy, and development*. Oxford: Oxford University Press.

Stewart, J., Gapenne, O., & Di Paolo, E. (2010). *Enaction: Toward a new paradigm for cognitive science.* Cambridge, MA: MIT Press.

Striano, T., Reid, V. M., & Hoehl, S. (2006). Neural mechanisms of joint attention in infancy. *European Journal of Neuroscience*, *23*(10), 2819–2823. https://doi.org/10.1111/j.1460-9568.2006.04822.x

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT Press.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, 10(1), 121-125.

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. doi:10.1017/S0140525X05000129

Terrace, H., Bigelow, A., and Beebe, B. (2022). Intersubjectivity and the origin of words. *Frontiers in Psychology*. In press.

Trevarthen, C. (1979). Communiaction and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech: The beginning of human communication*. (pp. 321-347). London: Cambridge University Press.

Trevarthen, C. (1986). Form, significance, and psychological potential of hand gestures in infants. In J. L. Nespoulous, P. Perron, & A. R. Lecours (Eds.), *The biological foundation of gestures: Motor and semiotic aspects* (pp. 149-202). Hillsdale, NJ: Erlbaum.

Trevarthen, C. (1999). Musicality and the intrinsic motive pulse: Evidence from human psychobiology and infant communication. *Musicae Scientiae, Special Issue Rhythms, Musical Narrative, and the Origins of Human Communication*, 157-213.

Trevarthen, C. (2005). Action and emotion in development of the human self, its sociability and cultural intelligence: Why infants have feelings like ours. In, J. Nadel and D. Muir (Eds.) *Emotional development*, (pp. 61-91) Oxford: Oxford University Press.

Trevarthen, C. (2011). What is it like to be a person who knows nothing? Defining the active intersubjective mind of a newborn human being. *Infant and Child Development*, *20*, 119–135. https://doi.org/10.1002/icd.689

Trevarthen, C. (2012). Embodied human intersubjectivity: Imaginative agency, to share meaning. *Journal of Cognitive Semiotics*, IV(1), 6-56.

Trevarthen, C., & Aitken, K. J. (2001). Infant intersubjectivity: Research, theory, and clinical applications. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(1), 3–48. https://doi.org/10.1017/S0021963001006552

Trevarthen, C, Aitken, K. J., Vandekerckhove, M., Delafield-Butt, J. T., & Nagy, E. (2006). Collaborative regulations of vitality in early childhood: Stress in intimate relationships and postnatal psychopathology. In D. Cicchetti and D. J. Cohen (Eds.), *Developmental Psychopathology* (pp. 65-126). New York: John Wiley and Sons.

Trevarthen, C., & Delafield-Butt, J. T. (2013). Biology of shared meaning and language development: Regulating the life of narratives. In M. Legerstee, D. Haley, & M. Bornstein (Eds.), *The infant mind: Origins of the social brain* (pp. 167-199). New York: Guildford Press.

Trevarthen, C., & Delafield-Butt, J. T. (2017). Development of Consciousness. In B. Hopkins, E. Geangu, & S. Linkenauger (Eds.), *Cambridge encyclopedia of child development* (pp. 821-835). Cambridge: Cambridge University Press.

Trevarthen, C., Delafield-Butt, J. T., & Schögler, B. (2011). Psychobiology of musical gesture: Innate rhythm, harmony and melody in movements of narration. In A. Gritten & E. King (Eds.), *Music and gesture II* (pp. 11-43). Aldershot: Ashgate.

Trevarthen, C., & Hubley, P. (1978). Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. In A. Lock (Ed.), *Action, gesture and symbol* (pp. 183-229). London: Academic Press.

Tronick, E. (2005). Why is connection with others so critical? The formation of dyadic states of consciousness and the expansion of individuals' states of consciousness: coherence governed selection and the co-creation of meaning out of messy meaning making. In J. Nadel & D. Muir (Eds.), *Emotional development: Recent research advances.* (pp. 293–

315). Oxford: Oxford University Press.

Urakawa, S., Takamoto, K., Ishikawa, A., Ono, T., & Nishijo, H. (2015). Selective medial prefrontal cortex responses during live mutual gaze interactions in human infants: An fNIRS study. *Brain Topography*, *28*(5), 691–701. https://doi.org/10.1007/s10548-014-0414-2

Valenza, E., Simion, F., Cassia, V. M., & Umiltà, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 892–903. https://doi.org/10.1037/0096-1523.22.4.892

Vallotton, C. D. (2009). Do infants influence their quality of care? Infants' communicative gestures predict caregivers' responsiveness. *Infant Behavior & Development*, *32*(4), 351–365. https://doi.org/10.1016/j.infbeh.2009.06.001

van der Meer, A. L. H., & van der Weel, F. R. (2011). Auditory guided arm and whole body movements in young infants. In P. Strumillo (Ed.), *Advances in sound localization*. InTech. https://doi.org/https://www.intechopen.com/chapters/15122

Vandekerckhove, M., & Panksepp, J. (2009). The flow of anoetic to noetic and autonoetic consciousness: A vision of unknowing (anoetic) and knowing (noetic) consciousness in the remembrance of things past and imagined futures. *Consciousness and Cognition*, 18, 1018-1028.

Væver, M., Beebe, B., Kirk, O., Snidmann, N., Harder, S., & Tronick, E. (2015). An automated approach for measuring infant head orientation in a face-to-face interaction. *Behavior Research Methods, 47*(2), 328-339. https://doi.org/10.3758/s13428-014-0487-6

Venkatraman, A., Edlow, B. L., & Immordino-Yang, M. H. (2017). The brainstem in emotion: A review. *Frontiers in Neuroanatomy*, *11*, 15. https://doi.org/10.3389/fnana.2017.00015

Vinciarelli, A., Pantic. M., & Bourlard. H. (2009). Social signal processing: survey of an emerging domain. *Image Vis Comput,* 27, 1743–1759

Voinov, P. V., Call, J., Knoblich, G., Oshkina, M., & Allritz, M. (2020). Chimpanzee coordination and potential communication in a two-touchscreen turn-taking game. *Scientific Reports*, 10(1), 3400.

von Hofsten, C. (1993). Prospective control -- A basic aspect of action development. *Human Development*, 36, 253-270.

von Hofsten, C. (2007). Action in development. *Developmental Science*, 10(1), 54-60.

Wass, S. V., Smith, C. G., Clackson, K., Gibb, C., Eitzenberger, J., & Mirza, F. U. (2019). Parents mimic and influence their infant's autonomic state through dynamic affective state matching. *Current Biology*, *29*(14), 2415-2422.e4. https://doi.org/10.1016/j.cub.2019.06.016

Waters, S. F., West, T. V., & Mendes, W. B. (2014). Stress contagion: Physiological covariation between mothers and infants. *Psychological Science*, *25*(4), 934–942. https://doi.org/10.1177/0956797613518352

Winn, P. (2012). Putting the brain into brainstem. *Physiology News, 88*, 29-32.

Zahavi, D. (2007). Self and other: The limits of narrative understanding. *Royal Institute of Philosophy Supplement*, *60*, 179–202. https://doi.org/DOI: 10.1017/S1358246107000094

Figures

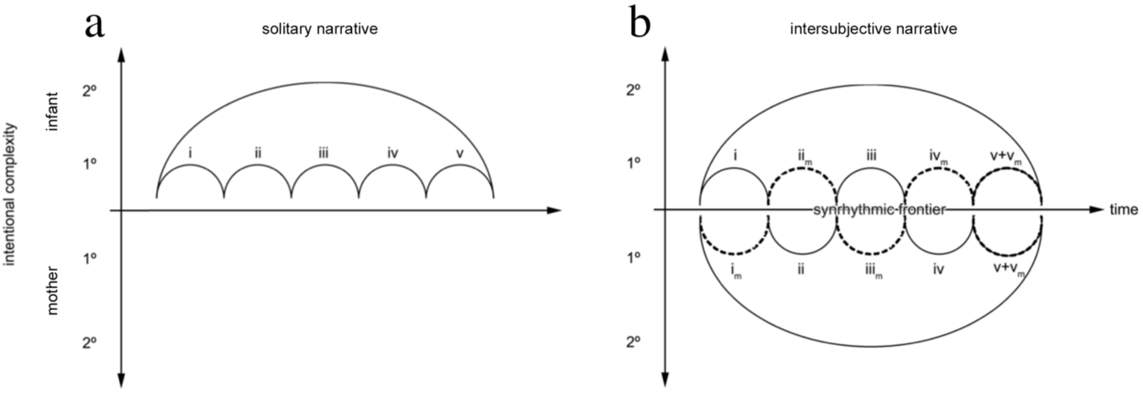


Figure 1.1. The motor foundations of narrative. (a) The nested hierarchy of solitary purposive movement to complete a task, such as ‘to eat’, requires the serial organisation of individual purposive actions, such as (i) a reach, (ii) grasp, (iii) transport, (iv) chew, and (v) swallow. (b) Made with an attentive social other, this serial motor organisation can be shared in time where the expressive actions of one actor (i, ii, iii…) are mirrored in the actions of the observer (i_m, ii_m, iii_m…), enabling turn-taking in protoconversational narrative (c.f. Delafield-Butt & Trevarthen, 2016).

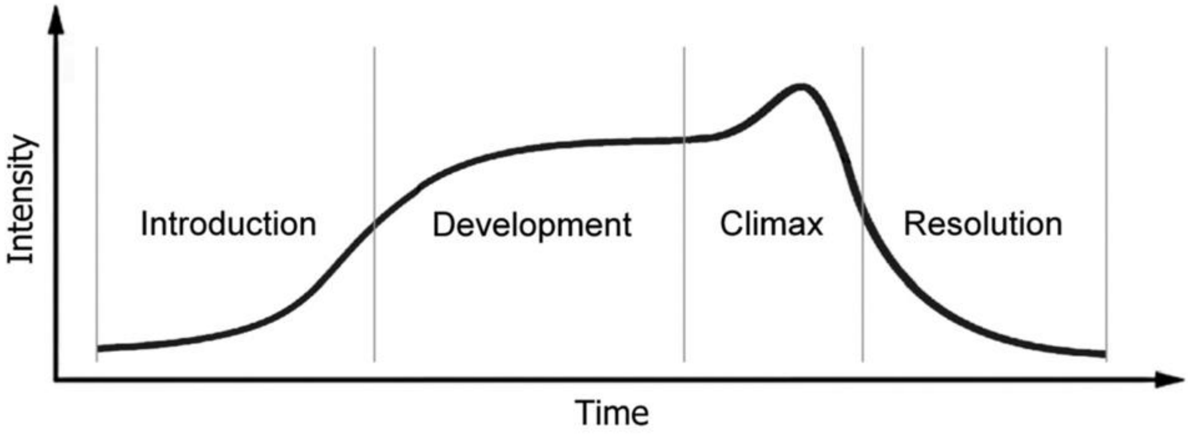


Figure 1.2. An intensity contour illustrating the four-part narrative structure: (i) Interest in the narrative begins at a low-intensity in the introduction, which invites participation in shared purposefulness; (ii) the coordination of the actions and interests of real and imagined agents intensifies over the development, as the plan or project is developed; (iii) a peak moment of excitation in mutual intention is reached at the climax; after which (iv) the intensity reduces as the purposes of the participants share a resolution, and those who were closely engaged may separate, extend, or repeat another cycle. Reproduced with permission from Trevarthen & Delafield-Butt (2013).

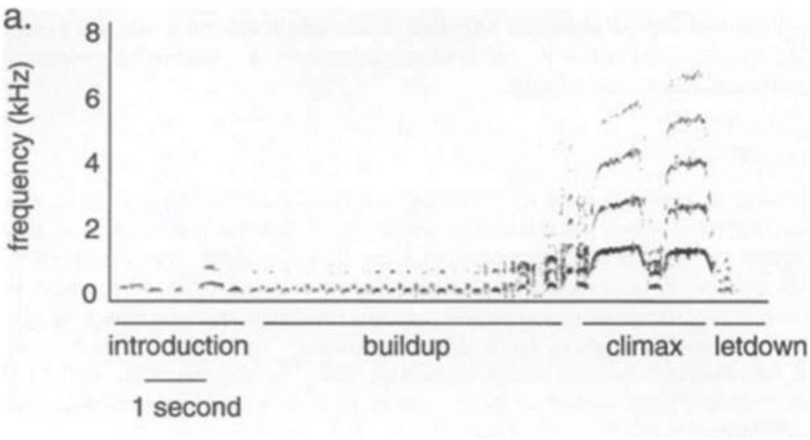


Figure 1.3. Audiospectrogram of a typical chimpanzee pant hoot, illustrating its narrative four-part structure of (i) introduction, (ii) development (buildup), (iv) climax, and (iv) resolution (letdown). Reproduced with permission from Mitani and Gros-Louis (1998).

Chapter 2 - Neonatal Participation in Neonatal Imitation: Narrative in Newborn Dialogues.

This chapter includes sections of previously published work (McGowan, T., & Delafield-Butt, J. (2023). Neonatal Participation in Neonatal Imitation: Narrative in Newborn Dialogues. *Human Development*, 67(3), 135–153. https://doi.org/10.1159/000531311) for which I have been responsible. I was the first author of this work and took the lead in the writing of the manuscript under the guidance of my supervisor.

Abstract

Since the publication of Meltzoff and Moore's seminal paper, neonatal imitation has been discussed, debated and scrutinised at considerable length. Despite this, the temporal structure within which the interaction sits, has received limited attention. We hypothesise underlying successful examples of neonatal imitation exists a narrative temporal structure, expressed and perceived not only through vocalisations but also (if not primarily) through movement. We contextualise neonatal imitation through a communicative lens, viewing the phenomenon as an early dialogue between adult and infant, underpinned by the same narrative structure as other 'proto-conversations' in infancy. From this perspective several of the leading and traditional theories that have been proposed to explain neonatal imitation are considered. Ultimately, we argue neonatal imitation is an innately dialogical phenomenon that forms one of the first examples of primary intersubjectivity, exemplifying the importance of the neonatal period in human psychological and social development. On this basis we propose further study is required into the temporal structure underlying neonatal imitation.

1.0 Narrative in the Neonatal Phase

As discussed in Chapter 1, infant research has shown the capabilities of newborn infants to engage in proto-conversations. These exchanges contain a prelinguistic meaning-making that utilises vocal and bodily gestures to create non-verbal, co-created narratives within proto-conversation (Delafield-Butt & Trevarthen, 2015; Malloch, 1999; Malloch & Trevarthen, 2009). Dautenhahn (2002) considered these pre-verbal narratives the basis of a consciousness in a purposeful social existence. For this to be true, narratives do not need to be based in language. Indeed, all narratives are rooted and expressed in movements of the body.

On this basis, it is also possible that a narrative architecture underpins successful examples of neonatal imitation. Neonatal imitation can be argued to represent sensitivity to another human being (Trevarthen, 2001) and may serve as an early example of early reciprocal communication present long before the development of language (Nagy & Molnar, 2004). It has been proposed that such imitative interaction provides an initial means of communication for infants with caregivers, and may aid in the development of conceptual understanding of others as those with whom goals and interests can be shared (Meltzoff & Moore, 1998). Viewing this phenomenon through a dialogical lens allows us to consider it as an early example of primary intersubjectivity, and of being a precursor to other proto-conversations already identified between adults and infants (e.g. Delafield-Butt & Trevarthen, 2015; Gratier, 2003; Jaffe et al., 2001; Stern, 1977, 1999). On this basis, we might expect imitative interactions to possess the same generative architecture and characteristics as other proto-conversations.

However, in a laboratory setting neonatal imitation is simply defined as movements (usually involving the face, hands or fingers) or vocalisations made by infants in the first month after birth that imitate those made by a human or non-human model. Within this strict experimental paradigm, a model presents a movement that can be a facial or bodily expression, or a sound, and an infant's response is monitored for attempts at responding with the same movement or sound. It is important to note that despite the ongoing debate surrounding neonatal imitation, it is not necessarily the neonate's ability to imitate that is of the greatest importance when considering the phenomenon. Indeed, there has been limited examination of neonatal imitation in natural settings with parents, and in the studies that have been completed it is a behaviour that is displayed rarely during interactions with carers (Simpson et al., 2014). It is perhaps more useful to view neonatal imitation as one of multiple ways neonates can express themselves, and so it forms part of much more nuanced, and complex, natural behavioural repertoire (Trevarthen, 1979).

1.1 Imitation as Communicative Social Learning

In this chapter, we propose that through study of the temporal organisation of neonatal imitation, it will become possible to delve more deeply into the fundamental capacity of neonates to engage with another individual's actions and agency, and therefore to better understand the role of this particular, peculiar element in neonatal social behaviour. The importance of imitative interactions in building attachment and in developing cultural intelligence has been supported in a wealth of psychological research (e.g. Nadel, 2014), and in recent neuroscientific findings (Ammaniti & Gallese, 2014). Indeed, imitation through infancy allows infants to socially connect with and learn from others. Meltzoff and Marshall (2018) argue that even though other animal species have been found to imitate, in humans there is a unique ability to imitate across modalities (from bodily movements to vocalisations), and that human children are highly motivated to be like their caregivers and other adults. This helps underpin the diverse cultural outcomes found in human societies, as imitation forms a bridge between infants and caregivers, from whom cultural information is gathered, re-embodied, then remembered and understood.

Imitation plays a fundamental role in establishing what Piaget came to term the child's developing 'sensorimotor intelligence' (Piaget, 1953, 1954). From the very beginning of child psychology, Baldwin recognised imitative mechanisms as fundamental for learning the patterns of one's own body, made with self-conscious awareness, "The self-repeating or 'circular' reaction... is seen to be fundamental and to remain the same, as far as structure is concerned, for all motor activity whatever: the only difference between higher and lower function being, that in the higher, certain accumulated adaptations have in time so come to overlie the original reaction, that the conscious state which accompanies it seems to differ per se from the crude imitative consciousness in which it had its beginning." (Baldwin, 1895, p. 23, italics our emphasis).

Baldwin's "circular reaction" is a conscious act that explores repetition and deviation from the model (antecedent) made creatively in subsequent self-generated movement. In this case, he is concerned with intra-personal imitation, *per se*. But we now know how self-generated movement and observation of others' actions harness the same neural systems (Rizzolatti & Sinigaglia, 2007). Indeed, Meltzoff and Marshall (2020) review and describe the neurological basis of infant bodily processing (in regard to the infant's own body and bodies belonging to other individuals) and, stemming from this, propose an early recognition of the correspondence between self and other in terms of bodily representations. This forms the foundation for the future generation of a shared social intelligence (Bråten, 2009; Rizzolatti & Sinigaglia, 2008).

The notion of neonatal imitation having a social or communicative function has been expressed for some time. Meltzoff & Moore's (1997) active intermodal matching mechanism (AIM), for example, described a mechanism underpinning neonatal imitation that also forms the basis for understanding the intentional acts of others. Meltzoff and Gopnik (1993) believed it is through this mechanism and the act of imitation that a neonate begins to identify a person as “like me” relative to other inanimate or physical objects. Similarly, Trevarthen (1979, 1980) proposed an innate readiness in infants to know another human being, which allows them to intrinsically appreciate the “other” as being separate from themselves and facilitates direct contact between them and their carers (alongside other adults). The phenomenon of neonatal imitation sits comfortably within this ideal of an infant mind being predisposed to social interaction with other human beings, with Bråten (1988), Maratos (1982), and Užgiris (1991) all linking neonatal imitation with the wider concept of social communication. The work of Nagy (Nagy & Molnar, 2004; Nagy 2006) has furthered this idea, describing neonatal imitation as “the first dialogue”, and demonstrating the infant’s ability to initiate imitative interactions, as well as simply imitating an adult or experimenter. When we view neonatal imitation in this way, as opposed to an action-response phenomenon, its importance with regard to human (and even non-human primate) social capacities becomes more prominent. Seen as an interaction, neonatal imitation can be considered the first step on the road to wider social cognition. Based on the argument presented by Fuchs and De Jaegher (2009), social understanding and social cognition are rooted in the embodied interactions we engage in from the earliest age. They argue it is the interaction process that gives rise to intersubjectivity: through the reciprocity, musicality, affective attunement, vocal expression, and movement of an interaction, social cognition emerges, and meaning is created and transformed in what is described as participatory sensemaking (De Jaegher & Di Paolo, 2007, 2008; De Jaegher, 2009). Looking at neonatal imitation through an intersubjective lens (where the temporal organisation of the interaction plays a fundamental role in the successful communicative creation of imitative acts of meaning) could allow for a deeper understanding not only of the function of neonatal imitation in human and non-human primates, but also a reappraisal of the current debate and the mixed findings that continue to fuel disagreement within the field.

2.0 Adult-Infant Attunement a Necessary Prerequisite for Narrative Meaning Making and Neonatal Imitation

For narrative structure to successfully form it is necessary for each of the participants to be aware, mindful and attentive to one another's movements, actions and vocalisations. Such expressions are always intentional (Delafield-Butt & Gangopadhyay, 2013), so it is necessary that the sensorimotor systems of each of the participants are appropriately attuned. This process of the participants becoming attuned to one another is of paramount importance as it is only when this is achieved that individuals are able to gain an understanding of another's expressive acts and respond appropriately. In the realm of neonatal imitation, Anisfeld (1991) found that a key component for successfully finding evidence of imitation is the length of time for which an action is modelled. When an action is modelled for less than 60 seconds, many pre-1990 studies failed to find evidence for neonatal imitation. Such short periods of modelling could prevent the attunement of neonate and experimenter, and hamper the development of a narrative-based interaction, which ultimately leads to a lower probability of finding evidence of neonatal imitation.

Debate surrounding the extent of infant awareness and self-awareness is informed by neonatal imitation. Evidence for the existence of neonatal imitation lends support to the possibility that there exists in human beings an innate awareness of self and other long prior to the development of language. In contrast to many earlier models of infant development, those informed by evidence of imitation in early infancy tend to support the presence of relatively complex cognitive concepts and mechanisms, including (to at least some degree) an awareness of the self, but also intersubjective abilities and motor control. That is not to say such theories argue in favour of a fully developed sense of selfhood, nor adult-like reflective and conceptually-backed understanding of the other. Rather, it is proposed there exists in infants the foundations onto which future reflective awareness is built. Meltzoff and Moore (2000) proposed neonates have access to "initial mental structures" that serve to develop more advanced mechanisms and concepts. For them, neonatal imitation is a "discovery procedure" possessed by infants that aids in their understanding other persons. Gallagher and Meltzoff (1996) followed a similar line suggesting that neonates are capable of a basic differentiation between self and other and have a rudimentary understanding that their own body is like that belonging to another human being. More recently, Meltzoff and Marshall (2020) have brought together neuroscientific and behavioural evidence to suggest that even very young infants can represent their own bodily actions with those of others in "commensurate terms" (Meltzoff & Marshall p. 11). For Meltzoff, the self and other are linked (Meltzoff 2007, 2013), with infants being able to learn about the internal states of others through perceptual representation of the bodily acts of others, and their internal monitoring.

Another key feature that appears to be required for narratives to successfully form is a "rhythmic temporal pattern"; the common tempo or pulse between participants that structures the organisation of an exchange, with actions, movements, and vocalisations occurring on or within a rhythm or beat (Malloch, 1999). This stems from the notion of communicative musicality that identifies and maps the reciprocal exchange that shares and reflects imitative acts within a common, shared time (see Chapter 1 for a more in depth discussion of this). Within adult-infant interaction, each partner adds creative flourish to enhance and contribute to the dialogue as two jazz musicians might riff off each other's expressive beat (Malloch & Trevarthen, 2009). In this way, imitation forms a basic building block of the exchange, but it is never a mechanical imitation devoid of affective and personal expression, but rather one that acts in mirror reciprocity with creative reflection in each generative expression (Malloch, 1999; Trevarthen, 1999). Sharing a pulse also means this shared timing can be manipulated to create dramatic effects. For example, tension can be raised by withholding an expression on the beat, or excitement increased by coming in early (Gratier & Trevarthen, 2008). Interestingly, Trevarthen (1979) suggests that social interaction and communication can be achieved through the imitation of such rhythm and musicality, and although subtle changes may be more difficult to consciously detect than the traditional mouth opening or tongue protrusion from the neonatal imitation paradigm, their role in creating a successful imitative trial could be paramount.

The importance of movement in the embodied interactions of communicative musicality highlights another important aspect of human communication: it is multi-modal. To fully appreciate the narrative structures that underpin communication in infancy or neonatal imitation, it is therefore necessary to consider not only the role of vocal expression in their creation but also that of movement (Trevarthen & Delafield-Butt, 2013; Delafield-Butt & Trevarthen, 2015). Through a consideration of this modality within the neonatal imitation paradigm, the role of shared rhythm of body and voice, their expressive qualities and their reciprocal exchanges between partners, altogether generate a narrative form in which the infant is able to parse and draw meaning from that interaction. This expected narrative form and its intersubjective contact is important for psychologists to appreciate in any assessment of an infant's ability to imitate a "model" action. It is our proposal that this modelling is intrinsically a part of the infant's attempt to form meaning, or understanding *vis-à-vis* its adult modeller, through participation in the shared narrative.

3.0 Movement is Communicative and the Temporal Nature of Adult-Infant Interactions

Within the history of research into adult-infant interactions, the role of bodily movement has played a distant second fiddle to a predominant focus on vocalisations, facial gestures, and eye contact (Adolph et al., 2010; Pérez & Español, 2016). However, the work of researchers such as Beebe, Stern, Condon, and Sanders serves to demonstrate the importance of movement as a mode of communication within adult-infant interactions. For example, in a landmark study, Condon and Sander (1974) demonstrated the sensitivity of infants to the motor impulses of adult speech and described the synchronised fashion in which they found infants moved to their mother's rhythm and speech as being like an interactive dance that preludes the development of language. Stern *et al.* (1977) further highlighted the importance of a multimodal analytical approach that considered the temporal nature of an interaction. However, they found that the vocal units and movement units within an interaction did not arise in synchrony but did share similarities in terms of frequency of occurrence, their temporal structural form, and their occurrence in unison and independent of one another. These findings demonstrate the need to consider movement units as independent units of communication and the necessity of measuring them when considering adult-infant interactions.

It is not just the movements and actions of the infant that are worthy of attention, though. Other studies have highlighted the communicative nature of adult movements alongside those of the infant (Beebe & Gerstman, 1980; Stern, 1974). It is in the movement of both participants within the dyadic interaction (as much as within their vocal patterning) that evidence of an underlying narrative structure can be found. In neonatal imitation studies, in particular, there is (understandably) an overwhelming focus on the behaviours of the infant (particularly upon their facial expressions), but by also including the temporal organisation of the behaviours of the adult in the analysis of these paradigms, a more complete understanding of the imitative interaction can be discerned.

The focus on infants' ability to imitate facial expressions within the neonatal imitation paradigm stems from the assertions of Piaget that an infant's ability to imitate actions they cannot view themselves arises later in development than observable self-body movements (Piaget, 1951). The analysis of neonatal imitative interactions has also primarily been focused on the infant's facial responses, with minimal attention paid to the infant's body (other than as an indicator of levels of arousal (Nagy et al., 2013)). However, the studies discussed in this section provide a strong rationale for a focus not only on the vocal exchanges and facial expressions that take place during adult-infant interactions but also on the role of movement originating from the bodies of both neonate and adult. Beyond the vocalisations and facial expressions of a parent and child, there is rich and purposeful communication, and it is only through a multimodal approach that it is possible to gain a complete picture of this and appreciate the infant's full and embodied response during the imitative exchange.

4.0 Neonatal Imitation as Primary Intersubjectivity

The interactions outlined above can be described as intersubjective, with a form of this (primary intersubjectivity) being present in humans from birth (evident in the proto-conversations of mother and infant (Delafield-Butt & Trevarthen, 2015)). Primary intersubjectivity is created within dyadic interactions in the earliest stages of postpartum life (it is important to keep in mind that the communication between mother and baby is not simply a one-way dialogue originating from the mother, but rather an interaction of joint attention between two individuals that contains joint meaning). It is through a process of turn taking and responsiveness to one another's emotions, bodily gestures, facial expressions as well as vocalisations that intersubjectivity is achieved (Stern, 2000). Tronick (2005) argues that it is through the back-and-forth responses of mother and infant, which build upon the emotions and purpose of the partner, that joint meaning is established, a so-called “dyadic state of consciousness”. In cases where individuals are unable to establish joint meaning and experience primary intersubjectivity, development can be severely negatively impacted (Delafield-Butt et al., 2020).

The importance of intersubjectivity stems from the narrative form that underpins its structure and function. Delafield-Butt and Adie (2016) describe how the interactive processes involved in intersubjectivity act as a mechanism for making sense of the world - " Meaning is co-created within the organization of the interaction, structured by the agencies and power of will with its expressions of interest and intention from both sides, constituting something unique: a dyadic (or greater) unit that is more than the sum of its parts" (Delafield-Butt & Adie, 2016, p. 118). Through the coming together of two independent beings with their own agency, feelings and aims, something original and new is created in their narrative-based interaction in a process that forms the foundation for learning. Just as this is true for the proto-conversations of mother and baby or the exchanges of teacher and student, so it could be true of the interaction of experimenter and neonate within the imitation paradigm.

4.1 Structures Potentially Responsible for the Communicative Nature of Movement in Intersubjectivity

Alternative suggestions have been made regarding the underlying biological structures responsible for intersubjectivity. One such structure believed to be influential is the mirror neuron system. Initially discovered in Macaque monkeys, the system allows for the organisation of actions whilst also producing “reflective” firing patterns when observing another individual's actions. The term reflective refers to the mirror neuron system's activation in the same manner when a specific action is performed by an individual and when it is being observed. It can be argued this allows for the development of a rudimentary understanding of the intentions of another individual. This would not require higher level cognitive development nor executive functions, but through synchronicity between the motor intelligence of two individuals and their nervous systems, there would be the potential for an understanding of another's motor intentions (Sinigaglia & Rizzolatti, 2011). The mirror neuron system has also been highlighted as a possible candidate in the discussion surrounding the underlying mechanisms of neonatal imitation. The immaturity of the prefrontal cortex in the neonatal brain also raises the possibility of subcortical components of the mirror neuron system, or other subcortical structures being at play in both neonatal imitation and early forms of intersubjectivity. When discussing the role of the mirror neuron system in this regard, it is important to think of it in the context in which it operates. It is within embodied interactions that its true importance comes to the fore, and the link between action and perception is most pronounced in the creation of social understanding (Fuchs & De Jaegher, 2009). However, an overreliance on mirroring and the mirror neuron system to explain social development has been criticised for being too reductive. For example, Reddy and Uithol (2016) argue it is insufficient for explaining action understanding, and suggest that this is better achieved through active engagement and direct interaction. This perspective emphasizes that learning and understanding occur through participatory social interactions, where contextual and relational factors play crucial roles – an argument with clear relevance for our reframing of neonatal imitation.

Another possible structure that might underpin the development of intersubjectivity is the polyvagal system. Porges and Daniel (2017) outline this as a structure that allows for the sharing of regulatory control of certain bodily functions (e.g., breathing and heart rate). However, whilst the initial purpose of the system in evolutionarily more primitive organisms was limited to autonomic physiology, it has evolved in mammals to play a much more dynamic function (Porges & Furman, 2011). Through the brainstem, what was once a system limited to autonomic physiology becomes a powerful social engagement system that influences the intonation of actions made by our hands, our facial expressions, and even our voices. This link between autonomic physiology and expressive forms of communication allows humans to socially demonstrate elements of our vitality, wellbeing, and needs. This in turn develops the basic form of mutual understanding seen in intersubjectivity.

The role of the brainstem in underpinning the development of narratives and intersubjectivity potentially extends far beyond the polyvagal system, as it functions across a variety of areas relevant for the development of sensorimotor activity and autonomic functioning. In the realm of the sensorimotor, it has a role in the up-take of proprioceptive sensory information in addition to hearing taste and touch (all of which play a part in the self-regulation of communication and movement (Dadalko & Travers, 2018; Merker, 2005, 2013; Panksepp & Northoff, 2009; Venkatraman et al., 2017). The autonomic nervous system, which is located in the brainstem, controls eating, alertness, sleep regulation, heart rate, and breathing (Ngeles Fernández-Gil et al., 2010). Moreover, the principal responsibility for actions that are responsive to rhythm lies with the brainstem (Delafield-Butt & Trevarthen, 2017). In addition to these functions, the brainstem has an important role in some higher-order functions such as sensory processing, goal-orientated behaviour, the modulation of emotions (Berntson & Micco, 1976; Venkatraman et al., 2017), and the regulation of social attention (Geva et al., 2017).

Of the varied functions the brainstem is involved in, its impact on higher-order functions, including sociability and attention, is clearly important with regard to intersubjectivity and communication; however, its role in regulating behaviour and emotion is also hugely relevant in the creation of a basic form of consciousness. The integration of sensory information that takes place in the brainstem is not a passive process, and it has been argued this is what creates basic psychological experiences and ultimately leads to Panksepp's (2005) notion of the primary, or “core self”. Indeed, Panksepp (2005) argues that the same brainstem functions that are responsible for sensory and motor information integration are responsible for the generation of basic conscious experiences. He calls these “primary process functions”, and they include core feelings as well as the desire and intention to act through movement. This brainstem-based primary consciousness, that is able to anticipate and perceive, represents an adaptable form of mental agency which generates purpose and meaning (Delafield-Butt & Trevarthen, 2015). If human language and thought are considered as deriving from the same neuro-anatomical systems that generate motor actions in response to stimuli in the environment, the basis of this could be the brainstem sensorimotor and affective integrative systems outlined here (Merker, 2007). Following birth, this primary self actively seeks to share and create meaning with social others, and this desire to build and share narratives is present in the earliest interactions of human life and can be seen, Kugiumutzakis and Trevarthen (2015) believe, in the infant’s power to imitate expressive actions and cooperate in their temporal sequencing. The ability of the infant neonate to imitate could be viewed as an expression of this primary consciousness to create meaning through a narratively structured imitative dialogic interaction.

5.0 Neonatal Imitation in Light of Narrative Temporal Structuring

The notion of a primary consciousness seeking to share and create meaning with others links well with the idea of neonates being “adapted” for narrative-based interactions rooted in communicative musicality. New-born infants display sensitivity and awareness of the purposeful movements and vocalisations of adults in relation to their own person (Condon & Sander, 1974; Nagy, 2011). In the hours after birth human infants have been shown to demonstrate an ability to take part in imitative exchanges with an adult (e.g., Kugiumutzakis, 1998; Meltzoff & Moore, 1983a, 1989), a conversation rooted in movement (Kugiumutzakis & Trevarthen, 2015). These imitative exchanges are considered to be a foundation for social-communication (Meltzoff, 2007) and even, as demonstrated by Nagy and Molnar (2004), a communicative act in and of themselves. They can also be considered the first step in the gradual development through infancy on the road to theory of mind (Meltzoff, 2007) and intersubjectivity (Trevarthen, 2011). Indeed, these imitative exchanges represent some of the first examples of cooperation between infant and adult, an initial form of the primary intersubjectivity outlined above. The infant’s ability to engage in these exchanges is one of multiple social skills (others include the ability to identify their mother's voice (Decasper & Fifer, 1980) and face (Burnham, 1993) and a preference for humanoid faces (Valenza et al., 1996)) that constitute what some consider an innate disposition for intersubjectivity (Nagy & Molnar, 2004). These interactions help the development of cooperative endeavours involving movement, which become more intricate and prospective in their nature (Nagy, 2011). Each imitative act can be thought of as forming part of a larger, encompassing narrative, which creates meaning over multi-second encounters in "mutually sustained expectation of a rhythmic project of communication, engaging two persons in a nonverbal 'narrative'" (Kugiumutzakis & Trevarthen, 2015, p. 487). The temporal architecture that frames these imitative exchanges remains a topic that has received extremely limited academic attention (where temporality has been considered, it is normally with a focus on the timing and organisation of infant response (e.g., Heimann & Tjus, 2019; Meltzoff & Moore, 1983a) and not the framework of the interaction as a whole), whilst the existence of the phenomenon is itself still hotly debated.

5.1 Neonatal Imitation in Human Infants

Whilst the study of neonatal imitation goes back over a hundred years, it is only in the last sixty that the phenomenon has received particular attention. It was Meltzoff and Moore's (1977) seminal work that propelled neonatal imitation as a major topic, spurring multiple studies clarifying its nature and supporting its existence (e.g., Field et al., 1982; Heimann et al., 1989; Heimann & Schaller, 1985; Heimann & Tjus, 2019; Kugiumutzakis, 1998; Meltzoff & Moore, 1983a; Meltzoff & Moore, 1989; Nagy & Molnar, 2004; Reissland, 1988). There have also been many studies that have failed to find evidence of the phenomenon (e.g., Davis et al., 2021; Fontaine, 1984; Hayes & Watson, 1981; Koepke et al., 1983; Lewis & Sullivan, 1985; Oostenbroek et al., 2016); and these have in turn been critiqued for using procedures that may have lacked sensitive conditions, both social and perceptual-cognitive, for eliciting imitation in very young infants (e.g., Kugiumutzakis, 1998; Meltzoff & Moore, 1983b; Meltzoff et al., 2018). Additionally, imitation has not only been studied in human neonates; numerous studies have also been conducted that have reported its presence in non-human primates (e.g., Bard, 2007; Ferrari et al., 2006; Myowa-Yamakoshi et al., 2004).

Nagy and Molnar (2004), however, went further than to simply support the existence of neonatal imitation in humans. Their study also reported an infant's ability to initiate interactions. In this study, 45 neonatal infants, aged 2-54 hours, were found to both imitate tongue protrusion, and initiate voluntary exchanges with an adult utilising the tongue protrusion action. The study utilised both behavioural and psychophysiological measures in order to differentiate initiation from imitation. The psychophysiological measure used was heart rate in order to ascertain levels of arousal (increased heart rate was considered reflective of increased levels of arousal) and of orientation, learning, and expectance (decreased heart rate was taken to be indicative of these during imitation). Following several cycles of imitation, it was found that the neonates not only imitated the modelled action but also initiated an exchange and waited for the adult to respond, therefore demonstrating a motivational desire to take part in the dialogue.

Such evidence suggests social motivation could underlie neonatal imitation. Several examples of extended exchanges between neonate and experimenter, which were initiated by the infant and grew into what the authors describe as reciprocal conversations, were reported during the Nagy and Molnar's study (Nagy, 2006). These exchanges were regarded as experimental examples of "the first dialogue" (Nagy & Molnar, 2004; Nagy, 2006, 2011) and allow us to consider neonatal imitation one of the foundation stones of primary intersubjectivity. Given the presence of narratives in these early proto-conversations and evidence from Nagy and Molnar (2004) regarding the social and emotional motivations for the infant in engaging with adults in imitative exchanges, the presence of a narrative structure in the interactions of the neonatal imitation paradigm would be a logical extension to the current state of understanding in intersubjectivity.

Indeed, it is within the paradigm utilised by Nagy and Molnar (2004) that narrative structures seem most likely to be found. There are key methodological differences between the work of Nagy and Molnar and other neonatal imitation studies, for example, Meltzoff and Moore (1977). In many such earlier studies, the modelling by experimenters was very rigid and experimentally controlled, with only limited consideration of the infant's engagement and response. In Nagy and Molnar (2004), the experimental design was more interactive and representative of imitative exchanges that occur naturally between adults and infants. The impact of this can be seen in the differences in response times of infants in such studies. In Nagy and Molnar (2004), imitation occurred at a rate that enabled the infants to participate in communicative exchanges with adults, with infants imitating in a shorter time frame than had previously been demonstrated in other studies. The responsiveness of infants in the study, and the ability they demonstrated to participate in these early forms of turn-taking dialogues and primary intersubjectivity, make this experimental structure a prime place to search for evidence of narrative structures within the neonatal imitation paradigm.

5.2 Non-human Primates

As noted, neonatal imitation has not only been observed in human neonates - there is also a body of evidence demonstrating the phenomenon's presence in chimpanzees (Myowa-Yamakoshi et al., 2004; Bard,2007) and rhesus macaques (Ferrari et al., 2006; Ferrari et al., 2009; Paukner et al., 2011). Myowa-Yamakoshi et al. (2004) and Bard (2007) both demonstrated the ability of neonatal chimpanzees to imitate mouth opening and tongue protrusion of a human modeller, believing that their results supported the notion of an innate neonatal imitation “mechanism” in chimpanzees as well as humans. With regard to macaques, Ferrari et al. (2006) utilised a sample of 21 animals aged three days and found evidence that tongue protrusion and lip smacking were regularly imitated when presented by a human model. Ferrari et al. (2009) utilised this data and as well as collecting data from an additional 20 neonatal macaques and found further evidence supporting the findings of the 2006 study (although the ability to imitate seemed to dissipate from age seven days). Additionally, it was found that at one month of age, macaques who were shown to be better imitators demonstrated more developed motor skills in goal directed movements. This further demonstrates the potential close link between movement, neonatal imitation, and intersubjectivity. The authors believed the mechanisms to explain this result lay in the maturation of the motor and parietal cortices, which are close in proximity to parts of the mirror neuron system.

Evidence has also continued to emerge from non-human primate studies linking neonatal imitation with future social development. For example, neonatal macaques who reliably and successfully imitate have been found to pay more attention visually to the eyes of others later in the first month of life (Paukner et al., 2014). Simpson et al. (2016) found evidence that imitative ability in neonatal macaques predicted gaze following at 7 months, therefore demonstrating the role of imitation in modulating a social cognitive skill later in development. Additionally, Kaburu et al. (2016) showed neonatal imitative ability to be a predictor of social behaviour and temperament (in the form of greater dominance behaviour and lower levels of anxiety among lip smacking imitators) at one year old. These studies are beginning to highlight the importance of neonatal imitation not only as a neonatal phenomenon, but also as a predictor of other wider social behaviours and point towards the underlying links between social/motor development and neonatal imitation (possibly stemming from shared underlying cognitive structures involving, for example, the brainstem).

Such non-human primate results are important as there are strong similarities between the observed imitation in human and macaque neonates (Ferrari et al., 2013). For one, imitation is recorded most prominently in the first few weeks of life in both humans and macaques. Additionally, in more naturalistic paradigms, the mother is found to imitate their infant more than the infant does the mother. It has also been found that in both species there is a large degree of individual difference with regard to the quality of imitation. This, it has been suggested, could be the result of differing levels of social predisposition between individuals (Paukner *et al.*, 2014) or an infant's sensitivity to social cues within the multi-modal interactional flow between adult and infant (Heimann, 1998). Other factors linked with individual differences have also been proposed such as levels of alertness and activity, visual acuity, temperament (Heimann, 2022), and expressiveness (Field, 1982).

As outlined earlier, the presence of narrative structures within chimpanzee pant-hoot chorusing already demonstrates the existing influence such temporal structures have on the social interactions of some non-human primates and how the temporal structure is actively manipulated to enhance social bonding (Fedurek et al., 2013). It is such similarities between human and non-human primates, and the potential links between imitation and potential downstream social development, that make the investigation of narrative structure in primate development an attractive area that could shed further light onto the development of human social interaction and neonatal imitation.

6.0 Theories Behind Neonatal Imitation

From the varied findings in the field there continues to be a high level of debate around the existence of neonatal imitation (e.g., Meltzoff et al., 2018; Oostenbroek et al., 2016), as well as multiple theories to explain the mechanisms that underlie it. Some of these theories attempt to explain neonatal imitation as a phenomenon that is not imitation per se, but that is simply an associative reaction devoid of any social meaning, whilst others consider it an innate ability. For the purposes of this paper a range of key theories will be discussed. However, rather than simply providing a summary of the theories and explanations that exist, we aim to contextualise them in light of intersubjectivity and consider how they align with the view of neonatal imitation as a proto-dialogue between two actively engaged persons with agency.

6.1 Imitation as a Reflex

One explanation of the underlying causes of neonatal imitation proposes that there is actually a relatively simplistic mechanism underpinning positive imitative results. This theory argues that the imitative acts recorded in studies of neonatal imitation are in fact non-mental reflexes (Anisfeld, 1996), a triggering of an involuntary motor response or a fixed action pattern that is activated upon witnessing the demonstration of a particular behaviour. Anisfeld (1996) believed that the only act where there was evidence of neonates imitating was tongue protrusion and claims of imitative abilities in early infancy were too “uncritically accepted”.

He took the view that such claims were actually facilitated by a trend to view infants as more advanced than they were in reality (Anisfeld actually points the finger very clearly at Dan Stern in this regard). It has been demonstrated in some studies that neonates will perform tongue protrusion actions in response to other similar stimulus events (such as a ball or felt tip pen being moved towards the face of the infant (Jacobson, 1979)), and such findings have been interpreted as signifying the potential for multiple protruding objects to release a sucking reflex which creates the illusion of an imitative response.

However, considering neonatal imitation results as reflex reactions does not entirely remove communication from the paradigm. Another interesting adaptive function of such an innate releasing mechanism, outlined by (Bjorklund, 1987), is that it facilitates social interaction between an infant and adult. Bjorklund believed that neonatal imitation was an example of a transient ontogenetic adaptation (Oppenheim, 1981) in that it serves a specific survival purpose for the neonate, which becomes redundant later in development. For Bjorklund, neonatal imitation helps maintain social interaction between adult and infant, but this ability becomes less important when infants develop greater control of head and mouth movements. Were the “imitation as reflex” hypothesis an accurate explanation of neonatal imitation, the role of narrative temporal structuring in its facilitation may still be compatible, adding temporal detail to the description. Despite this, the “imitation as reflex” theory risks doing the opposite of what Anisfeld accused Dan Stern of, by over-simplifying what is an apparent complex interaction between infant and adult.

6.2 Imitation Due to Arousal

Another popular explanation of the apparent empirical evidence of neonatal imitation relates to the excitation of arousal of the neonate in the paradigm. An example of this was proposed by Anisfeld (1991), who described how attentional competition (originating from the modelling adult) inhibits infant tongue protrusion during the modelling period. This results in a build-up of what can be defined as arousal, which is released following the completion of action modelling. A more recent examination of neonatal imitation, which employs the arousal hypothesis, looks at the phenomenon in the context of aerodigestive development (Keven & Akins, 2017). It is clear the arousal hypothesis remains a popular account of neonatal imitation (e.g., Anisfeld, 2005; Jones, 2009; Keven & Akins, 2017; Vincini et al., 2017), a fact not least due to continuing debate around the imitation of actions beyond tongue protrusion (acceptance of neonatal imitation in actions other than tongue protrusion makes it difficult to ascribe the behaviour of the infant to increased levels of arousal) but also other key factors. These include the aforementioned findings that other stimuli can also illicit tongue protrusion (balls and felt tip pens (Jacobson, 1979), lights and toys (although only until the infant had developed the ability to reach, at which point the rate of tongue protrusion was found to decrease significantly, leading to the suggestion that reaching may have become the infant's primary response to arousing stimuli) (Jones, 1996), and music (Jones, 2006)), and neonates looking for longer at displays of tongue protrusion than they do mouth opening (Jones, 1996).

This leads to one of the key weaknesses in the arousal hypothesis: much of the evidence given to support the theory stems from tongue protrusion, and it does not take full account of infants' abilities to imitate a variety of adult modelled actions. Additionally, Nagy et al. (2013) focused on infants' ability to imitate tongue protrusion and found imitation of this behaviour was not simply the result of arousal during the testing process. Her own findings from Nagy and Molnar (2004) found increasing heart rate when an infant imitates whilst decreasing heart rate was recorded during unprompted actions. As discussed above, this would suggest there to be different mechanisms underlying imitation when compared to infant initiated behaviours.

However, notwithstanding these limitations of the theory, as with “imitation as reflex”, the arousal hypothesis is not necessarily a non-communicative theory that is incompatible with the suggestion of underlying narrative patterns underpinning neonatal imitation. If adult facial modelling prompts arousal in an observing infant, and this triggers a response, the observing adult could be drawn into an exchange of facial expressions/movements that take the form of a non-verbal dialogue. Such an arousal-driven dialogue could be underpinned by the same narrative structures found to underlie other examples of early social interactions between adults and infants, and it would not be necessary for adults to be aware of the imitative nature of the interaction (indeed, Heimann (2002) suggested parents could be affected by the imitative responses of their infants without actually being aware imitation was taking place). One of the unanswered questions around the arousal hypothesis is exactly how the stimuli affect levels of arousal, and while multiple factors could affect how social stimuli affect the internal states of infants, one factor worthy of consideration is the temporal structure within which the stimuli are presented. Indeed, narrative temporal structuring could be a modulating factor on levels of infant arousal during these interactions.

6.3 Imitation Through Association by Similarity

A relatively new theory put forward to explain neonatal imitation is the association by similarity theory (AST) proposed by (Vincini & Jhang, 2018). Vincini and Jhang (2018) posit a process through which current experiences of cognitive events are connected to past events and experiences of a similar nature; in the words of Vincini and colleagues (2017, p. 9), “AST hypothesizes that the same resources that represent those action features in action planning and execution represent those features in perception”. Within AST, infant imitation is part of an existing repertoire of habitual and spontaneous actions which are “awoken” when they witness corresponding action being modelled (Vincini & Jhang, 2018). AST can sit alongside an arousal account of how infant reactions in the imitation paradigm are provoked, and in itself does not claim that neonates have an in-built ability for imitation, nor that neonatal imitation has a foundational role with regard to social cognition. The theory claims to neatly account for the varied findings across neonatal imitation literature (both in terms of the mixed evidence for the existence of neonatal imitation and any variations in the weight of evidence for imitation of particular acts), the fact it is more common in controlled laboratory than natural settings, and for evidence of imitation falling away after 2-3 months of age (Vincini et al., 2017).

Whilst AST does not view neonatal imitation as acting as a foundation for social cognition in the same way as some other theories of neonatal imitation, it does not disallow it being a communicative exchange. Vincini and colleagues are of the opinion that the same logic can be applied to AST as is applied to the arousal hypothesis with regard to social interaction. Therefore, whatever role arousal plays in facilitating adult-infant interactions in the arousal hypothesis can also be true for AST. They also suggest AST supports the Direct Social Perception hypothesis (Gallagher, 2015) and, therefore, Interaction Theory as an alternative to more traditional theories of mindreading.

6.4 Imitation Due to Associative Learning

Associative learning provides another popular account for neonatal imitation. This is based on the idea that infants can learn to associate their own actions with the movements of other individuals, which then allows for the development of imitative abilities. A major proponent of a version of this argument is Heyes, who in her Associative Sequence Learning theory (Catmur & Heyes, 2019; Heyes, 2001, 2018) proposes that associations between sensory and motor representations form during an infant's development due to correlated sensorimotor experiences. Heyes believes that as a result of these experiences, bidirectional associative links are formed so that the action observation can result in action execution. Such sensorimotor associations should be formed in everyday life, for example, when an infant is imitated by an adult or when adults and infants engage in synchronous activities (Heyes, 2018).

A key issue with the associative learning account for neonatal imitation is the time taken for sensorimotor associations to form within a child. Imitation has been demonstrated in the first few hours of birth for a variety of actions and expressions, presumably too early in the course of development for associative links to have formed in the manner described by Catmur and Heyes (2019) and other proponents of the associative learning account. This objection is equally valid with regard to macaque neonates, who are able to imitate before they have experienced relevant facial interactions with carers (Ferrari et al., 2006). The rate at which infants receive imitative feedback from carers is also not considered to be great enough to build such learned associations (Watson, 1979). Were this not the case it would not be impossible to imagine the bidirectional associations of ASL forming during the early imitative based dialogues and proto-conversations of the neonatal imitation paradigm. There is evidence of some infant attempts slowly building towards full imitative actions such as tongue protrusion (Kugiumutzakis, 1998; Maratos, 1973; Meltzoff & Moore, 1977), and the progressive attempts before successful imitation could be indicative of associative learning taking place. This could potentially indicate a social element to ASL, which could be underpinned by a narrative temporal patterning, ultimately leading to successful imitative responses. Again, this theory is also compatible with our proposition and proposes important angles of consideration.

6.5 Imitation Underpinned by an In-built Cognitive Mechanism

As well as the above explanations, there are others that consider neonatal imitation to be a genuine phenomenon underpinned by an inbuilt cognitive mechanism that allows imitative behaviours to be displayed from birth. One such hypothesis involves a direct sensory-motor matching mechanism underpinned by mirror neurons (Ferrari et al., 2006; Rizzolatti et al., 1999; Rizzolatti et al., 2002). This theory outlines that within the neonatal imitation paradigm the motor system of the observing infant is triggered by the process of watching the experimenter model target actions. For example, when an infant observes an experimenter opening and closing their mouth, the same mirror neurons are triggered in the infant that would be activated were the action performed. This directly leads to the execution of the observed action. This hypothesis has not only been described on a theoretical level but also at the neuroanatomical level in neonatal macaques (Ferrari et al., 2012; Ferrari et al., 2017). This explanation of the underpinning mechanism for neonatal imitation has several advantages when compared to other cognitive alternatives. For example, it requires relatively few cognitive processes and does not necessarily require the infant to have an understanding of why they are performing an action as the process is seemingly largely automated (Vincini et al., 2017). However, this does not remove social importance from neonatal imitation, as imitation underpinned by mirror neurons promotes interaction between neonates and adults (Heimann, 2002), which can become the foundation of future proto-conversations (e.g., Malloch & Trevarthen, 2009). Despite the attractiveness of this cognitive explanation, others have argued that such a resonance mechanism is insufficient to explain neonatal imitation (Meltzoff & Decety, 2003), with an alternative cognitive model being the active intermodal matching mechanism (AIM) (Meltzoff & Moore, 1997).

Meltzoff and Moore's AIM (1997) goes some way towards appreciating how the underlying mechanism for neonatal imitation could link with the future development of children's understanding of the intentional actions of others, joint attention, and theory of mind (Meltzoff, 2007). It proposes that from infants' self-generated movements, proprioceptive feedback is produced which can be compared by the infant to the visual representation perceived during modelling. Meltzoff and Moore proposed this comparison is viable due to a supramodal action space, in which the observation and execution of human movements are coded in a common framework, and that this coding in a common framework forms the basis for the understanding of the intentional actions of others, joint attention and theory of mind later in development (Meltzoff, 2007). Although AIM is one of the leading theories to explain neonatal imitation as innately present in human infants, Kugiumutzakis and Trevarthen (2015) propose that it is not enough to explain the phenomenon on its own. They argue it also "depends on matching organized intentional motor systems that seek confirmation by different forms of reafference." (Kugiumutzakis & Trevarthen, 2015, p. 487). For Kugiumutzakis and Trevarthen, there is a two-way exchange of complementary affective states which involves the regulation of motor impulses by a "hierarchy of rhythms" created in the brains of those involved. It is this element of musicality that directly parallels the underlying mechanisms of neonatal imitation with those of other dialogues and exchanges in human beings. Indeed, as mentioned, in naturalistic exchanges it is not only the neonates who imitate, but also their parents, with both parties timing their expressive movements so as to achieve an intersynchrony between intentions and experiences. This, Kugiumutzakis and Trevarthen believe, is necessary for the construction of joint projects and the creation of narrative structures.

From this perspective, neonatal imitation is more than a stimulus-response exchange, but rather an intersubjective phenomenon that is social in its very nature. Each imitative exchange is not a single, isolated experience, but part of a developing "story" that creates meaning over the multiple seconds in which the overall interaction is maintained

(Kugiumutzakis & Trevarthen, 2015). This interaction takes place within a jointly created intersubjective space in which the neonate is required to detect the adult's intention and motivation for communication, and respond reciprocally (Kugiumutzakis, 1998). Kugiumutzakis believes the factors that allow an infant to discern an adult's motivation for communication lie within the acoustical and kinematic features of the modelled actions. It is in this space that the temporal structure of the interactions becomes paramount. If the modelled actions or sounds are presented as though they are introducing or developing a narrative structure, this could form a key element of what Kugiumutzakis describes as the critical invariants that signify to the infant a desire within an adult for communication.

By moving neonatal imitation beyond a physical process, it can become a coupling of embodied agents coordinated through bodily movement and vocalisation, facial expression, and timing. Recent findings in infant neuroscience have also shown a specific neural pattern can be detected when an infant performs an action and the adult imitates back to the infant. The imitative "dialog," exhibited through the turn-taking of matching of motor actions, causes a change in the infant neural mu rhythm that is significantly different from that observed when the adult does a nonmatching action, indicating there is something particularly noticeable or communicative about the baby being in a mutual imitative interaction with another person (Saby et al., 2012). These early interactions, simplistic though they may seem, become imbued with affective meaning underpinned by the patterns of the interaction. From the implicit experience of this co-created meaning arises the foundation for social understanding and intersubjectivity (Fuchs & De Jaegher, 2009). Reddy (2008) argues this is what makes imitation relevant for the infant, as it allows a bi-directional interplay in which the two engaged parties influence one another in a shared affective resonance within an interactional framework. We extend this argument, proposing the specific underlying framework is narrative in form.

7.0 Conclusion

The potential communicative underlying nature of neonatal imitation makes it fertile ground with regard to developing narrative structures. Imitation represents an effective method of engagement for both young infants yet to develop language skills and slightly older toddlers. It is this that suggests imitation has a function that runs deeper than being purely reactionary to external stimuli. It enables infants to engage in meaningful and purposeful interactions with adults long before the onset of language. Within such interactions there is also the potential for an affective component resonating between neonatal and adult participants. Kugiumutzakis (1998) suggests that a feeling of enjoyment may be created as a consequence of the infant's recognition of the adult as another "other" entering into a shared interactive unit. Indeed, the interactive unit, or intersubjective companion space, acts as a "nest of emotions" (Kugiumutzakis, 1998, p. 79) in which adult and infant can potentially share emotions and create basic acts of meaning. Whilst such an outlook runs a risk of falling foul of Anisfeld's (1996) criticism of imbuing infants with levels of sophistication beyond their years, without considering the potential affective outcomes of such interactions and the infant's innate desire for companionship, the neonatal imitation debate becomes a poorer landscape.

By considering neonatal imitation as innately dialogical (Kugiumutzakis & Trevarthen, 2015; Nagy & Molnar, 2004), it is possible to further grasp its underlying importance in human development and why a temporal analysis of it, similar to that already performed on other interactional exchanges (Delafield-Butt & Trevarthen, 2015), could help us understand its underlying organisation and function. Arbib (2005) suggested that from an evolutionary standpoint, imitation can be viewed as a foundation stone onto which language development was laid, and Nagy (2006) proposed that the same could be true with regard to human development from the neonatal period, through infancy and into childhood. The intersubjective exchange becomes the template on which words and language form (Terrace et al., 2022). This foundational role is supported by the latest evidence stemming from infant macaques linking neonatal imitation with future social development (Kaburu et al., 2016; Simpson et al., 2016).

Building on the proposition that neonatal imitation is dialogical in nature and function, we propose that successful examples of neonatal imitation are underpinned by a narrative framework, in the same manner other adult-infant proto-conversations are (see also Heimann, 2022, p. 184). This narrative architecture would be expressed in a multimodal format (involving movement, vocalisation, and facial expression) and would require a multi-modal analysis to identify the peaks of intensity and arousal expressed in the precise kinematics of movement and acoustic intensity. It seems most likely that narratives will be found within imitation paradigms like that utilised by Nagy and Molnar (2004), in which the neonate was allowed the potential for expressive agency and interest through provocation or initiation. It was within such a paradigm that the role of the infant in creating lasting, intimate, imitative exchanges was first established. However, the same narrative structure could also be present in the model-response pattern utilised by other studies (especially when the movements of the neonate are also considered as part of the response), demonstrating the language-ready brain of the neonate.

The one-month period after birth, commonly referred to as the neonatal phase, is hugely important in terms of development. Nagy (2011) argues that it deserves a specific period of its own in development theory. Narrative structures in neonatal imitation could represent the first step in a developmental trajectory that builds towards larger projects of communication through primary intersubjectivity and secondary intersubjectivity, and from proto-conversations through to the fully developed dialogues of language of later childhood.

References

Adolph, K. E., Tamis-Lemonda, C. S., & Karasik, L. B. (2010). Cinderella indeed - A commentary on Iverson’s developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, *37*(2), 269–273. https://doi.org/10.1017/S030500090999047X

Ammaniti, M., & Gallese, V. (2014). *The birth of intersubjectivity. Psychodynamics, neurobiology and the self.* Norton.

Anisfeld, M. 2005. No compelling evidence to dispute Piaget’s timetable of the development of representational imitation in infancy. *Perspectives on Imitation: From Neuroscience to Social Science* 2, 107–131.

Anisfeld, M. (1991). Neonatal imitation. *Developmental Review*, *11*(1), 60–97. https://doi.org/10.1016/0273-2297(91)90003-7

Anisfeld, M. (1996). Only tongue protrusion modeling is matched by neonates. *Developmental Review*, *16*(2), 149–161. https://doi.org/10.1006/drev.1996.0006

Arbib, M. A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *The Behavioral and Brain Sciences*, 28(2), 105–167. https://doi.org/10.1017/s0140525x05000038

Baldwin, J. M. (1895). *Mental development in the child and the race: method and processes.* Macmillan Pub.

Bard, K. A. (2007). Neonatal imitation in chimpanzees (Pan troglodytes) tested with two paradigms. *Animal Cognition*, *10*(2), 233–242. https://doi.org/10.1007/s10071-006-0062-3

Beebe, B., & Gerstman, L. J. (1980). The “packaging” of maternal stimulation in relation to infant facial-visual engagement: a case study at four months. *Merrill-Palmer Quarterly of Behavior and Development*, 26(4), 321–339. http://www.jstor.org/stable/23084041

Berntson, G. G., & Micco, D. J. (1976). Organization of brainstem behavioral systems. *Brain Research Bulletin.*, *1*(5), 471–483.

Bjorklund, D. F. (1987). A note on neonatal imitation. *Developmental Review*, *7*(1), 86–92. https://doi.org/10.1016/0273-2297(87)90006-2

Bråten, S. (1988). Dialogic mind: The infant and the adult in protoconversation. In M.E. Carvallo (Ed.) *Nature, cognition and system I. Theory and decision library* (vol 2). Springer.

Bråten, S. (2009). *The intersubjective mirror in infant learning and evolution of speech.* John Benjamins Publishing Company.

Burnham, D. (1993). Visual recognition of mother by young infants: Facilitation by speech. *Perception*, *22*(10), 1133–1153. https://doi.org/10.1068/p221133

Catmur, C., & Heyes, C. (2019). Mirroring “meaningful” actions: Sensorimotor learning modulates imitation of goal-directed actions. *Quarterly Journal of Experimental Psychology*, *72*(2), 322–334. https://doi.org/10.1080/17470218.2017.1344257

Cobley, P. (2013). *Narrative* (2nd ed.). Routledge.

Condon, W. S., & Sander, L. W. (1974). Synchrony demonstrated between movements of the neonate and adult speech. *Child Development*, *45*(2), 456–462. https://doi.org/10.2307/1127968

Dadalko, O. I., & Travers, B. G. (2018). Evidence for brainstem contributions to autism spectrum disorders. *Frontiers in Integrative Neuroscience*, *12*, 47. https://doi.org/10.3389/fnint.2018.00047

Dautenhahn, K. (2002). The origins of narrative: In search of the transactional format of narratives in humans and other social animals. *International Journal of Cognition and Technology*, *1*(1), 97–123. https://doi.org/10.1075/ijct.1.1.07dau

Davis, J., Redshaw, J., Suddendorf, T., Nielsen, M., Kennedy-Costantini, S., Oostenbroek, J.,

& Slaughter, V. (2021). Does neonatal imitation exist? Insights from a meta-analysis of 336 effect sizes. *Perspectives on Psychological Science*, *16*(6), 1373–1397. https://doi.org/10.1177/1745691620959834

De Jaegher, H. (2009). Social understanding through direct perception? Yes, by interacting. *Consciousness and Cognition*, 18(2), 535–542.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, 6, 485–507.

De Jaegher, H., & Di Paolo, E. (2008). Making sense in participation. An enactive approach to social cognition. In F. Morganti, A. Carassa & G. Riva (Eds.), *Enacting intersubjectivity: A cognitive and social perspective to the study of interactions*. IOS Press.

Decasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928

Delafield-Butt, J. T., & Adie, J. (2016). The embodied narrative nature of learning: Nurture in school. *Mind, Brain and Education*, 10(2), 117–131. https://doi.org/10.1111/mbe.12120.

Delafield-Butt, J. T., & Gangopadhyay, N. (2013). Sensorimotor intentionality: The origins of intentionality in prospective agent action. *Developmental Review*, *33*(4), 399–425. https://doi.org/10.1016/j.dr.2013.09.001

Delafield-Butt, J. T., & Trevarthen, C. (2015). The ontogenesis of narrative: from moving to meaning. *Frontiers in Psychology*, *6*(September), 1–16. https://doi.org/10.3389/fpsyg.2015.01157

Delafield-Butt, J. T., & Trevarthen, C. (2017). On the brainstem origin of autism: disruption to movements of the primary self. In E. Torres, and C. Whyatt (Eds). *Autism: The movement sensing perspective* (pp. 119-138). Taylor & Francis, CRC Press.

Delafield-Butt, J. T., Zeedyk, M. S., Harder, S., Vaever, M. S., & Caldwell, P. (2020). Making meaning together: Embodied narratives in a case of severe autism. *Psychopathology*, 53(2), 60-73. https://doi.org/10.1159/000506648

Di Paolo, E., & De Jaegher, H. (Eds.). (2015). *Towards an embodied science of intersubjectivity: Widening the scope of social understanding research*. Frontiers Media. https://doi.org/10.3389/978-2-88919-529-9.

Fedurek, P., Schel, A. M., & Slocombe, K. E. (2013). The acoustic structure of chimpanzee pant-hooting facilitates chorusing. *Behavioral Ecology and Sociobiology*, *67*(11), 1781–1789. https://doi.org/10.1007/s00265-013-1585-7

Ferrari, P. F., Visalberghi, E., Paukner, A., Fogassi, L., Ruggiero, A., & Suomi, S. J. (2006). Neonatal imitation in rhesus macaques. *PLoS Biology*, *4*(9), 1501–1508. https://doi.org/10.1371/journal.pbio.0040302

Ferrari, P.F., Paukner, A., Ruggiero, A., Darcey, L., Unbehagen, S., & Suomi, S. J. (2009). Interindividual differences in neonatal imitation and the development of action chains in rhesus macaques. *Child Development*, *80*(4), 1057–1068. https://doi.org/10.1111/j.1467-8624.2009.01316.x

Ferrari, P.F., Paukner, A., & Suomi, S.J. (2013). A comparison of neonatal imitation abilities in human and macaque infants. In S.A. Gelman and M.R. Banaji, (Eds.), *Navigating the social world: what infants, children, and other species can teach us* (pp. 133-138.). Oxford University Press.

Ferrari, P. F., Gerbella, M., Coudé, G., & Rozzi, S. (2017). Two different mirror neuron networks: The sensorimotor (hand) and limbic (face) pathways. *Neuroscience*, *358*, 300–315. https://doi.org/10.1016/j.neuroscience.2017.06.052

Ferrari, P. F., Vanderwert, R. E., Paukner, A., Bower, S., Suomi, S. J., & Fox, N. A. (2012). Distinct EEG amplitude suppression to facial gestures as evidence for a mirror mechanism in newborn monkeys. *Journal of Cognitive Neuroscience*, *24*(5), 1165–1172.

https://doi.org/10.1162/jocn_a_00198

Field, T. (1982). Individual differences in the expressivity of neonates and young infants. In R. S. Feldman (Ed.), *Development of nonverbal behavior in children* (pp. 279–298). Springer-Verlag.

Field, T. M., Woodson, R., Greenberg, R., & Cohen, D. (1982). Discrimination and imitation of facial expressions by neonates. *Science*, *218*(4568), 179–181. https://doi.org/10.1126/science.7123230

Fontaine, R. (1984). Imitative skills between birth and six months. *Infant Behavior and Development*, *7*(3), 323–333. https://doi.org/10.1016/S0163-6383(84)80047-8

Fuchs, T., & De Jaegher, H. (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the Cognitive Sciences*, 8(4), 465–486. https://doi.org/10.1007/s11097-009-9136-4

Gallagher, S. (2015). The new hybrids: Continuing debates on social perception. *Consciousness and Cognition*, 36, 452–465. https://doi.org/10.1016/j.concog.2015.04.002

Gallagher, S., & Meltzoff, A. N. (1996). The sense of self and others: Merleau-Ponty and recent developmental studies. *Philosophical Psychology*, 9, 211–233.

Geva, R., Dital, A., Ramon, D., Yarmolovsky, J., Gidron, M., & Kuint, J. (2017). Brainstem as a developmental gateway to social attention. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *58*(12), 1351–1359. https://doi.org/10.1111/jcpp.12746

Gratier, M. (2003). Expressive timing and interactional synchrony between mothers and infants: Cultural similarities, cultural differences, and the immigration experience. *Cognitive Development, 18*, 533-554.

Gratier, M., & Trevarthen, C. (2008). Musical narrative and motives for culture in mother-infant vocal interaction. *Journal of Consciousness Studies*, *15*(10–11), 122–158.

Hayes, L. A., & Watson, J. S. (1981). Neonatal imitation: Fact or artifact? *Developmental Psychology*, *17*(5), 655–660. https://doi.org/10.1037/0012-1649.17.5.655

Heimann, M. (1989). Neonatal imitation, gaze aversion, and mother-infant interaction. *Infant Behavior and Development*, *12*(4), 495–505. https://doi.org/10.1016/0163-6383(89)90029-5

Heimann, M. (1998). Imitation in neonates, older infants and in children with autism. Feedback to theory. In S. Bråten (Ed.), *Intersubjective communication and emotion in ontogeny* (pp. 89–104). Cambridge University Press.

Heimann, M. (2002). Notes on individual differences and the assumed elusiveness of neonatal imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind: Development, evolution and brain bases* (pp. 74–84). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511489969.005

Heimann, M. (2022). *Imitation from infancy through early childhood typical and atypical development.* Springer. https://doi.org/10.1007/978-3-031-08899-5

Heimann, M., Nelson, K. E., & Schaller, J. (1989). Neonatal imitation of tongue protrusion and mouth opening: methodological aspects and evidence of early individual differences. *Scandinavian journal of psychology*, 30(2), 90–101. https://doi.org/10.1111/j.1467-9450.1989.tb01072.x

Heimann, M., & Schaller, J. (1985). Imitative reactions among 14-21 days old infants. *Infant Mental Health Journal*, *6*, 31–39.

Heimann, M., & Tjus, T. (2019). Neonatal imitation: Temporal characteristics in imitative response patterns. *Infancy*, *24*(5), 674–692. https://doi.org/10.1111/infa.12304

Heyes, C. M. (2001). Trends in cognitive sciences: Causes and consequences of imitation. *Trends in Cognitive Sciences*, *5*(6), 253–261.

Heyes, C. M. (2018). *Cognitive gadgets the cultural evolution of thinking*. The Belknap Press of Harvard University Press.

Hutto, D. D. (2007). Narrative and understanding persons. *Royal Institute of Philosophy Supplement*, *60*(August), 1–16. https://doi.org/10.1017/s135824610700001x

Jacobson, S. W. (1979). Matching behavior in the young infant. *Child Development*, *50*(2), 425–430. https://doi.org/10.2307/1129418

Jaffe, J., Beebe, B., Feldstein, S., Crown, C., & Jasnow, N. (2001). Rhythms of dialogue in infancy. *Monographs of the Society of Research and Child Development., 66*, 1-132.

Jones, S. S. (1996). Imitation or exploration? Young infants' matching of adults' oral gestures. *Child Development*, *67*(5), 1952–1969. https://doi.org/10.1111/j.1467-8624.1996.tb01837.x

Jones, S. S. (2006). Exploration or imitation? The effect of music on 4-week-old infants' tongue protrusions. *Infant Behavior & Development*, *29*(1), 126–130. https://doi.org/10.1016/j.infbeh.2005.08.004

Jones, S. S. (2009). The development of imitation in infancy. *Philosophical Transactions of the Royal Society, Series B: Biological Sciences*, *364*(1528), 2325–2335. https://doi.org/10.1098/rstb.2009.0045

Kaburu, S. S. K., Paukner, A., Simpson, E. A., Suomi, S. J., & Ferrari, P. F. (2016). Neonatal imitation predicts infant rhesus macaque (Macaca mulatta) social and anxiety-related behaviours at one year. *Scientific Reports*, *6*(October), 4–11. https://doi.org/10.1038/srep34997

Keven, N., & Akins, K. A. (2017). Neonatal imitation in context: Sensorimotor development in the perinatal period. *Behavioral and Brain Sciences*, *40*, 16000911. https://doi.org/10.1017/S0140525X16000911

Koepke, J. E., Hamm, M., Legerstee, M., & Russell, M. (1983). Neonatal imitation: Two failures to replicate. *Infant Behavior and Development*, *6*(1), 97–102. https://doi.org/10.1016/S0163-6383(83)80012-5

Kugiumutzakis, G. (1998). Neonatal imitation in the intersubjective companion space. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 63–88). Cambridge University Press.

Kugiumutzakis, G., & Trevarthen, C. (2015). Neonatal imitation. In J. Wright (Ed.) *International encyclopedia of the social & behavioral sciences* (2nd ed,. pp. 481–488). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.23160-7

Lewis, M., & Sullivan, M. W. (1985). Imitation in the first six months of life. *Merrill-Palmer Quarterly, 31*(4), 315–333.

Malloch, S. N. (1999). Mothers and infants and communicative musicality. *Musicae Scientiae, Special Issue Rhythms, Musical Narrative, and the Origins of Human Communication*, 29-57. https://doi.org/10.1177/10298649000030s104

Malloch, S., & Trevarthen, C. (2009). Musicality: Communicating the vitality and interests of life. In S. Malloch and C. Trevarthen (eds.) *Communicative musicality: Exploring the basis of human companionship*. (pp. 1–11). Oxford University Press.

Maratos, O. (1973, April*). The Origin and development of imitation in the first six months of life* [Paper presentation]. *Annual Meeting of the British Psychological Society, Liverpool, England*. https://eric.ed.gov/?id=ED096001

Maratos, O. (1982). Trends in the development of imitation in early infancy. In T. G. Bever (Ed.), *Regressions in mental development: Basic phenomena and theories* (pp. 81–101). Erlbaum.

Marler, P., & Tenaza, R. (1977). Signaling behavior of apes with special reference to vocalization. In T. A. Sebeok (Ed.), *How animals communicate* (pp. 965–1033). Indiana University Press.

Meltzoff, A. N. (1988). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. *Child Development*, *59*(1), 217–225. https://doi.org/10.1111/j.1467-8624.1988.tb03210.x

Meltzoff, A. N. (2007). “Like me”: A foundation for social cognition. *Developmental Science*, *10*(1), 126–134. https://doi.org/10.1111/j.1467-7687.2007.00574.x

Meltzoff, A. N. (2013). Origins of social cognition: bidirectional self-other mapping and the “Like-Me” hypothesis. In M. Banaji & S. Gelman (Eds*.) Navigating the social world: What infants, children, and other species can teach us* (pp. 139–144). New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199890712.003.0025

Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1431), 491–500. https://doi.org/10.1098/rstb.2002.1261

Meltzoff, A. N., & Marshall, P. J. (2018). Human infant imitation as a social survival circuit. *Current Opinion in Behavioral Sciences, 24,* 130-136. https://doi.org/10.1016/j.cobeha.2018.09.006

Meltzoff, A. N., & Marshall, P. J. (2020). Importance of body representations in social-cognitive development: New insights from infant brain science. *Progress in Brain Research*, *254*, 25–48. https://doi.org/10.1016/bs.pbr.2020.07.009

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*(4312), 75–78.https://doi.org/10.1126/science.198.4312.75

Meltzoff, A. N., & Moore, M. K. (1983a). Newborn infants imitate adult facial gestures. *Child Development*, 54(3), 702–709.

Meltzoff, A. N., & Moore, M. K. (1983b). The origins of imitation in infancy: Paradigm, phenomena, and theories. In L. P. Lipsitt (Ed.), *Advances in infancy research* (Vol. 2, pp.

265–301). Norwood, NJ: Ablex.

Meltzoff, A. N., & Moore, M. K. (1989). Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology*, 25(6), 954–962. https://doi.org/10.1037/0012-1649.25.6.954

Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Early Development & Parenting*, *6*(3–4), 179–192. https://doi.org/10.1002/(SICI)1099-0917(199709/12)6:3/4<179::AID-EDP157>3.0.CO;2-R

Meltzoff, A. N., & Moore, M. K. (1998). Infant intersubjectivity: Broadening the dialogue to include imitation, identity and intention. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 47–62). Cambridge University Press.

Meltzoff, A., & Moore, M. K. (2000). Resolving the debate about early imitation. In D. Muir & A. Slater (Eds.), *Infant development: The essential readings* (pp. 176-182). London: Blackwell.

Meltzoff, A., & Gopnik, A. (1993). The role of imitation in understanding persons and developing a theory of mind. In S. BaronCohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), Understanding other minds (pp. 335-366). New York: Oxford University Press.

Meltzoff, A. N., Murray, L., Simpson, E., Heimann, M., Nagy, E., Nadel, J., Pedersen, E. J., Brooks, R., Messinger, D. S., Pascalis, L. De, Subiaul, F., Paukner, A., & Ferrari, P. F. (2018). Re-examination of Oostenbroek et al. (2016): evidence for neonatal imitation of tongue protrusion. *Developmental Science*, *21*(4), 1–8. https://doi.org/10.1111/desc.12609

Merker, B. (2005). The liabilities of mobility: A selection pressure for the transition to consciousness in animal evolution. *Consciousness and Cognition*, *14*(1), 89–114. https://doi.org/10.1016/S1053-8100(03)00002-3

Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences*, *30*(1), 63–81. https://doi.org/10.1017/S0140525X07000891

Merker, B. (2013). The efference cascade, consciousness, and its self: Naturalizing the first person pivot of action control. *Frontiers in Psychology*, *4*(AUG), 501. https://doi.org/10.3389/fpsyg.2013.00501

Mitani, J. C., & Brandt, K., L. (1994). Social factors influence the acoustic variability in the long-distance calls of male chimpanzees. *Ethology*, *96*(3), 233–252.

Mitani, J. C., & Gros-Louis, J. (1998). Chorusing and call convergence in chimpanzees: Tests of three hypotheses. *Behaviour*, *135*(9–10), 1041–1064. https://doi.org/10.1163/156853998792913483

Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., & Matsuzawa, T. (2004). Imitation in neonatal chimpanzees (Pan troglodytes). *Developmental Science*, *7*(4), 437–442. https://doi.org/10.1111/j.1467-7687.2004.00364.x

Nadel, J. (2014). *How imitation boosts development: In infancy and autism spectrum disorder.* (E. Corbett, Trans.). Oxford University Press.

Nagy, E. (2006). From imitation to conversation: The first dialogues with human neonates. *Infant and Child Development*, *15*, 223–232. https://doi.org/10.1002/icd.460

Nagy, E. (2011). The newborn infant: A missing stage in developmental psychology. *Infant and Child Development*, *20*, 3–19. https://doi.org/10.1002/icd.683

Nagy, E., & Molnar, P. (2004). Homo imitants or homo provocans? Human imprinting model of neonatal imitations. *Infant Behavior and Development*, *27*(1), 54–63. https://doi.org/10.1016/j.infbeh.2003.06.004

Nagy, E., Pilling, K., Orvos, H., & Molnar, P. (2013). Imitation of tongue protrusion in human neonates: Specificity of the response in a large sample. *Developmental Psychology*, *49*(9), 1628–1638. https://doi.org/10.1037/a0031127

Negayama, K., Delafield-Butt, J. T., Momose, K., Ishijima, K., Kawahara, N., Lux, E. J., Murphy, A., & Kaliarntas, K. (2015). Embodied intersubjective engagement in mother-infant tactile communication: A cross-cultural study of Japanese and Scottish mother-infant behaviors during infant pick-up. *Frontiers in Psychology*, *6*(FEB), 1–13. https://doi.org/10.3389/fpsyg.2015.00066

Ngeles Fernández-Gil, M., Palacios-Bote, R., Leo-Barahona, M., & Mora-Encinas, J. P. (2010). Anatomy of the brainstem: A gaze into the stem of life. *Seminars in Ultrasound, CT and MRI*, *31*(3), 196–219. https://doi.org/10.1053/j.sult.2010.03.006

Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., Clark, S., & Slaughter, V. (2016). Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Current Biology*, *26*(10), 1334–1338. https://doi.org/10.1016/j.cub.2016.03.047

Oppenheim, R. W. (1981). Ontogenetic adaptations and retrogressive processes in the development of the nervous system and behavior: A neuroembryological perspective. *Maturation and Development: Biological and Psychological Perspectives*, 73–109.

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, *14*(1), 30–80. https://doi.org/10.1016/j.concog.2004.10.004

Panksepp, J., & Northoff, G. (2009). The trans-species core SELF: The emergence of active cultural and neuro-ecological agents through self-related processing within subcortical-cortical midline networks. *Consciousness and Cognition*, *18*(1), 193–215. https://doi.org/10.1016/j.concog.2008.03.002

Paukner, A., Ferrari, P. F., & Suomi, S. J. (2011). Delayed imitation of lipsmacking gestures by infant rhesus macaques (Macaca mulatta). *PLoS ONE*, *6*(12). https://doi.org/10.1371/journal.pone.0028848

Paukner, A., Simpson, E. A., Ferrari, P. F., Mrozek, T., & Suomi, S. J. (2014). Neonatal imitation predicts how infants engage with faces. *Developmental Science*, *17*(6), 833–840. https://doi.org/10.1111/desc.12207

Pérez, S. C., & Español, S. A. (2016). Multimodal study of adult-infant interaction: A review of its origins and its current status. *Paideia*, *26*(65), 377–385. https://doi.org/10.1590/1982-43272665201613

Piaget, J. (1953). *The origin of intelligence in the child* (M. Cook, Trans.). Routledge and Kegan Paul.

Piaget, J. (1954). *The construction of reality in the child* (M. Cook, Trans.). Basic Books.

Piaget, J. (1951). *Play, dreams and imitation in childhood* (1st ed.). Routledge. https://doi.org/10.4324/9781315009698.

Porges, S., & Daniel, S. (2017). Play and the dynamics of treating pediatric medical trauma. In S. Daniel & C. Trevarthen (Eds.), *Rhythms of relating in children's therapies: Connecting creatively with vulnerable children* (pp. 113 – 124). Jessica Kingsley.

Porges, S. W., & Furman, S. A. (2011). The early development of the autonomic nervous system provides a neural platform for social behaviour: A polyvagal perspective. *Infant and Child Development*, *20*(1), 106–118.

Reddy, V. (2008). *How infants know minds*. Harvard University Press.

Reddy, V., & Uithol, S. (2015). Engagement: Looking beyond the mirror to understand action understanding. *British Journal of Developmental Psychology, 34*(1), 101-114. https://doi.org/10.1111/bjdp.12106.

Reissland, N. (1988). Neonatal imitation in the first hour of life: Observations in rural Nepal. *Developmental Psychology*, *24*(4), 464–469. https://doi.org/10.1037/0012-1649.24.4.464

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1999). Resonance behaviors and mirror neurons. *Archives italiennes de biologie*, *137*(2-3), 85–100.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (2002). From mirror neurons to imitation: Facts and speculations. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind: Development, evolution, and brain bases* (pp. 247–266). Cambridge University Press. https://doi.org/10.1017/CBO9780511489969.015

Rizzolatti, G., & Sinigaglia, C. (2007). Mirror neurons and motor intentionality. *Functional Neurology*, *22*(4), 205–210.

Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the brain: How our minds share actions and emotions.* (F. Anderson, Trans.). Oxford University Press.

Saby, J. N., Marshall, P. J., & Meltzoff, A. N. (2012). Neural correlates of being imitated: an EEG study in preverbal infants. *Social neuroscience*, 7(6), 650–661. https://doi.org/10.1080/17470919.2012.691429

Simpson, E. A., Miller, G. M., Ferrari, P. F., Suomi, S. J., & Paukner, A. (2016). Neonatal imitation and early social experience predict gaze following abilities in infant monkeys. *Scientific Reports*, *6*(February), 1–6. https://doi.org/10.1038/srep20233

Simpson, E. A., Murray, L., Paukner, A., & Ferrari, P. F. (2014). The mirror neuron system as revealed through neonatal imitation: Presence from birth, predictive power and evidence of plasticity. *Philosophical Transactions of the Royal Society, Series B: Biological Sciences*, *369*(1644). https://doi.org/10.1098/rstb.2013.0289

Sinigaglia, C., & Rizzolatti, G. (2011). Through the looking glass: Self and others. *Consciousness and Cognition*, *20*(1), 64–74. https://doi.org/10.1016/j.concog.2010.11.012

Stern, D. N. (1974). Mother and infant at play: The dyadic interaction involving facial, vocal, and gaze behaviors. In M. Lewis & L. A. Rosenblum (Eds.), *The effect of the infant on its caregiver* (pp. 187-214). Wiley.

Stern, D. N. (1977). *The first relationship*. Harvard University Press.

Stern, D. N. (1999). Vitality contours: The temporal contour of feelings as a basic unit for constructing the infant's social experience. In *Early social cognition: Understanding others in the first months of life.* (pp. 67–80). Lawrence Erlbaum Associates Publishers.

Stern, D. N. (2000). *The interpersonal world of the infant: a view from psychoanalysis and developmental psychology.* (2nd ed.). Basic Books.

Stern, D. N., Beebe, B., Jaffe, J., & Bennett, S. L. (1977). The infant's stimulus world during social interaction: A study of caregiver behaviors with particular reference to repetition and timing. In H. R. Schaffer (Ed.), *Studies in mother-infant interaction: The Loch Lomond symposium* (pp. 177–202). Academic Press.

Terrace, H. S., Bigelow, A. E., & Beebe, B. (2022). Intersubjectivity and the emergence of words [Review]. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.693139

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech: The beginning of human communication* (pp. 321-347). Cambridge University Press.

Trevarthen, C. (1980). The foundations of intersubjectivity: Development of interpersonal and cooperative understanding in infants. In D. Olson (Ed) *The social foundation of language and thought* (pp 316–342). W. W. Norton.

Trevarthen, C. (1999). Musicality and the intrinsic motive pulse: Evidence from human psychobiology and infant communication. *Musicae Scientiae, Special Issue Rhythms, Musical Narrative, and the Origins of Human Communication*, 157-213. https://doi.org/10.1177/10298649000030S109

Trevarthen, C. (2001). The neurobiology of early communication: Intersubjective regulations in human brain development. In Kalverboer, A.F., & Gramsbergen, A. (Eds.) *Handbook on brain and behavior in human development* (pp. 841–882). Kluwer.

Trevarthen, C. (2011). What is it like to be a person who knows nothing? Defining the active intersubjective mind of a newborn human being. *Infant and Child Development*, *20*, 119–135. https://doi.org/10.1002/icd.689

Trevarthen, C., & Delafield-Butt, J. T. (2013). Biology of shared meaning and language development: Regulating the life of narratives. In M. Legerstee, D. Haley, & M. Bornstein (Eds.), *The infant mind: Origins of the social brain* (pp. 167-199). Guildford Press.

Tronick, E. (2005). Why is connection with others so critical? The formation of dyadic states of consciousness and the expansion of individuals' states of consciousness: Coherence governed selection and the co-creation of meaning out of messy meaning making. In J Nadel & D. Muir (Eds.), *Emotional development: Recent research advances.* (pp. 293–315). Oxford University Press.

Užgiris, I. Č. (1991). The social context of infant imitation. In M. Lewis & S. Feinman (Eds.), *Social influences and socialization in infancy* (pp. 215–251). Plenum Press. https://doi.org/10.1007/978-1-4899-2620-3_10

Valenza, E., Simion, F., Cassia, V. M., & Umiltà, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 892–903. https://doi.org/10.1037/0096-1523.22.4.892

Venkatraman, A., Edlow, B. L., & Immordino-Yang, M. H. (2017). The brainstem in emotion: A review. *Frontiers in Neuroanatomy*, *11*, 15. https://doi.org/10.3389/fnana.2017.00015

Vincini, S., & Jhang, Y. (2018). Association but not recognition: An alternative model for differential imitation from 0 to 2 months. *Review of Philosophy and Psychology* 9 (2) 395–427. https://doi.org/10.1007/s13164-017-0373-0

Vincini, S., Jhang, Y., Buder, E. H., & Gallagher, S. (2017). Neonatal imitation: Theory, experimental design, and significance for the field of social cognition. *Frontiers in*

Psychology, *8*(AUG), 1–16. https://doi.org/10.3389/fpsyg.2017.01323

Watson, J. S. (1979). Perception of contingency as a determinant of social responsiveness. In E. Thoman (Ed.), *Origins of the infant's social responsiveness.* Erlbaum.

Part 2

Prologue

The empirical work that forms the basis of Part 2 of this thesis was undertaken in the shadow of Covid 19. Following issues in acquiring data relating to the neonatal imitation phenomenon, we explored alternative avenues to investigate the temporal architecture of adult-infant interaction. With in-person experimentation still heavily restricted at the time of study, we opted to pursue secondary data that would allow us to analyse the presence and development of narrative in early interactions between adults and infants. A dataset held by the Center for Early Intervention and Family Studies at the University of Copenhagen provided an ideal opportunity for this. Following discussions with researchers at the University of Copenhagen, permission was granted for their data to be used for our analysis. However, due to ethical permissions obtained at the time of testing, and data protection regulations in Denmark, it was not possible for the recordings to be shared across international borders. As such, it was necessary for me to travel directly to Copenhagen to conduct our planned analysis.

This began a lengthy process of applying for funding to cover expenses associated with living costs, travel and accommodation in Copenhagen, as well as permission to conduct overseas fieldwork. To cover costs and expenses I was awarded an Overseas Fieldwork Grant by the Scottish Graduate School of Social Science. However, Covid-19 had created additional internal restrictions on travel at the University of Strathclyde, and a formal application process also needed to be completed within the Institute of Education for permission to be granted from the Head of Institute. I was one of the first individuals to follow this procedure. Following permission being granted I spent a total of 3 months in Copenhagen (split across three separate visits each lasting 1 month) and a total of 1 year conducting research that formed the basis of this chapter.

In addition to the work completed across the following two chapters (the human rater coding that formed the basis of Chapter 3, and the training of a pose estimation model for the tracking of movement in Chapter 4), I also had the opportunity to work as part of an international team, presenting my research plans and papers to faculty and researchers at the Center for Early Intervention and Family Studies. This experience not only provided access to valuable data, but also a wealth of experience, for which I am grateful to colleagues at the University of Copenhagen.

Chapter 3 - The Temporal Architecture of Narrative in Adult-Infant Interaction Through Infancy

Abstract

Human interaction is structured in time by acts that initiate an engagement and develop it over cycles of affective expression with shared interest. A temporal structure underpins each act of participation in shared time, and is fundamental to the creation of social meaning. Narrative is a critical structure for human learning, culture and intersubjectivity, and has been shown to be present in the earliest proto-conversations of infancy. However, it's precise temporal structure and development in the first year of life has not been mapped. We measured the temporal structure of narratives in dyadic interactions between mothers and their infants in a longitudinal cohort of 18 mother-infant dyads when the infants were 4, 7 and 10 months old. Video microanalysis of infant gaze, affect, engagement duration and narrative phase were recorded. Counter-intuitively, we found the durations of continuous mother-infant engagement decreased as infants grew older. Commensurately, dyads with older infants engaged in more interactions with greater narrative development that reached a climatic peak of arousal and resolution. These resulted in increased durations of positive infant affect. Altogether, these results define a changing temporal structure of interaction in the first year of life that increasingly parses interaction into rich narrative engagements with a defined 4-part structure of introduction, development, climax, and conclusion. These results support improved understanding of pre-linguistic narrative, its development in the first year of life and its close relationship with infant emotional states. Further, this study affords a new baseline for understanding the effective of disturbances to development, including for example, postnatal maternal depression.

1.0 Introduction

1.1 Narrative

From infancy, human beings are exposed to narratively structured content. We generally understand these narratives to be within the stories and songs we share with our children; stories and songs that grow in complexity and meaning as a child develops and they are able to more readily access the intricate narrative structures of their culture (often in the form of the temporal arts - film, literature, dance and opera, to name a few). However, infants experience and embody narrative structure long before they are exposed to language imbued storytelling.

Bruner (1990) proposed narrative to be the organising life principle of human cognition and the core structure underlying communication. He understood this narrative to be composed of the typical four-part structure common to narrative storytelling (introduction, development, climax, and resolution – see Figure 1.2 [Chapter 1]), and as with traditional storytelling, the narrative is characterised by cyclical patterns of increasing intensity, arousal and interest (Delafield-Butt & Trevarthen, 2015). Within an engagement between two individuals each narrative phase has distinctive characteristics. The introduction phase initiates an engagement and establishes shared attention between two persons. In the development phase the narrative builds with rising levels of energy, arousal and intensity as bodily gesture and vocalisation are exchanged in reciprocal acts until a crescendo is reached, and the narrative enters its climax phase. After this point a narrative will enter its resolution as the energy of the engagement subsides, the memory of the exchange is consolidated, and the participants prepare to renew their shared focus or alternatively disengage from one another (Delafield-Butt & Trevarthen, 2015; Gratier, 2003; Gratier & Trevarthen, 2008; Malloch, 1999; Malloch & Trevarthen, 2009; McGowan & Delafield-Butt, 2022). Infants experience and embody these narratives in their interactions with caregivers and engagements with their environment. These interactions take place from the earliest stages of infancy when intersubjective events, created by an infant and attuned adult, allow subjective experiences to become entwined.

These interactions form "proto-conversations" (Trevarthen, 1979) involving multi-modal exchanges of reciprocal movement and vocalisation, which generate rising levels of intensity and interest whilst creating shared affective meaning within a narratively structured framework (Gratier & Trevarthen, 2008). Initially, the narrative structures that framed these interactions were described as "proto-narratives" (Stern, 1999) that over time become permeated by language to form linguistic engagements. These grow in complexity and intricacy through childhood building on the foundation of regularly structured patterns of emotions and intentions (Delafield-Butt & Trevarthen, 2013).

Indeed, for many years the dominant view in narrative theory and developmental psychology was that narrative structured engagement needed language and an abstract rationale form of intelligence to be brought into being (Hutto, 2007) – see Chapter 1 for a more detailed discussion of this area. However, as discussed, recent work has demonstrated the abilities of even newborn infants to partake in the creation of fully-fledged narrative through gesture, movement and vocalisation (thus demonstrating that a narrative interaction does not require language for formation). Dautenhahn (2002) proposes pre-verbal narratives to be foundation of consciousness in a purposeful social existence, whilst McGowan & Delafield-Butt (2023) suggest that all narratives are rooted in movement. Evidence of narrative in the earliest interactions between adults and infants lends weight to the notion of narrative not necessarily being the creation of intersubjective experience but an innate foundational structure that forms the basis of learning patterns, and underpins the formation of embodied cultural practices and co-operative activity (Cobley, 2013; Delafield-butt &

Adie, 2016; Delafield-Butt & Trevarthen, 2015; Gratier & Trevarthen, 2008; Negayama et al., 2015).

1.2 Dyadic Interaction

Dyadic interactions in infancy are an infant's first experience of face-to-face dialogue. During these engagements there is an attention to the emotions, intentions, interests and arousal of another individual (without any reference to external objects), and a state of primary intersubjectivity is created (Trevarthen, 1979, 2012). Such interactions can vary dramatically in duration with some lasting only several seconds whilst others might last 30 seconds or more and contain prolonged and consistent engagement. Such interactions play a key function in the regulation of arousal and affect in both infant and adult, with regulation of the autonomic system in both parties becoming closely aligned to their expressive social engagement systems (Porges & Furman, 2011). These physiological systems of metabolic and autonomic regulation support the psychological interests and intentions of the infant, helping to structure their capacity for activity and learning. The coupling of these psycho-biological processes exemplifies the embodied nature of narrative engagements and underpins early childhood learning of the expected regulations in a family and culture (Gratier, 2003; Gratier & Apter-Danon, 2008).

Infants are born able to participate in these dyadic interactions and begin building narratively structured engagements with their caregivers from birth. They are equipped with a repertory of behaviours that can be deployed to facilitate attunement and co-regulation. This includes a range of emotional expressions as well as oral movements (involving the lips, mouth and tongue) and vocalisation (Trevarthen, 1979). The infant's ability to attune to caregivers is enhanced as they grow through the development of capabilities including following another person’s gaze, and perceiving movements as purposeful and goal-directed.

This ultimately facilitates the creation of intersubjective states of increasing complexity, which are direct product of dyadic engagement. The continuing development of communicative skills, alongside broader socio-cognitive development also facilitates increasingly complex narratives involving a wider range of communicative modalities. At around 9 months of age this will allow for the creation of secondary intersubjectivity (where adult and infant can bring into their intersubjective experience attention to objects and interests beyond the dyad) (Trevarthen & Hubley, 1978), and then ultimately language imbued conversation. Despite this trajectory of increasing complexity, and the role of narrative in regulating infant affective states and underpinning intersubjectivity, the mapping of narrative engagements through infancy has yet to be undertaken.

Understanding the role of narrative in dyadic regulation through infancy has important consequences for wider adaptive child development that extend beyond an immediate emotional benefit to the child. For example, poor regulatory capacities have been linked with academic performance, psychopathology, social competency, and health risk behaviour in later life (Calkins & Hill, 2007; Moffitt et al., 2011; Zeman et al., 2006). Despite mechanisms that can be deployed by the infant to aid in self-regulation (such as averting gaze during an interaction when intensity and arousal reach an uncomfortable level [Field, 1981]) they remain reliant on adult input and feedback to support affective regulation (Calkins & Hill, 2007; E. Tronick & Beeghly, 2011). This dyadic regulation also supports the development of more complex and effective self-regulatory procedures that can be deployed in later childhood and adult life (Beebe & Lachmann, 1998; Feldman et al., 1999; Tronick & Beeghly, 2011).

Narratives form a fundamental architecture within human interaction throughout life. They, and the type of meaning making they constitute, have their origins in the very earliest stages of human development and it is within a narrative framework that we develop our understanding of object properties, as well as the motivations and intentions of others. Narrative provides a temporal framework to our scientific and broader understanding of the world, and forms a fundamental architecture to human intelligence (Bruner, 1990). It is especially essential to early dyadic interaction when the first examples of shared meaning making occur. These narratively structured dyadic engagements are essential for healthy communicative and socio-cognitive development (Delafield-Butt et al., 2020). As such, understanding the structure of such engagements, how they develop through infancy and how they impact infant affect is an important point of enquiry within developmental psychology, and has the potential to influence areas including adult-infant entrainment, neonatal imitation, but also mother-infant health and early childhood education.

1.3 The Current Study

This study will aim to a) explore the changing temporal nature of dyadic narrative interactions through the first year of life, and b) examine the impact of these changes on the duration positive affect experienced within an interaction. This will be achieved through an analysis of mother-infant naturalistic interactions at three age points in a child's infancy.

Although past research has identified narrative in the earliest proto-conversations between adults and infants (as described more extensively in Chapter 1), no work has yet attempted to explore its development across infancy and examine how this impacts infant affect during dyadic interactions. As such, this study will help to not only further develop our understanding of narrative in infancy but also address the narrative gap identified in Chapter 1 (the absence of a narrative dimension from much current adult-infant interaction research). The study will expand our understanding of the link between infant affect and narrative, and will form the foundation of future work investigating how this impacts broader social and emotional development. We believe the duration of narrative engagements will increase through infancy as infant cognition develops, enabling longer and more complex interactions. We expect this to be true not only for overall engagement duration but also when engagements are subsetted according to the narrative phase at which they terminated. We expect engagements involving infants aged 7 or 10 months to be more likely to reach the climax or resolution phases of the narrative cycle, whilst a greater proportion of engagements would be expected to terminate in the introduction and development phase in early infancy. With regard to positive affect duration, we expect this to be a function of narrative completion, meaning increased positive affect resulting from engagements that complete full narrative cycles.

2.0 Materials and Methods

2.1 Participants

The data analysed in this study were collected at the University of Copenhagen Babylab and held at the Center for Early Intervention and Family Studies. It is longitudinal in nature and was originally collected in order to explore mother-infant interaction and relationships through infancy. Although the original dataset included both healthy mothers and those diagnosed with postnatal depression, the data for this study is comprised only of mother-infant dyads where the mother did not present as suffering postnatal depression (mothers were asked to complete the Edinburgh Postnatal Depression Scale 6-9 weeks after giving birth with those scoring ≥ 10 being excluded). Physically healthy first-time mothers were recruited to the project during pregnancy, with the following exclusion criteria applied: premature birth of the infant, non-singleton pregnancy, mental or physical disabilities in the infant following birth, drug or alcohol abuse in the mother, psychotic psychopathology in the mother, the development of a severe neurological condition in the mother in the first year following birth, and the development of a severe somatic health condition in the mother in the first year following birth. Mother and infant could not be included in the cohort if the mother was a student at the University of Copenhagen Department of Psychology (the location of the study), if they lived too far from the testing location, or if pre-birth interviews administering the Present State Examination (Wing et al., 1974), as well as assessments of attachment status and personality pathology, were not completed. The study was approved by the Ethical Review Board at the University of Copenhagen, and all mothers gave written informed consent prior to participation.

Recruitment resulted in a sample of 60 non-depressed mother infant dyads. In the current study the application of additional exclusion criteria based on the useability of audio-video recordings and available auxiliary data created the final sample for analysis (see Figure 3.1 for a breakdown of the final exclusion process with further details provided at relevant points within section 2.2). This was to ensure the integrity of the video footage analysed, and that the included engagements were a true representation of mother-infant interactions and not overly impacted by missing data. The result was a longitudinal cohort of 18 mother-infant dyads. The mean age of mothers from these dyads was 31.2 (range 23 to 43) with 8 of the infants being boys. All mothers were Caucasian, spoke Danish as there first language, and had a mean level of education of 15.7 years including primary school (range 12 to 16 years).

2.2 Procedure

Mother-infant dyads attended an observation room at the University of Copenhagen Babylab when the infants were aged 4 months, 7 months, and 10 months. Each visit was timed with the infant's eating and sleeping routines to ensure they were in an alert state. Mother and infant sat opposite one another in a standard face-to-face set up with the infant in an infant seat and the mother on a chair (Tronick & Cohn, 1989), beginning their interaction approximately 50cm apart (Figure 3.2 illustrates the precise nature of the experimental set up). Mothers were instructed to interact with their infants in a natural fashion as they normally would. No toys were provided and mothers were requested to avoid using a pacifier during the engagement. The interaction was filmed using Panasonic NV-GS300 from two perspectives at a rate of 25 frames per second. The precise set up of these two perspectives varied between two possible options: Option one utilised one camera recording a frontal view of the infant and a second recording a lateral view of the dyad, whilst option two utilised one camera behind the left shoulder of mother (facing the infant), and a second camera positioned behind the right shoulder of the mother (also facing the infant). For a dyad to be included in this study at least one recording from behind the mother needed to be in the data archive (dyads with only lateral recordings were excluded and were counted as recording not present). Further to this, two mirrors placed behind the infant allowed cameras to capture a frontal view of the mother (dyads where the mothers or infant's faces were obscured for more than 20% of the recording were excluded – see Figure 3.1). The interaction lasted 10 minutes in total but was terminated early if the infant cried for more than 30 seconds continuously or if the mother felt the infant to be too unsettled.

Our analysis of these interactions was conducted during the Covid-19 pandemic, requiring multiple visits to the University of Copenhagen to access the video footage (ethical permissions acquired when the video footage was collected prevented it from being shared outside of the European Union). This entailed a lengthy approval procedure before permission was granted by the University of Strathclyde and the Scottish Graduate School for Social Science for overseas fieldwork to be conducted. This process was further complicated through fluctuating restrictions on international travel. In total, 3 months was spent on site at the University of Copenhagen conducting our analysis.

2.2.1 Gaze Coding

Infant gaze was coded on a frame-by-frame basis over the first 5 minutes of each interaction using the video annotation software ELAN (Lausberg & Sloetjes, 2009). Gaze was coded into 3 categories: gaze on, gaze off, and uncodable. For gaze on to be coded the infant needed to have visually attended to the mother’s facial area for a minimum period of two frames (80ms). If this criterion was not met then gaze off was recorded (this approach was adapted from Beebe et al., 2008). The third criteria, uncodable, was used where infant gaze could not be ascertained (for example if the infant's face was blocked from view by the mother). In such cases the code immediately preceding the uncodable period was recorded. If during more than 20% of an interaction it was not possible to ascertain infant gaze, the dyad was excluded from analysis. Blinking was not coded as gaze off, but if an infant's eyes remained closed for more than 280ms gaze, off was recorded. 20% of dyadic engagement was dual-coded and inter-rater reliability was calculated (time-based $\kappa = 0.83$, event-based $\kappa = 0.69$; Bakeman & Quera, 2011).

2.2.2 Affect Coding

Infant and Mother affect was coded on a second-by-second basis to assess affective displays over the course of five minutes of interaction (aligned with gaze coding). Affect was classified according to 6 categories: high positive, low positive, neutral interest, mild negative, high negative and uncodable. The protocol was an adapted criteria developed by Høskuldson & Smith-Nielsen (2012) based on a similar procedure used by Koulomzin et al. (2002) and Beebe et al. (2010). Behavioural Observation Research Interactive (BORIS) software (Friard & Gamba, 2016) was used to annotate video recordings. Recordings where the infant displayed high levels of fussiness and minimal engagement with the mother and no positive affect were removed so as to mitigate outlier impact. Finally, 20% of recordings were dual-coded and inter-rater reliability was calculated (time-based $\kappa = 0.79$, event-based $\kappa = 0.81$; Bakeman & Quera, 2011). For the purposes of our analysis, the two positive affect categories were collapsed into one.

2.2.3 Narrative Coding

Narrative phase (introduction, development, climax, or resolution) was also coded on a second-by-second basis over the same five minutes of recorded interaction and temporally aligned with infant affect and gaze coding. The coding protocol for narrative structure was developed by Delafield-Butt et al (2020) and based on non-verbal narrative paradigms in neonates (Delafield-Butt & Trevarthen, 2015; Malloch & Trevarthen, 2009) and children (Delafield-butt & Adie, 2016). The procedure first identified periods of active engagement between mother and infant. For our purposes an engagement was considered to have begun when infant gaze was coded as 'gaze on' for a period ≥ 2 seconds, and terminated when infant gaze was coded as 'gaze off' ≥ 2 seconds. Gaze was utilised to delineate engagement as observing a dyadic partner's face indicates the potential for interaction and thus forms the basis of a face-to-face encounter (Beebe et al., 2010). Both infants and mothers are attuned to their dyadic partner’s gaze direction, with the infant's head movements and visual system nearly reaching adult proficiency by 3 to 4 months (Stern, 1974). Moreover, gaze control and head tilting enable the infant to achieve “subtle instant-by-instant regulation of contact” (Stern, 1971, p. 502). A gaze period lasting ≥ 2 seconds was selected as our engagement boundary. This provided a prolonged period of engagement that excluded fleeting or glancing gaze exchanges during the initiation of an interaction, but was long enough to include infant attention to other aspects of the dyadic partner (e.g. hands) during the engagement. Infant gaze was the focus of our analysis as mothers have been instructed to interact with their infants resulting in mother gaze being predominantly ‘gaze on’ (pilot coding of a subset of interactions showed mother gaze was ‘on’ for ≥ 92% of frames). Additionally coding for mother gaze was not possible to the same accuracy as infant gaze due to the positioning of cameras behind the mother. Interaction periods were thus defined as periods of adult-infant interaction initiated with a period of infant gaze at their mother lasting a minimum of 2 seconds and terminated when gaze is broken for a period longer than 2 seconds.

These engagements were then labelled according to the most developed narrative phase reached (for example, if a period of engagement between mother and infant followed a narrative form and reached a climatic point of energy and intensity whilst not proceeding to a joint resolution, the engagement would be labelled 'climax') as identified by the level of intensity and arousal displayed during the engagement. More precisely, and building on the protocol developed by Delafield-Butt et al (2020), the introduction phase was coded when an initial action from one member of the dyad was performed that could be understood (by human rater or dyadic partner) to have communicative intent was received and reciprocated by the partner. This phase could contain multiple reciprocal actions but with the absence of increasing levels of arousal and intensity. However, the development phase was coded when gradually increasing levels of arousal and intensity were expressed through increased movement, vocalisation and/or expression (or increasing intensity within such mediums, for example through increasing volume or pitch in acoustic expression, or increasing acceleration or velocity in the kinematics of expressive action) in reciprocal exchange and building on the introduction phase. The climax phase was coded when an energetic apex, following the period of building intensity in the development phase, was witnessed (again, expressed across any of the modalities associated with infant or adult expression). Finally, the resolution phase was coded according to a decrease in energy and intensity following a climatic peak, in which a resulting calm was experienced and shared by the dyad. It is also important to acknowledge in such a coding scheme, that as narrative beings, we expect human raters to be innately well attuned to identifying narrative in human engagements. It is inherent within our definition of narrative that classifying an interaction as having reached any specific phase means that preceding phases must already have been identified as present (for example, an interaction that reached the climax phase must have already passed through the introduction and development phases). We did not code for nor attempt to identify narrative phase boundaries within a period of engagement (for example, the point at which a development phase transitions to a climax phase). This is currently a weakness in this area of research, as there has been poor inter-rater reliability when attempting to code for the boundaries between narrative phases. This is something we hope the analysis pipeline in Part 3 of this thesis will help address. Periods of uncodable gaze lasting longer than 2 seconds were not taken as a break in engagement unless other factors supported a disintegration of the interaction (*for e.g.* the infant's head was directed away from mother, or the mother disengaged from infant). As with gaze and affect coding, 20% of recordings were dual-coded and inter-rater reliability was calculated (time-based κ = 0.96, event-based κ = 0.68; Bakeman & Quera, 2011).

2.2.4 Analysis and Generalised Linear and Linear Mixed Models

We were interested in the extent to which the narrative temporal structure of an engagement changed and developed through infancy as measured at 3, 7, and 10 months of age, and whether infant affect during an engagement was related to the narrative phase reached during interaction. Generalised linear models (GLM) and generalised linear mixed models (GLMM) were conducted utilising the log link (duration data, family = poisson) or logit link (binomial data, family = binomial) and the lme4 package glm and lmer functions for GLMs and GLMMs using R (R Core Team, 2023).

Each model included a by-participant slope for variables that varied within participant in an effort to keep random effects structures 'maximal' where possible (Barr et al., 2013). In cases of non-convergence or singular fit within the maximal model, random slopes followed by random intercepts were removed until a convergent, non-singular model was created. The final structure utilised in our analysis is outlined in each section below.

3.0 Results

3.1 Duration

3.1.1 Infant-Mother Engagement Durations Reduce with Age

To understand engagement duration as a function of age, we built a linear mixed-effects model. This model had engagement duration as our dependent variable, infant age as a fixed effect, a random effect for participant and included a total of 467 observations. These results are presented in Figure 3.3. Interactions between 4 month old infants and their mothers (M=11.90[1.30]) were significantly longer than when the infants were 7 months (M=9.16[0.88]), (β=-2.65, SE=1.28, t(462.68)=-2.06, **p=.004**) and 10 months of age (M=7.03[0.45]), (β=-5.04, SE=1.17, t(461.81)=-4.33, **p<.001**). We next compared length of engagements when infants were 7 and 10 months old (with a total of 335 observations included in model). Infant-mother engagements were significantly longer when infants were 7 months old (β=-2.25, SE=0.89, t(332.16)=-2.50, **p=.013**).

3.1.2 Engagement Durations Reduce with Age, Despite their Narrative Complexity

Having established the impact of age upon engagement duration, we next sought to understand the effect of age on engagement duration when interactions were grouped according to the narrative phase the engagement reached (i.e. if the interactions within an engagement reached the climax phase, it would be grouped accordingly). As such, we built a series of linear mixed effects models with each focusing on a different group (introduction phase group –182 observations included; development phase group – 166 observations included; climax phase group – 78 observations included; resolution phase group – 38 observations included). Each model followed the same structure, with a dependent variable of interaction duration, a fixed effect of infant age, and a random effect of participant. These results are presented in Figure 3.4.

Interactions between 4 month old infants and their mothers (M=3.79[0.22]) that only reached the introduction phase were significantly longer than when the infants were 10 months of age (M=2.95[0.12]), (β=-0.81, SE=0.24, t(178.10)=-3.40, **p<.001**) but there was no significant difference when infants were 7 months old (M=3.46[.21]), (β=0.29, SE=0.26, t(175.91)=-1.10, p=.274). However, in the development phase (10m(M=5.35[.27]) v 4m(M=10.6[1.03]): β=-5.27, SE=0.81, t(159.52)=-6.52, **p<.001**; 7m(M=6.67[.38]) v 4m: β=-3.83, SE=0.89, t(159.57)=-4.32, **p<.001**), climax phase (10m (M=11[0.68]) v 4m (M=28.7[4.98]): β=18.69, SE=3.16, t(73.85)=-5.91, **p<.001**; 7m (M=16[0.95]) v 4m: β=12.93, SE=3.41, t(73.95)=3.79, **p<.001**) and resolution phase (10m (M=20.4[1.88]) v 4m (M=47.3[0.89]): β=-27.22, SE=6.25, t(35.00)=-4.35, **p<.001**; 7m (M=31.2[6.51]) v 4m: β=-17.07, SE=7.08, t(34.41)=-2.41, **p=.021**) both the 7 month age group and 10 month age group had significantly shorter interactions than the 4 month age group.

Each model was also subsetted to exclude engagements when infants were 4 months old, and compare the length of interactions when infants were 7 and 10 months old (introduction phase group –124 observations included; development phase group – 118 observations included; climax phase group – 59 observations included; resolution phase group – 32 observations included). In the introduction phase interactions were significantly longer when infants were 7 months old (β=0.52, SE=0.22, t(121.99)=2.32, **p=.022**), a pattern repeated for the development phase (β=-1.43, SE=0.45, t(114.41)=-3.16, **p=.002**) and climax phase (β=-4.80, SE=1.36, t(10.22)=-3.52, **p=.005**). The difference between interactions was approaching significance between age groups in the resolution phase (β=10.13, SE=4.98, t(26.02)=-2.03, p=.052). Altogether, these results demonstrate a pattern of decreasing engagement duration with age, despite how complex those narratives might be.

3.2 Narrative Phase Reached within an Engagement

3.2.1 Distribution of Completed Narrative Cycles by Age

To begin our analysis, we first wanted to ascertain whether narrative cycles that reached the resolution phase (completed narratives) were evenly distributed across all age groups. A chi-square goodness of fit was performed and completed narratives were not found to be evenly distributed across age groups (χ^2 [2] = 12.67; **p = .002**).

3.2.2 Narrative Complexity Increases as Infants Age

To explore this further we investigated whether the narrative phase ultimately reached during an interaction was itself a function of infant age. To do this we again designed a series of general linear mixed effect models that considered this with regard to each narrative phase separately. The narrative phase reached variable was dummy coded so that it consisted of two levels in each model - the narrative phase that was the focus of the model and then all other narrative phases collapsed. Each model had the narrative phase reached as a dependent variable, infant age group and mean centred engagement duration as fixed effects, participant ID as a random effect and contained a total of 467 observations.

As shown in Figure 3.5 there was a significant difference in the number of interactions that only reached the introduction phase between age groups (see table 1 for distribution of engagements according to narrative phase reached) - there were significantly more of these interactions when the infants were 4 months of age than when infants were 7 months of age (β=-0.78, SE=0.37, z=-2.13, **p=.034**), and significantly less when infants were 4 months old than when infants were 10 months (β=-1.13, SE=0.34, z=-3.56, **p<.001**). These results were reflected for the number of engagements that reached the development phase (10m v 4m: β=1.26, SE=0.35, z=3.56, **p<.001**; 7m v 4m: β=0.83, SE=0.38, z=2.17, **p=.030**).

However, in the climax phase (10m v 4m: β=2.97, SE=0.58, z=5.11, **p<.001**; 7m v 4m: β=2.49, SE=0.59, z=4.24, **p<.001**) and resolution phase (10m v 4m: β=4.58, SE=1.16, z=3.94, **p<.001**; 7m v 4m: β=3.64, SE=1.08, z=3.36, **p<.001**) there were significantly more engagements that reached these phases when infants were 7 months and 10 months of age than when infants were 4 months old.

As in previous analyses, each model was then subsetted to exclude engagements when infants were 4 months old, and so compare the number of engagements that reached the introduction phase, development phase, climax phase and resolution phase when infants were 7 and 10 months old (see Figure 3.5 and Table 1). Each model contained a total of 335 observations. Significantly more interactions reached the resolution phase for infants at 10 months of age than 7 months of age (β=1.61, SE=0.65, z=2.46, **p=.014**). However, there was no significant difference between age groups concerning the number of interactions that reached the introduction (β=-0.40, SE=0.32, z=-1.24, p=.216), development (β=0.48, SE=0.33, z=1.43, p=.152) or (β=0.59, SE=0.39, z=1.52, p=.129) climax phases.

3.3 Narrative and Affect

3.3.1 Complex Narratives Result in Increased Duration of Infant Positive Affect

To explore the relationship between positive affect and narrative we first examined the duration of infant positive affect during an engagement as a function of the furthest narrative phase reached in an interaction. We built a linear mixed effects model with mean centred positive infant affect duration as our dependent variable; narrative phase reached, mean centred interaction duration, and mean centred positive mother affect duration as fixed effects; and participant, age group, Bailey cognitive composite score, Bailey language composite score, and Bailey motor composite score as random effects. Interaction duration was included as a fixed effect to control for the impact of longer engagements upon positive affect duration. The model had a total of 442 observations. As shown in Figure 3.6, an engagement contained significantly different levels of positive infant affect depending on the furthest narrative phase reached during the interaction. There were significant differences in this regard between all narrative phases with the exception of the introduction (M=1.07[.12]) and development (M=1.93[0.21]) phases (β=0.46, SE=0.39, t(419)=1.16, p=.652).

Engagements that reached the climax phase (M=5.11[0.63]) resulted in significantly longer durations of positive infant affect than those that reached the development phase (β=1.74, SE=0.55, t(423)=3.22, **p=.008**) and those that only reached the introduction phase (β=2.2, SE=0.59, t(428)=3.72**, p=.001**). Engagements that completed a full narrative cycle - reaching the resolution phase (M=8.14[1.54]) - resulted in significantly longer durations of positive infant affect than engagements that reached the climax phase (β=-2.04, SE=0.77, t(416)=-2.65, p=.041), development phase (β=-3.78, SE=0.83, t(422)=-4.57, **p<.001**) or the introduction phase (β=-4.24, SE=0.89, t(423)=4.76, **p<.001**).

3.3.1 Complete Narrative Cycles Result in Longer Durations of Positive Mother Affect

Our final analysis explored the relationship between mother positive affect and the furthest narrative phase reached in an interaction. We built a linear mixed effects model that mirrored the structure of the model utilised to analyse infant positive affect and narrative phase reached. Our model had mean centred positive mother affect duration as our dependent variable; narrative phase reached, mean centred interaction duration and mean centred positive infant affect duration as fixed effects; and participant, age group, Bailey cognitive composite score, Bailey language composite score, and Bailey motor composite score as random effects.

Interaction duration was again included as a fixed effect to control for the impact of longer engagements upon positive affect duration and there were 442 observations. There was no significant difference in the duration of positive mother affect (see Figure 3.7) regardless of the narrative phased reached in an interaction (Introduction (M=2.89[0.12]) v Development (M=6.22[0.34]): β=-0.06, SE=0.25, t(424)=-0.24, p=.995; Introduction v Climax (M=14.24[1.5]): β=0.03, SE=0.38, t(388)=0.09, p=.999; Development v Climax: β=0.09, SE=0.35, t(417)=0.26, p=.994) except with regard to the resolution phase, which resulted in significantly longer durations of positive mother affect (Introduction v Resolution (M=22.38[2.49]): β=2.16, SE=0.58, t(316)=3.74, **p=.001**; Development v Resolution: β=2.1, SE=0.53, t(350)=3.93, **p<.001**; Climax v Resolution: β=2.19, SE=0.48, t(412)=4.54, **p<.001**).

4.0 Discussion

The temporal architecture within which infants and their caregivers interact structures early intersubjective engagements, which themselves form the foundation upon which social and emotional development builds throughout infancy and childhood (Kokkinaki et al., 2023; Trevarthen & Hubley, 1978; Tronick, 2001). The precise nature of this structure - how it shifts and alters through development - has not as yet being fully explored. A body of literature has proposed narrative patterning as being foundational in early interactions, and the proto-conversations of early human life that prelude the onset of language (e.g. Delafield-Butt & Trevarthen, 2015; Malloch, 1999; Malloch & Trevarthen, 2009). We reason that this narrative architecture is not a static structure in human development, but rather its precise nature will change through early childhood depending on the abilities and interests of the engaged parties. To explore this, mothers and infants were recorded in naturalistic interactions at three time points (when infants were 4 months old, 7 months old, and 10 months old) that reflected the engagements they would partake in during their everyday lives.

Each interaction lasted for 10 minutes, of which the first 5 minutes was subject to further analysis.

From these recorded observations we coded and time stamped infant eye gaze (whether the infant was looking at their dyadic partner), infant and mother affective state, and the furthest narrative phase reached in interactions that took place within the 5-minute analysis period. We predicted that the amount of time spent in a positive affective state would be a function of the narrative phase reached during an engagement (more precisely, that interactions that completed more phases of a narrative cycle would generate greater duration of positive affect), and the narrative phase an engagement reached would itself be a function of the age of the infant (with older infants expected to engage in a greater number of interactions that reached the climax or resolution phase of the narrative cycle). Furthermore, we predicted infant age would also affect the overall duration of the engagements, with older infants engaging in longer interactions, a pattern we expected to be true when engagements were categorised according to the furthest narrative phase reached in the interaction.

With the exception of engagement durations, our results were broadly consistent with these predictions. Although, as predicted, the general duration of interactions was found to be a product of infant age, it was younger infants who experienced longer engagements with interaction duration decreasing with age. This pattern was also true when engagements when categorised by the narrative phase they reached, with our analysis only failing to find significant differences in the length of introduction phases between infants aged 4 and 7 months, and the resolution phase between infants aged 7 and 10 months.

When considering the relationship between the narrative phase an interaction reached and infant age, we found that the number of completed narrative cycles (those that reached the resolution phase) was not distributed evenly across age groups and in fact increased as infants aged. There were also more engagements that reached the climax phase before terminating (except between 7 months and 10 months). There were a greater number of interactions that failed to develop beyond the introduction or development phase in the 4-month age group compared to the 7-month age group, and more engagements that terminated in both of these phases for infants aged 10 months relative to those aged 4 months. Finally, we found that the duration of positive infant affect during interactions that completed a full narrative cycle was greater than in those that only reached the introduction, development or climax phases. Interactions that reached the climax phase also resulted in significantly longer positive infant affect than those that reached the introduction or development phase. For mothers, only interactions containing complete narrative cycles resulted in increased duration of positive affect.

A decrease in engagement duration as infants aged (both with regard to general engagement duration and when engagements were subsetted according to narrative phase reached) was unexpected as this initially seems to go against the grain of general cognitive and social development in the first year of life. We expected general development (both physiological and psychological) to increase infant ability to engage in longer and more complex interactions. However, our results could still be interpreted as part of the wider spectrum of development, whilst also potentially as a consequence of the methodological limitations of the experimental set up. For example, as part of the study all infants wore a one-piece baby grow and hat with optical motion tracking markers attached across the head (two markers positioned atop the infant's head, arms (shoulders, elbows and wrists), hips and legs (knees and ankles). This clothing became more disruptive in older infants, who would at times become distressed at having to wear the clothing or often become distracted by the presence of the markers (this was despite clothing being specifically chosen to camouflage the markers as much as possible). This potentially acted as a barrier to more prolonged engagements for some older infants. Another interpretation draws on the expanding world of the infant. Physical developments such as greater postural and motor control, but also psychological developments such as the onset of secondary intersubjectivity and joint attention allow the infant to actively engage with and observe their environment with increasing understanding and independence. An increased interest in their surroundings when in a novel environment could impede the recording of longer and more complex interactions involving older infants in the experimental setting. To overcome these issues a future study could collect data in the infants home using emerging motion tracking technologies (such as markerless motion tracking (e.g. Lauer et al., 2021; Nath et al., 2019) or wireless, miniaturised, wearable motion sensors (Chen et al., 2016) removing the potential impact of a novel environment on the infant's engagement with their mother.

Our results considering the narrative phase reached during an interaction and the age of the participating infant indicate the bulk of change to the temporal architecture of engagements taking place between 3 months and 7 months. These findings lend credence to primary intersubjective thesis (Moll et al., 2021) regarding the development of human unique sociality that argues human beings relate to one another in a distinctive and unique manner from early infancy. This is in contrast to shared intentionality theory which proposes a cognitive revolution towards the end of the first year of life (when joint attention develops and more complex co-operative projects are undertaken - it is at this point that the very nature of adult-infant interaction alters to include shared attention to the wider environment) represents the onset of a sociality that defines us (Moll & Tomasello, 2010; Tomasello, 2018, 2019; Tomasello et al., 2005). The significant impact primary intersubjective engagements potentially have on the development of narrative temporal architecture (as well as other developmental milestones within the first 7 months of life) further highlights the importance of early infancy in laying the foundations of future social and general development.

Our results also demonstrate the close relationship between narrative and affect, with complete narrative cycles resulting in increased positive affect. Completed narrative cycles create a sense of satisfaction and completion for the mother and infant, which can be processed in memory and support the learning of cultural norms of engagement (Delafield-Butt & Trevarthen, 2015) as well as wider learning (Delafield-Butt & Adie, 2016). The fact interactions that terminate in a climatic peak of arousal also result in increased duration of positive infant (but not mother) affect is also significant. This suggests 'partial-narratives' can also be beneficial for the social and emotional development of the infant. Given infant propensity for utilising gaze aversion to self-regulate (infants avert gaze to regulate their own state by reducing autonomic arousal [Field, 1981] whilst also signalling to their dyadic partner their disengagement [Gianino & Tronick, 1988]) it is also possible that these engagements are concluded independently by the infant in a ‘solo-resolution’ phase. The result of this in our coding would be that an engagement terminated at its climatic peak, when in fact the infant had chosen to disengage, regulate their emotional state, and consolidate their memory of the experience independently rather than bringing the narrative to completion in a joint resolution with their dyadic partner.

At a more foundational level, our findings demonstrate the presence of discernible and fully formed narratives in the first year of life, rather than proto-narratives described by Stern (1999) or pre-narratives suggested by Gallagher (in press). It therefore also supports the notion of narrative before the onset of language, and that narrative architecture is not dependent on language for creation. Furthermore, it potentially inverts the argument some have proposed regarding the relationship between language formation and narrative formation. Recent work by Di Liberto and colleagues (2023) has demonstrated the role rhythm (and therefore musicality) plays in language acquisition during early childhood, and the key role of narrative as a component of communicative musicality (Malloch, 1999) potentially infers the core role of narrative structure not only in the creation of intersubjective engagement but also language formation.

A key assumption underlying these findings is that narrative is present in the movement and actions (in the case of this study in the intersubjective space between adult and infant, but this can be broadened to include prospective movement that operates in the wider environment) of the infant. Gallagher (in-press) disputes this, suggesting that action and movement do not have an intrinsic narrative structure, but rather they appear to have a narrative temporal architecture because narratives take on the structure of the action that they are reflecting upon. This means that narratives need to be a reflective retelling of experience, and cannot be generated by the infant in their movement and actions. Furthermore, Gallagher argues that even when infant action is taken as part of a multi-modal communicative package in a primary intersubjective interaction, and shown to be part of a coherent narrative, the infant is not fully part of the narrative. On this basis, narratives are distinct from movement and interaction, even where movement is expressive of affect and intent. Instead of being the core architecture of primary intersubjectivity, narratives are not properly formed by an infant until later in development, with primary and secondary intersubjectivity being key components in their onset and emergence. Rather than being innately narrative, human infants only a possess a general narrative ability which is built upon in the first years of life through engagement with others and the environment until the emergence of full narrative capacity.

Our findings offer an initial empirical rebuttal to this interpretation of narrative in infancy. They support the presence of fully formed narrative as early as 3 months of age. That these narratives are recognisable and discernible as complete narrative cycles offers evidence against the notion of infant 'pre-narrative' ability. We instead view narrative as the foundational architecture of infant engagement with others (through primary intersubjectivity), as well in their engagement with the world. It forms an architecture within which learning can unfold, and shared affective meaning making can take place (McGowan & Delafield-Butt, 2022). In this sense, narratives do not need to be a re-presentation or re-telling of past experiences. Narrative is instead an integral architecture that brings structure to generative, expressive and reflective experience and engagement. This is as true for generative sequences of movement as it is for primary intersubjective engagements. The form and kinematics of movement when sequenced prospectively expresses forms of vitality (Stern, 2010) that demonstrate to the world the affect and intention of the individual from which an action originated. Ultimately though, our findings can not wholly exclude the possibility that action and movement appear to have a narrative temporal architecture because narratives take on the structure of an action. To explore further this point further, an experimental examination of earlier adult-infant interactions (for example in the neonatal imitation paradigm) would be beneficial, as would extending our findings beyond the experimental paradigm utilised here and into more naturalistic settings (such as the infant's home).

Whilst this distinction between pre-narrative and full narrative capacity might seem academic, it has real world consequences in how we engage with and view the learning of pre-verbal children. This does not only impact infant care and development, but extends into realm of disability and special needs provision in education. Practitioners who work with children yet to develop linguistic capacity correctly consider movement and behaviour as communicative expressions (Bunning et al., 2013). Understanding narrative as the structure which underpins this expression allows practitioners and parents an inlet into structured engagement (Delafield-Butt et al., 2020), and can help support such young individuals in regulating their internal states (McGowan and Delafield-Butt (2022) and facilitate learning (Bunning et al., 2013; Delafield-Butt & Adie, 2016). The importance of narrative in learning also extends beyond infancy and into the classroom. Narrative plays a crucial role in allowing individuals to evaluate actions and comprehend intentions (Cunliffe & Coupland, 2012) whilst organising and interpreting experiences (Kearney, 2002), thereby enhancing learning episodes. Children navigate learning tasks by breaking them into manageable segments with identifiable beginnings, middles, and ends, forming larger projects of sense-making (Egan & Ling, 2002). By successfully completing these smaller units, they achieve a sense of accomplishment, building confidence and memory of task completion (Delafield-Butt & Adie, 2016). These narrative episodes foster personal connectedness, aiding in emotional, intentional, and action-oriented ordering and orientation (Egan & Ling, 2002). As demonstrated by this study, this basic structure of this learning is present in complete form from the earliest months of infancy indicating its core role in infant as well as later childhood learning.

A likely criticism of our work, and a likely rebuttal from those who view infant narrative as pre or proto narrative, involves the use of human raters to identify narrative phases within engagements. If we accept that human adults are narrative organisms (we structure our understanding of and engagement with the world through narrative), then it could be argued that we are biased towards identifying narrative where no such structure exists. That is, we are inherently biased towards interpreting what we see in the world through a narrative structure. If such an argument were valid, it would undermine the use of human beings to identify early infant narrative. However, such an argument does not preclude the existence of infant narrative. If we are inherently narrative beings, we would expect narrative to structure human engagement with others from infancy and not necessarily be the result of cultural experience or the reflective use of language.

To fully rebut this criticism, we propose that future work should seek to automate the identification of narrative in infant engagements, therefore removing a potential confound created through the use of human raters. Such an approach could utilise an unsupervised machine learning based neural network that would enable automated detection of narrative phases that is not contingent on any human-coded ground truth. This approach could organise audio and movement data (as well as other applicable data sources) into a temporal sequence of tiles (a tile representing a package of all relevant data within a specified time window) with each tile having a predetermined temporal overlap with its previous neighbour. This tiling approach allows a network to learn the features necessary to derive levels of similarity between tiles. The clustering of these tiles according to their level of similarity makes it possible to identify any underpinning temporal framework or architecture.

For future work we would also aim to build on our approach for classifying periods of active engagement between mother and infant. In our current study we utilised infant gaze towards their mother's face as a tool for identifying periods where the infant was attentive to and engaged in interaction. However, as with broader adult-infant interaction, the initiation of engagement is a multimodal process. For example, infant attention to an adult's hands or expressive bodily expression could also be indicative of infant engagement in reciprocal interaction. In older infants, this could also include joint attention to objects in the wider environment. We would therefore aim to build on the coding protocols used in this study to implement a more holistic identification of infant social engagement.

As far as we are aware this study represents the first mapping of narrative engagements through infancy, as well as the first attempt to understand the shifting characteristics of narrative engagements and the impact they have on infant emotional states through the first year of life. Our results can be understood as part of wider development in infancy, and a progression in social engagement beginning with primary intersubjective interaction. Based on our findings we postulate that the temporal architecture of adult-infant interaction is a foundational aspect of early child development, supporting wider social and emotional development, as well as early meaning making and biobehavioural synchrony. More specifically though, our findings add weight to the body of evidence supporting the existence and importance of narrative structures in early infancy.

References

Bakeman, R., & Quera, V. (2011). Sequential analysis and observational methods for the behavioral sciences. In *Sequential analysis and observational methods for the behavioral sciences*. Cambridge University Press. https://doi.org/10.1017/CBO9781139017343

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. https://doi.org/https://doi.org/10.1016/j.jml.2012.11.001

Beebe, B., Jaffe, J., Buck, K., Chen, H., Cohen, P., Feldstein, S., & Andrews, H. (2008). Six-week postpartum maternal depressive symptoms and 4-month mother-infant self- and interactive contingency. *Infant Mental Health Journal*, 29(5), 442–471. https://doi.org/10.1002/imhj.20191

Beebe, B., Jaffe, J., Markese, S., Buck, K., Chen, H., Cohen, P., Bahrick, L., Andrews, H., & Feldstein, S. (2010). The origins of 12-month attachment: a microanalysis of 4-month mother-infant interaction. *Attachment & Human Development*, 12(1–2), 3–141. https://doi.org/10.1080/14616730903338985

Beebe, B., & Lachmann, F. M. (1998). Co-constructing inner and relational processes: Self- and mutual regulation in infant research and adult treatment. *Psychoanalytic Psychology*, 15(4), 480–516. https://doi.org/10.1037/0736-9735.15.4.480

Bruner, J. S. (1990). *Acts of meaning*. Harvard University Press.

Bunning, K., Smith, C., Kennedy, P., & Greenham, C. (2013). Examination of the communication interface between students with severe to profound and multiple intellectual disability and educational staff during structured teaching sessions. *Journal of Intellectual Disability Research*, 57(1), 39–52. https://doi.org/https://doi.org/10.1111/j.1365-2788.2011.01513.x

Calkins, S. D., & Hill, A. (2007). Caregiver Influences on Emerging Emotion Regulation: Biological and Environmental Transactions in Early Development. In *Handbook of emotion regulation*. (pp. 229–248). The Guilford Press.

Chen, H., Xue, M., Mei, Z., Bambang Oetomo, S., & Chen, W. (2016). A Review of Wearable Sensor Systems for Monitoring Body Movements of Neonates. *Sensors* (Vol. 16, Issue 12). https://doi.org/10.3390/s16122134

Cobley, P. (2013). *Narrative* (2nd ed.). New York : Routledge.

Cunliffe, A., & Coupland, C. (2012). From hero to villain to hero: Making experiences sensible through embodied narrative sensemaking. *Human Relations*, 65(1), 63–88.

Dautenhahn, K. (2002). The origins of narrative: In search of the transactional format of narratives in humans and other social animals. *International Journal of Cognition and Technology*, 1(1), 97–123. https://doi.org/10.1075/ijct.1.1.07dau

Delafield-butt, J. T., & Adie, J. (2016). of Learning : Nurture in School. *Mind, Brain and Education*, 10(2), 117–131.

Delafield-Butt, J. T., & Trevarthen, C. (2015). The ontogenesis of narrative: from moving to meaning. *Frontiers in Psychology*, 6(September), 1–16. https://doi.org/10.3389/fpsyg.2015.01157

Delafield-Butt, J. T., Zeedyk, M. S., Harder, S., Vaever, M. S., & Caldwell, P. (2020). Making Meaning Together: Embodied Narratives in a Case of Severe Autism. *Psychopathology*, 53(2), 60–73. https://doi.org/10.1159/000506648

Delafield-Butt, J., & Trevarthen, C. (2013). Theories of the development of human communication. In P. Cobley & P. J. Schulz (Eds.), *Theories and models of communication* (pp. 199–222). Mouton de Gruyter.

Di Liberto, G. M., Attaheri, A., Cantisani, G., Reilly, R. B., Ní Choisdealbha, Á., Rocha, S., Brusini, P., & Goswami, U. (2023). Emergence of the cortical encoding of phonetic features in the first year of life. *Nature Communications*, 14(1). https://doi.org/10.1038/s41467-023-43490-x

Egan, K., & Ling, M. (2002). We begin as poets: Conceptual tools and the arts in early childhood. In L. Bresler & T. M. Thompson (Eds)., *The arts in children's lives: Context, culture and curriculum* (pp. 93–100). Dordrecht, The Netherlands: Kluwer Academic Publishers.

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

Field, T. (1981). Gaze Behavior of Normal and High-Risk Infants during Early Interactions. *Journal of the American Academy of Child Psychiatry*, 20(2), 308–317. https://doi.org/https://doi.org/10.1016/S0002-7138(09)60991-2

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

Gallagher, S (in-press) Music first; narrative later: Distinguishing basic and higher-order empathies. In V. Reddy & J. Delafield-Butt (Eds)., *Trevarthen*.

Gianino, A., & Tronick, E. Z. (1988). The mutual regulation model: The infant's self and interactive regulation and coping and defensive capacities. In *Stress and coping across development*. (pp. 47–68). Lawrence Erlbaum Associates, Inc.

Gratier, M. (2003). Expressive timing and interactional synchrony between mothers and infants: Cultural similarities, cultural differences, and the immigration exprerience. *Cognitive Development*, 18(4), 533–554. https://doi.org/10.1016/j.cogdev.2003.09.009

Gratier, M., & Apter-Danon, G. (2008). The improvised musicality of belonging: Repetition and variation in mother–infant vocal interaction. In S. Malloch & C. Trevarthen (Eds.), *Communicative Musicality: Exploring the basis of human companionship* (pp. 301–327). Oxford University Press. https://doi.org/10.1093/oso/9780198566281.003.0014

Gratier, M., & Trevarthen, C. (2008). Musical narrative and motives for culture in mother-infant vocal interaction. *Journal of Consciousness Studies*, 15(10–11), 122–158.

Høskuldson, C., & Smith-Nielsen, J. (2012). Vejledning til kodning af infant facial affect (IFA) ved brug af ELAN.[Manual for coding infant facial affect (IFA) using ELAN]. Denmark: University of Copenhagen (Unpublished Manual).

Hutto, D. D. (2007). Narrative and Understanding Persons. *Royal Institute of Philosophy Supplement*, 60(August), 1–16. https://doi.org/10.1017/s135824610700001x

Kearney, R. (2002).*On stories*. London, UK: Routledge.

Kokkinaki, T., Delafield-Butt, J., Nagy, E., & Trevarthen, C. (2023). Editorial: Intersubjectivity: recent advances in theory, research, and practice. *Frontiers in Psychology*, 14. https://doi.org/10.3389/fpsyg.2023.1220161

Koulomzin, M., Beebe, B., Anderson, S., Jaffe, J., Feldstein, S., & Crown, C. (2002). Infant gaze, head, face and self-touch at 4 months differentiate secure vs. avoidant attachment at 1 year: a microanalytic approach. *Attachment & Human Development*, 4(1), 3–24. https://doi.org/10.1080/14616730210123120

Lauer, J., Zhou, M., Ye, S., Menegas, W., Nath, T., Rahman, M. M., Di Santo, V., Soberanes, D., Feng, G., Murthy, V. N., Lauder, G., Dulac, C., Mathis, M. W., & Mathis, A. (2021). Multi-animal pose estimation and tracking with DeepLabCut. *BioRxiv*, 2021.04.30.442096. https://doi.org/10.1101/2021.04.30.442096

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES--ELAN system. *Behavior Research Methods*, 41(3), 841–849. https://doi.org/10.3758/BRM.41.3.841

Malloch, S. N. (1999). Mothers and infants and communicative musicality. *Musicae*

Scientiae, 3(1_suppl), 29–57. https://doi.org/10.1177/10298649000030s104

Malloch, S., & Trevarthen, C. (2009). Musicality: Communicating the vitality and interests of life. In *Communicative musicality: Exploring the basis of human companionship.* (pp. 1–11). Oxford University Press.

McGowan, T., & Delafield-Butt, J. (2022). Narrative as co-regulation: A review of embodied narrative in infant development. *Infant Behavior and Development*, 68, 101747. https://doi.org/https://doi.org/10.1016/j.infbeh.2022.101747

McGowan, T., & Delafield-Butt, J. (2023). Neonatal Participation in Neonatal Imitation: Narrative in Newborn Dialogues. *Human Development*, 67(3), 135–153. https://doi.org/10.1159/000531311

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, 108(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108

Moll, H., Pueschel, E., Ni, Q., & Little, A. (2021). Sharing Experiences in Infancy: From Primary Intersubjectivity to Shared Intentionality. *Frontiers in Psychology*, 12(July). https://doi.org/10.3389/fpsyg.2021.667679

Moll, H., & Tomasello, M. (2010). Infant cognition. *Current Biology*, 20(20), R872–R875. https://doi.org/https://doi.org/10.1016/j.cub.2010.09.001

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, 14(7), 2152–2176. https://doi.org/10.1038/s41596-019-0176-0

Negayama, K., Delafield-Butt, J. T., Momose, K., Ishijima, K., Kawahara, N., Lux, E. J., Murphy, A., & Kaliarntas, K. (2015). Embodied intersubjective engagement in mother-infant tactile communication: A cross-cultural study of Japanese and Scottish mother-infant behaviors during infant pick-up. *Frontiers in Psychology*, 6, 1–13. https://doi.org/10.3389/fpsyg.2015.00066

Porges, S. W., & Furman, S. A. (2011). The early development of the autonomic nervous system provides a neural platform for social behaviour: A polyvagal perspective. *Infant and Child Development*, 20(1), 106–118.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Stern D. Mother and infant at play: The dyadic interaction involving facial, vocal and gaze behaviors. In: Lewis M, Rosenblum L, (Eds.), *The effect of the infant on its caregiver*. New York: Wiley; 1974. pp. 187–213.

Stern D. A microanalysis of the mother-infant interaction. *Journal of the American Academy of Child Psychiatry*. 1971;10:501–507.

Stern, D. N. (1999). Vitality contours: The temporal contour of feelings as a basic unit for constructing the infant’s social experience. In *Early social cognition: Understanding others in the first months of life*. (pp. 67–80). Lawrence Erlbaum Associates Publishers.

Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences of the United States of America*, 115(34), 8491–8498. https://doi.org/10.1073/pnas.1804761115

Tomasello, M. (2019). Becoming human: A theory of ontogeny. In *Becoming human: A theory of ontogeny*. Belknap Press of Harvard University Press. https://doi.org/10.4159/9780674988651

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *The Behavioral and Brain Sciences*, 28(5), 675–735. https://doi.org/10.1017/S0140525X05000129

Trevarthen, C. (1979). Communiaction and Cooperation in Early Infancy: A Description of primary Intersubjectivity. In M. Bullowa (Ed.), *Before Speech: The beginning of Human Communication*. Cambridge University Press.

Trevarthen, C. (2012). Embodied Human Intersubjectivity: Imaginative Agency, To Share Meaning. *Cognitive Semiotics*, 4(1), 6-56. https://doi.org/10.1515/cogsem.2012.4.1.6

Trevarthen, C., & Hubley, P. (1978). Secondary Intersubjectivity: Confidence, Confiding, and Acts of Meaning in the First Year. In J. Lock (Ed.), *Action, Gesture and Symbol* (pp. 183-229). London: Academic Press.

Tronick, E., & Beeghly, M. (2011). Infants' meaning-making and the development of mental health problems. *The American Psychologist*, 66(2), 107–119. https://doi.org/10.1037/a0021631

Tronick, E Z, & Cohn, J. F. (1989). Infant-mother face-to-face interaction: age and gender differences in coordination and the occurrence of miscoordination. *Child Development*, 60(1), 85–92.

Tronick, Edward Z. (2001). Emotional Connections and Dyadic Consciousness in Infant-Mother and Patient-Therapist Interactions: Commentary on Paper by Frank M. Lachmann. *Psychoanalytic Dialogues*, 11(2), 187–194. https://doi.org/10.1080/10481881109348606

Wing, J. K., Cooper, J. E., & Sartorius, N. (1974). Measurement and classification of psychiatric symptoms; An instruction manual for the PSE and Catego program. In *Measurement and classification of psychiatric symptoms; An instruction manual for the PSE and Catego program*. Cambridge University Press.

Zeman, J., Cassano, M., Perry-Parrish, C., & Stegall, S. (2006). Emotion regulation in children and adolescents. *Journal of developmental and behavioral pediatrics*, 27(2), 155–168.

Figures and Tables

Figure 3.1

Flow and Exclusion of Participants in Present Study

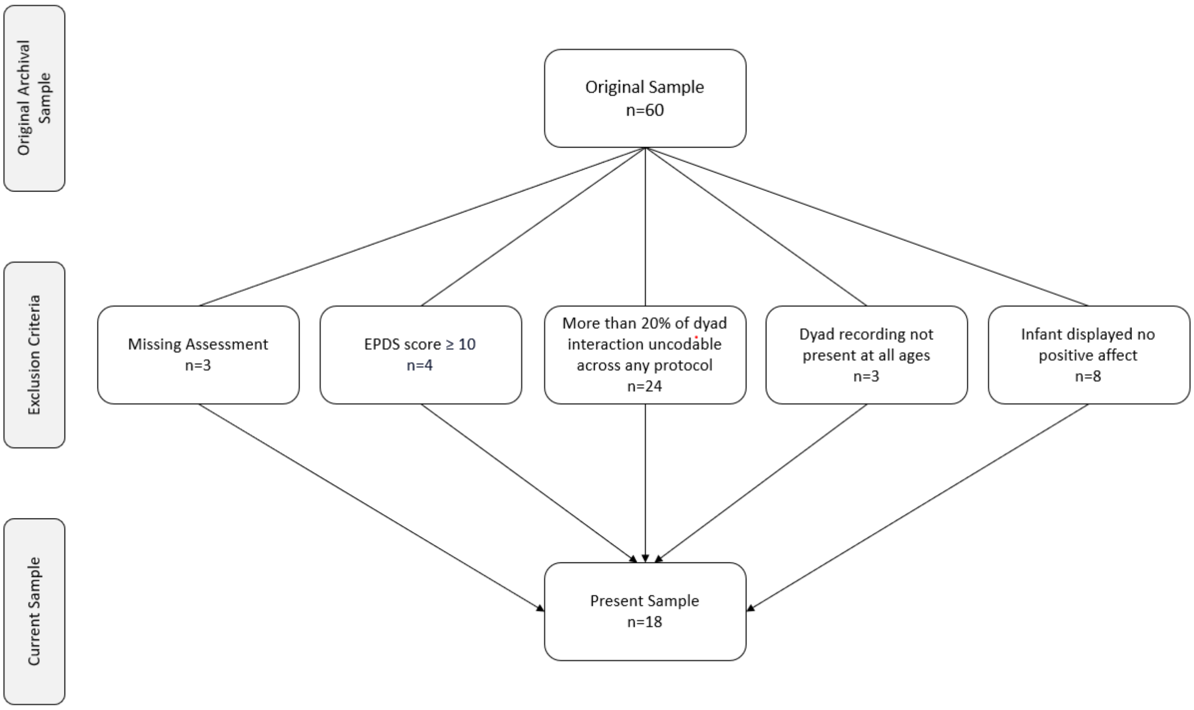
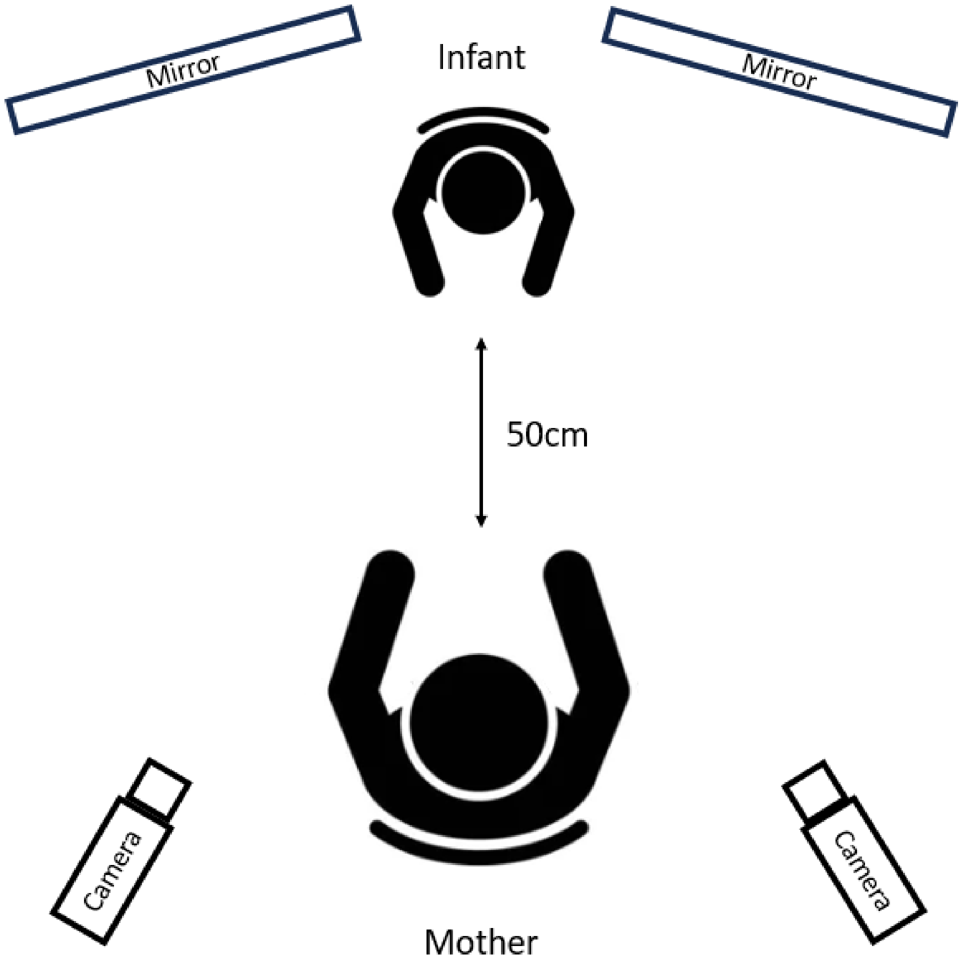


Figure 3.2

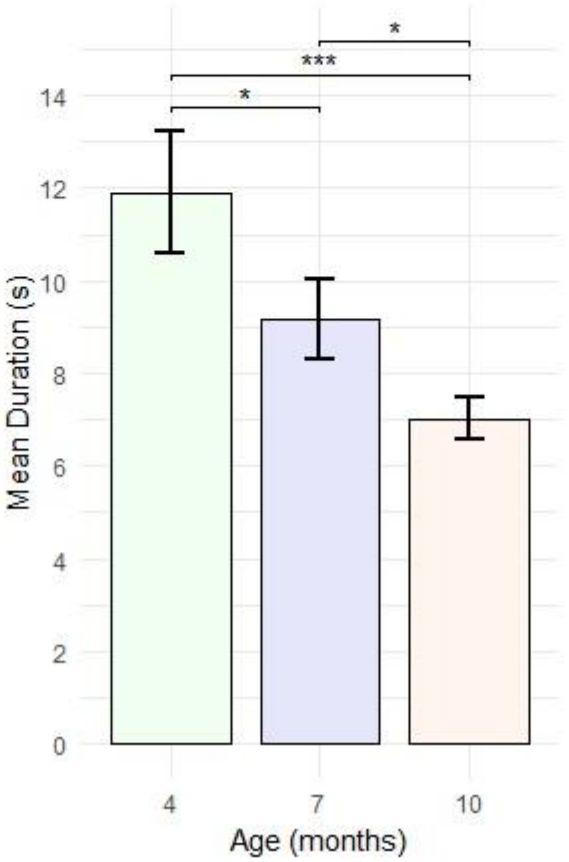
Experimental Set Up for Mother - Infant Interaction



Note. This figure illustrates the manner in which mother and infant began their period of engagement. Mother and infant were initially seated 50cm apart, with recording equipment set up behind the left and right shoulder of the mother. Mirrors were placed behind the left and right shoulder of the infant so as to enable the recording equipment to capture a frontal view of the mother during interaction. Mothers were instructed to behave with their infants as they normally would.

Figure 3.3

Mean Engagement Duration of Mother-Infant Dyads by Infant Age

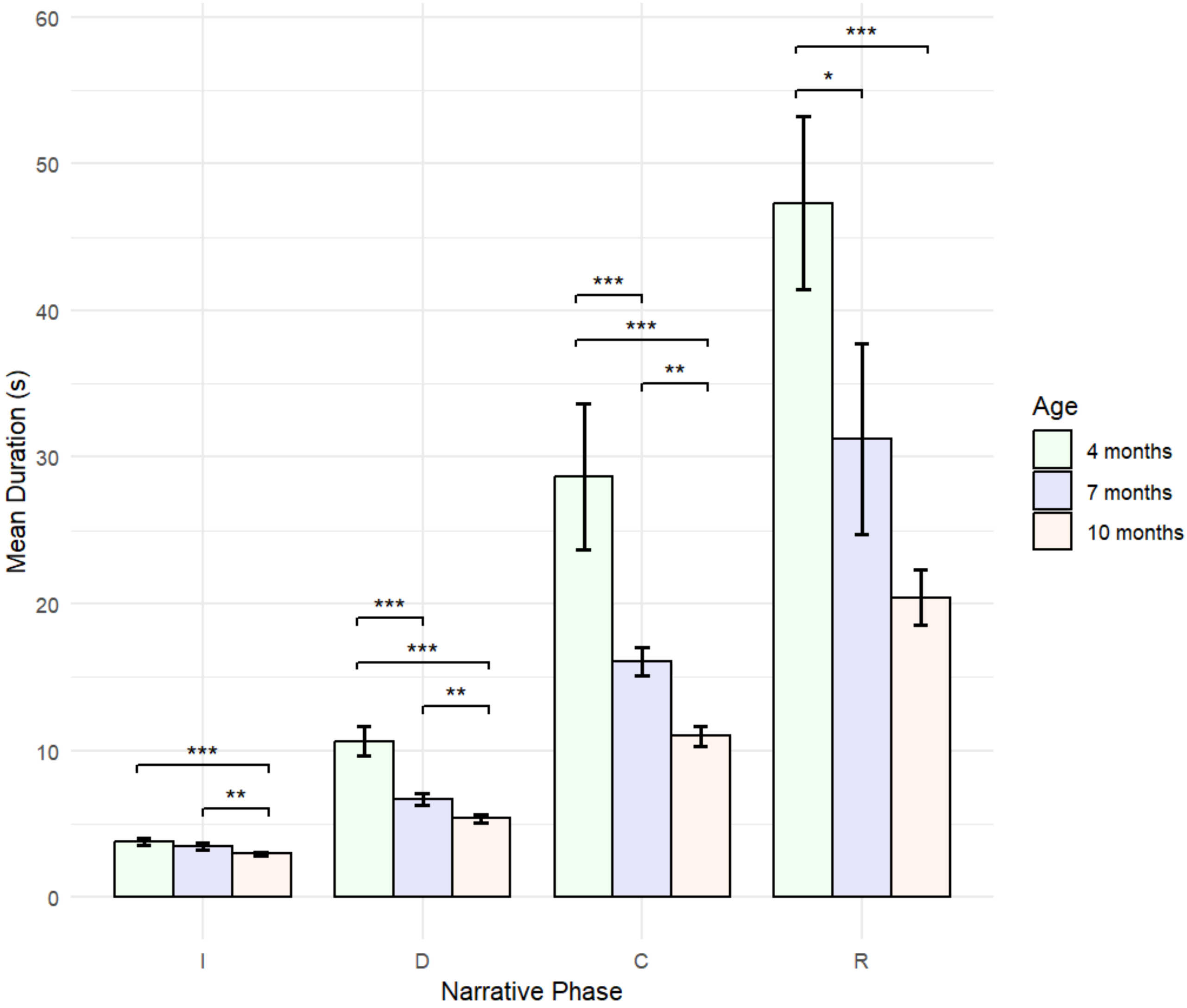


Note. Error bars represent standard error of the mean

*p < .05, ***p < .001

Figure 3.4

Mean Engagement Duration by Narrative Phase Reached and Age

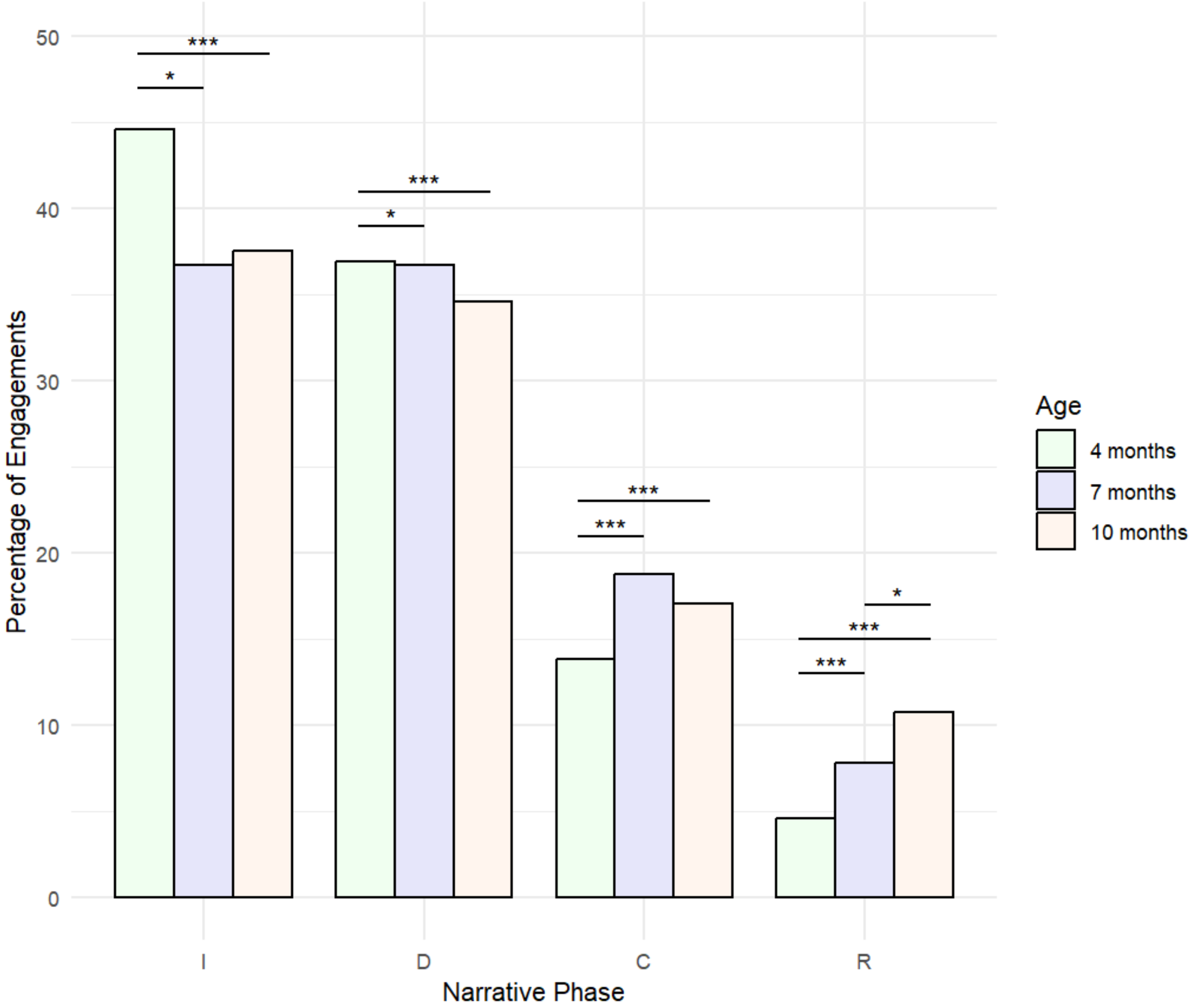


Note. Error bars represent standard error of the mean

*p < .05, **p < .01, ***p < .001

Figure 3.5

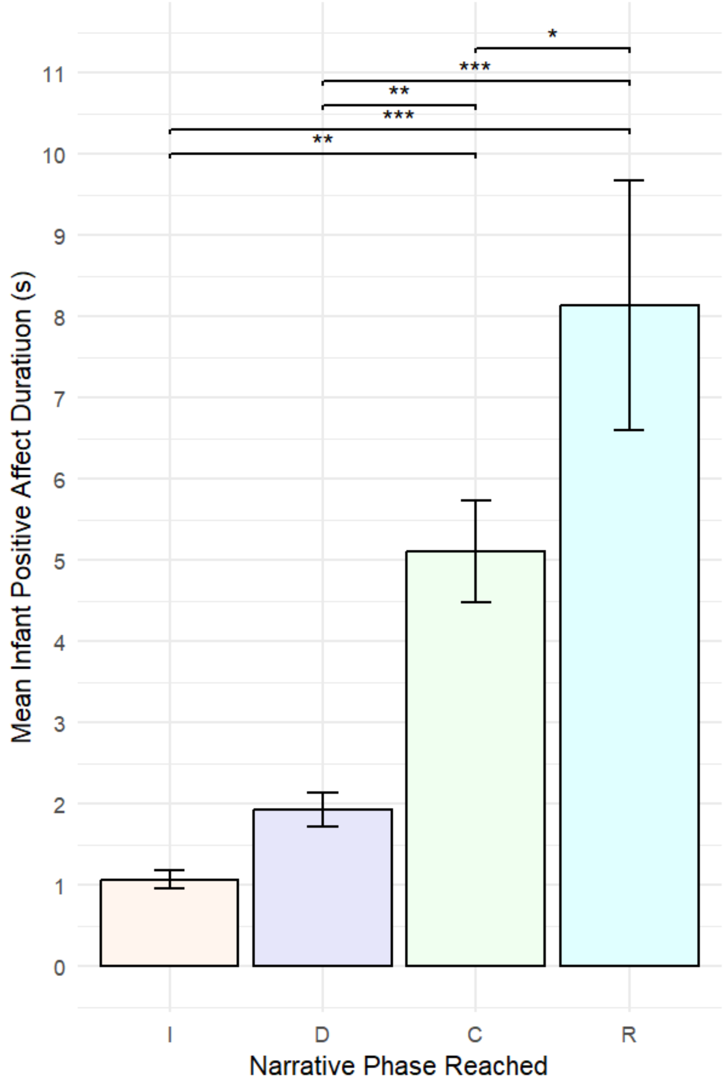
Percentage of Engagements that Reached a Given Narrative Phase



Note. *p < .05, ***p < .001

Figure 3.6

Mean Duration of Positive Infant Affect by Narrative Phase

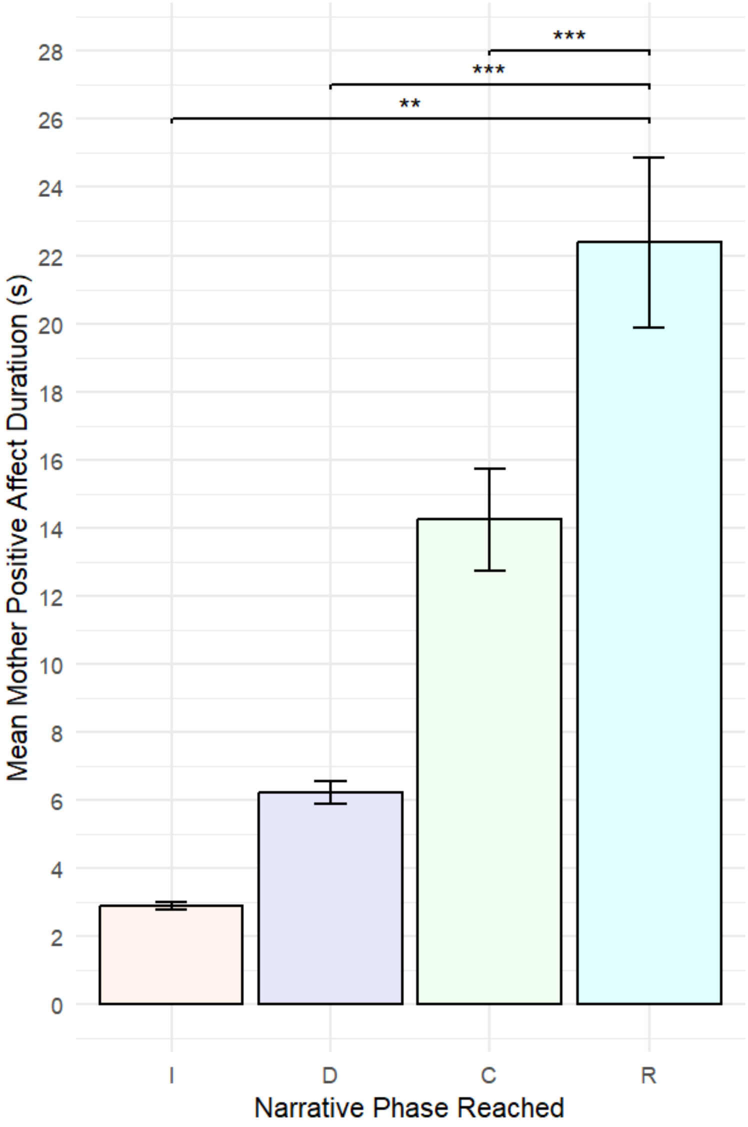


Note. Error bars represent standard error of the mean

*p < .05, **p < .01, ***p < .001

Figure 3.7

Mean Duration of Positive Mother Affect by Narrative Phase



Note. Error bars represent standard error of the mean

**p < .01, ***p < .001

Table 3.1

Percentage of Engagements that Reached a Given Narrative Phase within Each Age Group.

Age	Narrative Phase Reached	Percentage of Engagements
4	Introduction	45
4	Development	37
4	Climax	14
4	Resolution	5
7	Introduction	37
7	Development	37
7	Climax	19
7	Resolution	8
10	Introduction	38
10	Development	35
10	Climax	17
10	Resolution	11

Part 3

Prologue

My academic background could be interpreted as an interdisciplinary journey from the humanities to the social sciences, with my doctoral work has continuing in this vein. In Part 3 of this thesis, I aimed to further expand the interdisciplinary nature of my training through the application of machine learning based approaches to motion tracking and the development of a novel machine learning pipeline for the identification of narrative.

The application of deep learning based pose estimation was a direct response to the challenges presented by Covid-19. In my initial doctoral research plan, which involved data collection in Madeira (outlined in the prelude to Part 1), we had aimed to track the movements of neonates using inertial measurement units placed strategically on key body parts. Due to Covid-19 preventing in person data collection, an alternative methodological approach was needed that could be applied to secondary data composed of video recordings of mother-infant interaction. To this end, markerless motion tracking was explored (Chapter 4) and deep learning computer vision settled upon as the most effective approach to overcome the restrictions placed upon our research by the pandemic. Moreover, our work highlighted the broader utility of this approach as an exciting methodological advancement in its own right.

The development of a machine learning pipeline for the identification of narrative phases in adult-infant interaction (Chapter 5) represented the greatest departure from my past academic training. This was a significant learning curve that took me into the fields of computer science and engineering. However, the application of machine learning approaches to narrative theory is a novel approach that holds considerable potential for the future.

Abstract

A criticism often raised against proponents of narrative theory (and discussed in detail in Chapter 2) is that, as narrative creatures, human beings will be biassed towards seeing narrative in projects and interactions where no narrative exists. When studying narrative, we propose moving from a solely human identification approach to a more quantifiable methodology for the identification of narrative patterns in human (and non-human) interactions. At the heart of this approach is the automated tracking of movement. Movement is a core expressive modality within an interaction. Developments in the field of artificial intelligence and machine learning open the possibility to analyse these data without the need for human input. Here we outline the theory and pipeline for such an approach, presenting a logical automated empirical extension the work conducted in Chapter 2 while also addressing a fundamental criticism levelled against current human-coded studies of narrative in infancy.

Chapter 4 - Movement, Motion Tracking, and Narrative

1.0 Movement

“Motor coordination is the sole product of brain function" (Sperry 1952, p.297).

Movement is a basic expression of human thought, emotion and intention (Delafield-Butt and Gangopadhyay, 2013; Langer, 1967). It is closely tied with a wide range of cognitive and emotional processes, including learning, memory, decision-making, and social interaction, and as such, provides a valuable window into internal states and cognitive functions. For example, Thelen and Smith (1994) proposed that cognitive abilities emerge from the interaction between an individual and their environment, whilst Adolph (2008) discussed how the process of developing complex motor skills involves cognitive processes such as attention, perception, and memory. A review by Hötting and Röder (2013) demonstrated that physical exercise can enhance cognitive functioning and neuroplasticity, and Dell’Anna et al. (2021) explored the embodied nature and underpinnings of musical interactions. Motor coordination and development are tightly linked with cognitive processes, with a body of evidence suggesting that motor development plays a critical role in shaping cognitive development (Diamond, 2000).

Beyond cognitive abilities, movement also forms a fundamental expression of internal affective states. The body and its movements serve as a primary medium through which individuals express their arousal, intentions and emotions with others. Non-verbal expression is fundamental in conveying core emotional and internal states with Stern (1985) identifying three forms of vitality that are expressed and understood through movement: affect attunement, vitality affects, and enactments. Affect attunement refers to the ability to detect and respond to the emotional states of others through subtle movements, such as changes in tone of voice or body posture, vitality affects are the internal sensations that arise when an individual moves, and enactments are purposeful movements that are intended as an act of communication. Stern argues that these forms of vitality are essential for social communication and that they are deeply rooted in the human experience. He suggests that movement is not only a product of neural function but also a necessary component for neural development itself. Through movement, individuals learn to perceive and respond to the world around them, forming connections between sensory input and motor output.

Given the importance of movement in human cognition and communication, the ability to accurately measure and quantify movement is an increasingly attractive approach for understanding the human experience and child development. Methodological advances have led to the development of a variety of movement tracking technologies, such as optical, inertial, and markerless video based tracking that allow researchers to capture and quantify movement in a precise and objective manner. These technologies enable a deeper understanding of the neural and cognitive mechanisms underlying movement, and the relationship between movement and development (van Schaik & Dominici, 2020). However, they not only advance our understanding of the relationship of movement with cognition and development, but point towards new avenues for intervention and early identification of movement-related disorders. For example, motion tracking has been used to develop objective measures of motor function in individuals with Parkinson's disease (e.g. Ťupa et al., 2015), and to investigate the relationship between motor coordination and cognitive abilities in individuals with neurodevelopmental disorders, such as autism (e.g. Caruso et al., 2020). Indeed, there is a growing body of evidence demonstrating an underlying disruption of movement in those diagnosed with ASD (Fournier et al., 2010; Trevarthen & Delafield-Butt, 2013; Bhatt 2021), opening the possibility for the application of new approaches employing tablet game play to capture the kinematics of screen interactions to identify a specific movement signature which can then be used to aid in early diagnoses (e.g., Anzulewicz et al., 2016; Chua et al., 2022; Lu et al., 2022).

Sperry's assertion that "motor coordination is the sole product of brain function" (Sperry 1952, p.297) underscores the close relationship between movement and brain function, and the critical role of movement in shaping our thoughts, emotions, and behaviours. The development of new and innovative movement tracking technologies has opened up new avenues for exploring this relationship in populations and paradigms in which precise movement tracking would previously have been difficult or impossible

2.0 Markerless Video Based Motion Tracking - Computer Vision

Video based tracking is one such innovative approach. It is easily accessible and widely available (many software options are open source and free to use) and can be applied to already existing video data (opening up a huge quantity of secondary data for further analysis). Its non-invasive nature makes it particularly suitable to populations where traditional movement capture methods (e.g., marker or device-based movement tracking) are not suitable. Multiple forms of videography are available for users though, with the field seeing rapid developments in recent years with the rise of machine learning and AI.

These advances have begun to revolutionise the field of markerless video-based tracking and allow for the creation of accurate, robust, and highly generalisable (in the technical sense regarding an algorithms ability to process previously unseen footage) machine learning based computer vision approaches (Mathis & Mathis, 2020). Pose estimation is a method for measuring the positional configuration of body parts, which when considered over time allow for the extraction and calculation of kinematics (Egnor & Branson, 2016). These approaches are underpinned by deep neural networks (computational algorithms composed of units, organised into layers and then stacked to create networks).

Until recently the creation of these algorithms was extremely data intensive, requiring thousands of manually labelled images, making them unviable for creation and use in small to mid-sized laboratory settings (Mathis & Mathis, 2020). However, by leveraging a process known as transfer learning some modern algorithms have been able to cut substantially the amount of data required to train and tailor computer vision tracking for the needs of specific populations or studies. This process of transfer learning involves the leveraging of existing networks that have been trained using large data sets on one task or population, for deployment on another task or population.

3.0 DeepLabCut

The software proposed for use across this project, DeepLabCut (DLC; Mathis et al., 2018; Mathis & Mathis, 2020), makes use of the feature detectors from an algorithm named DeeperCut (Insafutdinov et al., 2016), retraining them according to the user's specific needs. When selecting DLC for use in our proposed approach there were several key parameters for consideration. These were the algorithm's accuracy, speed, adaptability to track the movements of less common research populations, and the amount of researcher input required in the training process (the training process requires a human to manually label the points of interest [PoI] for tracking on still images). With regard to accuracy and levels of researcher input, DLC has been shown to achieve accuracy within 5 pixels of the target PoI with as few as 100 manually labelled frames, and achieve human level accuracy with as few as 200-500 labelled images (Mathis et al., 2018; Nath et al., 2019). With regard to the software's adaptability, for the tracking of human adults there are numerous fully trained off-the-shelf options (e.g., Cao et al., 2017), but for infant and non-human primate populations a partially trainable network is required. DLC has been applied to a range of animals including mice (Mathis et al., 2018), rats (Clemensson et al., 2020), fruit flies (Zhan et al., 2021), cheetahs (Nath et al., 2019), dogs (Ferres et al., 2022) and non-human primates (Labuguen et al., 2021). It has also been applied to ultrasound images of the human calf muscle (Krupenevich et al., 2021), human children and young adults with cerebral palsy (Haberfehlner et al., 2022), human adults when running under water (Cronin et al., 2019), and is beginning to be used within developmental research paradigms (Perez et al., 2021; Solby et al., 2021).

For example, Solby et al. (2021) examined the factors impacting the exploratory or imitative problem-solving approaches of infants (specifically the effect of the level of effort exerted and success achieved by an adult model, and an infant's personal experience of failure). The authors applied DLC to secondary video data (Lucca et al., 2020) and tracked the handle of a rope as it was pulled by 1 month old infants in order to solve a problem task. This tracking generated robust positional data, showing imitative similarity and spatial variability, and revealed infant exploratory or imitative problem solving was affected by their first-hand experience of failure but not the level of effort exerted or success achieved by an adult model. This study not only highlights the insights that can be garnered from a focus on movement in traditional developmental research paradigms, but also the versatility and utility of video-based motion tracking, and DLC in particular, in developmental research.

3.1 DeepLabCut: Speed and Accuracy

When considering speed, DLC is not the fastest algorithm with regard to motion tracking, but the reason for this stems from its adaptability, accuracy and generalisability. DLC utilises ResNets, a form of deep network that make an algorithm better suited for generalisation with fewer training images (DLC does offer the use of MobileNetV2 - an alternative underpinning network - but these exchange accuracy for speed) (Kornblith et al., 2018). Overall, when utilising GPU hardware offline, DLC is capable of speeds of several hundred frames per second (Mathis & Mathis, 2020), which makes the balance between speed, accuracy, generalisation and researcher input ideal.

Importantly, DLC has also been shown to perform well against device-based motion tracking. Pouw and colleagues (2019) compared the performance of DLC to a Polhemus Liberty wired motion tracking system, and found that even when relying on a single camera perspective, DLC was comparatively high performing to its wired counterpart. Pouw and colleagues did this by estimating the time at which the maximum downward speed for a gesture was achieved and calculating the correlation of performance between DLC and Polhemus. Similarly, Perez and colleagues (2021) compared DLC to Xsens MTw Awinda wearable sensors, implementing a cross-correlation analysis that showed comparable tracking results between the approaches. This study conducted the comparison upon infants during both structured and partially restricted (through the use of supported infant seating), and unstructured and unrestricted play sessions. The authors found DLC achieved high accuracy (particularly considering the resolution of the videos) in both paradigms (although interestingly there was greater error in the unrestricted paradigm, possibly caused by the greater range of movements performed by the infant). Given such results, DLC can be considered as accurate as human annotators and capable of replacing physical markers. It is a viable and effective alternative to traditional motion tracking methodologies. Indeed, the weight and size of many wearables make them difficult to use in populations such as neonatal infants, making computer vision tracking not only comparable, but in many cases preferable. The ability to apply this approach to existing video recordings (including those recording many years in the past) make it ideal for the pipeline proposed in this chapter.

4.0 DeepLabCut Implementation on Adult-Infant Dyadic Interaction

As part of our proposed pipeline for the automatic identification of narrative, we created a computer vision model for the tracking of adults and infants during interactions with one another. For the construction of this tracking model, we trained DLC (Version 2.3) using labelled video frames taken from a selection of the videos drawn from the same archive as those analysed in Chapter 2. This model represents one of the core contributions of this thesis to infant development research, and at the time of writing was unique in its focus and nature.

4.1 Material Requirements

One of the key advantages of DLC is not only that it is free and open source, but also that it requires minimal and readily available specialist computer hardware to run (Mathis et al., 2018; Nath et al., 2019). We operated DLC on a Windows OS laptop containing a Nvidia GPU with 8GB of memory (whilst possible to run DLC using a CPU, this is substantially slower than when utilising a GPU (Nath et al., 2019)). The process of implementing DLC began with the selection of a set of videos from which a training dataset was be created.

4.2 Training Data Selection

Our primary goal was to create a model focused on tracking movement in the videos analysed in Chapter 2, but we also wanted the model to be as generalisable as possible within the limitations of the training data. With that in mind, training videos were drawn from the archival database held by the University of Copenhagen, Center for Early Intervention and Family Studies. The database is longitudinal in nature and so we ensured that training videos were drawn equally from each of the three focus age brackets (when infants were aged 3 months, 7 months and 10 months) to account for changes in infant size and proportions.

Normally, it is necessary to ensure training videos reflects the diversity of variables within the experimental recordings (e.g., gender and appearance of target participants, background conditions, and lighting) but as infants were dressed uniformly and interactions took place in a controlled experimental setting, these variables are largely accounted for. However, it was necessary to account for different camera perspectives (whether the video was shot from behind the mother's left or right shoulder - see Figure 3.2, Chapter 3) and so training videos were drawn equally from each perspective. This resulted in a training set consisting of a total 15 videos, drawn equally from each age bracket, with seven videos shot from behind the mother's left shoulder, and eight videos shot from behind the mother's right shoulder.

4.3 Procedure

From these videos we extracted 510 frames (34 frames per video) (200 – 500 frames labelled frames have been shown to produce accuracy comparable to human labelling [Mathis et al., 2018; Nath et al., 2019], although more can be required to achieve desired accuracy in the completed model). These frames were manually labelled with the points of interest that we wished to track (these were the head, shoulders, elbows, wrists, hips, knees, torso and ankles). A training dataset was then created by DLC by shuffling the annotated frames into subsets of test and train frames (95% of frames were used in training set whilst 5% were reserved for testing). DLC's pretrained underlying network is then refined using the researcher labelled images, with the performance of the resulting network being evaluated on the training and test frames. This newly tailored model can then be used to analyse unseen videos selected by the researcher. Initially, our newly refined network did not adapt successfully to unseen video footage, and so we extracted outlier frames with poor results and manually corrected the annotations made by the network. These frames were then merged back into the training dataset, and the model retrained. This process was repeated twice so as to improve the final accuracy of the model.

4.4 Model Accuracy

Following 320,000 training iterations (during which our model iteratively adjusted its parameters to minimize training error by propagating forward input data through the network, calculating the loss function [a measure of the difference between predicted and ground truth values], and backward propagating to update the model parameters using optimization techniques such as stochastic gradient descent) our model achieved a training and test accuracy of under 5 pixels. Through DLC, both a training error and test error rate can be calculated. The training error represents level of discrepancy between the prediction output of the deep learning model (the estimated pose co-ordinates) and the ground truth annotations (the frames we manually labelled) on the training dataset. A low training error indicates that the model has learnt to accurately predict the co-ordinates of PoI in the training dataset. However, it's essential to ensure that the model generalizes successfully to unseen data. The test error represents this ability to generalise, showing the performance of our model on video frames that were not used during the training process and are therefore novel. Similar to the training error rate, the test error is calculated through a comparison of the model's PoI co-ordinate predictions and our labelled ground truth annotations. A low test error indicates that the model generalizes well and can make accurate predictions on unseen data. However, if the test error is significantly higher than the training error, it may indicative of overfitting or variability in the human annotations.

For our own model, we achieved a training error rate of 2.1 pixels and a test error rate 2.6 pixels (image size was 800 x 800 pixels), demonstrating greater than human level accuracy (human level accuracy is 2.7 pixels [Nath et al., 2019]). This model allowed us to gather positional data for each point of interest from our experimental recordings (see Figure 4.1 for example [this example is not taken from the experimental recordings but from a pilot model due to ethical permissions]). DLC will produce x and y coordinates for all points of interest when visible in each frame of footage analysed by the model. From this positional data further analysis can be performed to extract kinematic variables, and explore the temporal structure of the recorded interactions.

5.0 The Automatic Measuring of Other Metrics

Whilst our pipeline will focus on the creation of an unsupervised machine learning analysis using the pose data created by DLC, this approach could ideally be augmented to include additional data sources, also gathered through automated means, that would capture other key facets of adult-infant interactions. These methods would in many respects replace the microanalysis techniques utilised in Chapter 2 of this thesis and some wider developmental research. Below we briefly discuss some of the latest approaches in automating these aspects of interaction coding and their potential for inclusion in our current pipeline.

5.1 Facial Affect

One of the most important aspects (in addition to movement) for inclusion in an automated analysis of narrative in adult-infant interaction is the type and intensity of the affective state of the involved participants. Whilst it could be possible to potentially identify affective states on the basis of pose and motion data (Chang et al., 2022), facial expressions represent a stronger indicator of such states in human beings. Automatic affect recognition has most commonly been studied in human adults and within psychological literature it is often measured using scales such as Facial Action Units (FAU), which is based on the activation of facial muscles and their movements. OpenFace (Baltrusaitis et al., 2018) is currently a leading example of an automated open-source tool capable of providing dependable estimates of facials features, action units and FAU based intensity. Unfortunately, such a tool would not be applicable to the recordings utilised in Chapter 2. This is due to the constraints associated with utilising OpenFace on secondary data that was not intended for such analysis. These constraints include our footage quality, the distance of recording equipment from participants, and the regularity of partial facial occlusion of mother or infant.

Another, perhaps more nuanced, issue concerns the application of affect recognition tools trained on adult facial data to infants whose facial features differ quite dramatically. Infant expressions can also be difficult to differentiate based purely on facial expression, and the context in which an expression is performed can be an important que for the observer. To overcome these challenges specific tools, trained and optimised for use with infants, would need to be considered (Maroulis et al., 2019; Oster, 2016).

5.2 Visual Focus of Attention

In Chapter 2 we coded for infant and adult mutual gaze (i.e. whether adult and infant were looking at one another), which allowed us to identify periods of shared attention. Ascertaining the visual focus of attention (VFOA) extends this and allows for the identification of focus beyond the dyadic pair (this can include external objects from the environment depending on the focus of the research design). A focus on VFOA rather than simply shared gaze allows us to capture and fold into our analysis additional aspects of interactions that extend beyond the primary intersubjective and include joint attention to aspects of the wider environment. Whilst gaze coding would classify such periods as 'gaze off', VFOA allows us to appreciate them as episodes of secondary intersubjectivity.

Whilst optimally identified using eye tracking (Vehlen et al., 2021), the application of such an approach can be difficult to apply to certain populations, experimental paradigms, and secondary data. However, VFOA can also be inferred based on head pose data (Ba & Odobez, 2008). Both head pose and eye gaze can be estimated using OpenFace, but this is again not without challenges. For VFOA to be properly identified, it must be possible to ascertain the focus objects in the wider environment, meaning they must also be in camera shot or at least inferable from wider contextual data, and their characteristics must be known. Noise in the environment (in the form of object clutter) can also create interference in this process. These factors prevent such an approach being applied to our own video data.

5.3 Proximity and Contact Detection

The proximity of individuals within the dyadic partnership, and how this ebbs and flows within the temporal architecture of an interaction, is also of interest. For example, Avril and colleagues (2014) utilised proximity in their exploration of relevant social signals in adult-infant engagement and calculated this from pose estimation data. The authors were exploring the relationship between different aspects of proximity (such as hand proximity, and shoulder orientation with proximity) with other factors (such as child engagement, and parental sensitivity and intrusiveness). This study not only demonstrates the importance of proxemics when attempting to capture a complete measure of an interaction, but also the different aspects of proximity which are important. It is not necessarily enough to only measure a single point on each individual (e.g. head or torso) and track their proximity through an engagement. A far richer picture of proximity can emerge when bodily orientation, hand proximity and general body proximity are considered, but for this to be accurately and reliably measured requires tracking in 3D space.

Linked with dyadic proximity, is physical contact between individuals within an engagement. The regularity, duration and form of contact is itself reflective of the underlying nature of the interaction. For example, both maternal touch (Peláez-Nogueras et al., 1996) and infant self-touch (Weinberg & Tronick, 1996) play important roles in infant regulation, and there is a bi-directional contingent relationship between maternal touch and infant vocalisation (Van Egeren et al., 2001).

The identification of physical contact, it's associated kinematics, and duration can be extrapolated from pose data given certain conditions in recording are met. For example, multi-perspective tracking using two or more cameras provides pose data with the greatest reliability for estimating physical contact. In 2D pose data derived from single camera perspective tracking, occlusions caused by one individual blocking another from camera view might be interpreted as physical contact when none in fact took place (Fieraru et al., 2020). It is even feasible to classify the type of touch that is taking place (e.g. hugging, caressing, or hitting) through an analysis of the features and characteristics of the kinematics and context, which provide additional insights into the nature of an interaction.

5.4 Speech and Vocalisation: Recognition and Classification

Within dyadic social engagements, infant and adult vocalisations play a crucial role in the creation and regulation of joint affective meaning. Basic vocal expressions exhibited by infants (including screeching, crying, laughing and babbling) hold substantial communicative value, expressing internal states of affect and arousal. As infants progress in their development and enhance their capacity for interaction with caregivers, these expressions gradually begin to transform from protoconversational to fully fledged language imbued interaction in later development.

Speech recognition has a significant research background spanning several decades and whilst it was once dependent upon on statistical modelling (for both acoustic and linguistic analysis) (Haton, 2004), developments in machine learning have brought substantial advancements in the field resulting in sophisticated systems capable of interpreting and transcribing speech both rapidly and accurately (Alharbi et al., 2021). Whilst the automatic recognition and classification of adult speech is robust and accurate, infant vocalisation has received far less attention but has nonetheless experienced significant development. For example, acoustic analysis of infant vocalizations has often employed traditional acoustic measures drawn from adult speech acoustics, but Warlaumont et al. (2010) analysed spectrographic features using a neural network consisting of a self-organizing map and a single-layer perceptron. Warlaumont et al. (2010) represents part of a rise in the use of deep learning methodologies in infant vocalisation classification. These methodologies involve replacing manually engineered feature sets with more fundamental, task-agnostic audio representations such as spectrograms or raw waveforms. These representations are then inputted into neural networks, usually CNNs or Recursive Neural Networks (Wagner et al., 2018) and have been successfully deployed in in a range of infant vocalisation recognition tasks including, for example, cry reason classification (Liang et al., 2022).

Despite these advancements in the field, the analysis of vocalisation in adult-infant interaction still requires recordings of sufficient quality that both adult and infant vocalisation can be audibly detected. For our own dataset, we were dependent on the recordings produced by the handheld video recorders which did not reliably detect and record infant vocalisation.

References

Adolph, K. E. (2008). Learning to move. *Current Directions in Psychological Science*, 17(3), 213-218.

Alharbi, S., Alrazgan, M., Alrashed, A., Alnomasi, T., Almojel, R., Alharbi, R., Alharbi, S., Alturki, S., Alshehri, F., & Almojil, M. (2021). Automatic Speech Recognition: Systematic Literature Review. *IEEE Access*, *9*, 131858–131876. https://doi.org/10.1109/ACCESS.2021.3112535

Anzulewicz, A., Sobota, K., & Delafield-Butt, J. T. (2016). Toward the autism motor signature: Gesture patterns during smart tablet gameplay identify children with autism. *Scientific Reports, 6*. https://doi.org/10.1038/srep31107

Avril, M., Leclère, C., Viaux, S., Michelet, S., Achard, C., Missonnier, S., Keren, M., Cohen, D., & Chetouani, M. (2014). Social signal processing for studying parent-infant interaction. *Frontiers in Psychology*, *5*, 1437. https://doi.org/10.3389/fpsyg.2014.01437

Ba, S. O., & Odobez, J. M. (2008). Visual focus of attention estimation from head pose posterior probability distributions. *2008 IEEE International Conference on Multimedia and Expo*, 53–56. https://doi.org/10.1109/ICME.2008.4607369

Bhat, A. N. (2021). Motor Impairment Increases in Children With Autism Spectrum Disorder as a Function of Social Communication, Cognitive and Functional Impairment, Repetitive Behavior Severity, and Comorbid Diagnoses: A SPARK Study Report. *Autism Res, 14*(1), 202-219. https://doi.org/10.1002/aur.2453

Cao, Z., Simon, T., Wei, S.E., Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. *Proceedings of the IEEE Conference on Computer Visionand Pattern Recognition*

Caruso, A., Gila, L., Fulceri, F., Salvitti, T., Micai, M., Baccinelli, W., Bulgheroni, M., et al.

(2020). Early Motor Development Predicts Clinical Outcomes of Siblings at High-Risk for Autism: Insight from an Innovative Motion-Tracking Technology. *Brain Sciences*, 10(6), 379.

Chang, A., Klein, L., Rosales, M. R., Deng, W., Smith, B. A., & Mataric, M. J. (2022). Evaluating Temporal Patterns in Applied Infant Affect Recognition. *2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 1–8. https://doi.org/10.1109/ACII55700.2022.9953842

Chua, Y. W., Lu, S. C., Anzulewicz, A., Sobota, K., Tachtatzis, C., Andonovic, I., Rowe, P., & Delafield-Butt, J. (2022). Developmental differences in the prospective organisation f goal-directed movement between children with autism and typically developing children: A smart tablet serious game study. *Developmental science*, 25(3), e13195. https://doi.org/10.1111/desc.13195

Clemensson, E. K. H., Abbaszadeh, M., Fanni, S., Espa, E., Cenci, M. A. (2020). Tracking Rats in Operant Conditioning Chambers Using a Versatile Homemade Video Camera DeepLabCut. *J. Vis. Exp.* (160), e61409, doi:10.3791/61409.

Cronin, N. J., Rantalainen, T., Ahtiainen, J. P., Hynynen, E., & Waller, B. (2019). Markerless 2D kinematic analysis of underwater running: A deep learning approach. *Journal of Biomechanics*, *87*, 75–82. https://doi.org/10.1016/j.jbiomech.2019.02.021

Delafield-Butt, J. T., & Gangopadhyay, N. (2013). Sensorimotor intentionality: The origins of intentionality in prospective agent action. *Developmental Review, 33*(4), 399-425. https://doi.org/10.1016/j.dr.2013.09.001

Dell'Anna, A., Leman, M., & Berti, A. (2021). Musical Interaction Reveals Music as Embodied Language. *Frontiers in neuroscience*, 15, 667838. https://doi.org/10.3389/fnins.2021.667838

Diamond, A. (2000). Close interrelation of motor development and cognitive development and of the cerebellum and prefrontal cortex. *Child development*, 71(1), 44–56. https://doi.org/10.1111/1467-8624.00117

Egnor, S. E. R., & Branson, K. (2016). Computational Analysis of Behavior. *Annual Review of Neuroscience*, *39*(1), 217–236. https://doi.org/10.1146/annurev-neuro-070815-013845

Ferres, K., Schloesser, T., & Gloor, P. A. (2022). Predicting Dog Emotions Based on Posture Analysis Using DeepLabCut. *Future Internet*, *14*(4), 97. https://doi.org/10.3390/fi14040097

Fieraru, M., Zanfir, M., Oneata, E., Popa, A.-I., Olaru, V., & Sminchisescu, C. (2020). Three-Dimensional Reconstruction of Human Interactions. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7212–7221. https://doi.org/10.1109/CVPR42600.2020.00724

Fournier, K. A., Hass, C. J., Naik, S. K., Lodha, N., and Cauraugh, J. H. (2010). Motor coordination in autism spectrum disorders: a synthesis and metaanalysis. *J. Autism Dev. Disord*. 40, 1227–1240.

Haberfehlner, H., van de Ven, S. S., van der Burg, S., Aleo, I., Bonouvrié, L. A., Harlaar, J., Buizer, A. I., & van der Krogt, M. M. (2022). Using DeepLabCut for tracking body landmarks in videos of children with dyskinetic cerebral palsy: a working methodology. *MedRxiv*, 2022.03.30.22272088. https://doi.org/10.1101/2022.03.30.22272088

Haton, J. P. (2003). Automatic speech recognition: A Review. *ICEIS 2003 - Proceedings of the 5th International Conference on Enterprise Information Systems*, *1*, IS5–IS10. https://doi.org/10.5120/9722-4190

Hötting, K., & Röder, B. (2013). Beneficial effects of physical exercise on neuroplasticity and cognition. *Neuroscience & Biobehavioral* Reviews, 37(9), 2243-2257.

Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., & Schiele, B. (2016). *DeeperCut: A Deeper, Stronger, and Faster Multi-person Pose Estimation Model BT - Computer*

Vision – ECCV 2016 (B. Leibe, J. Matas, N. Sebe, & M. Welling (eds.); pp. 34–50). Springer International Publishing.

Kornblith, S., Shlens, J., & Le, Q. V. (2018). Do Better ImageNet Models Transfer Better? *arXiv.* https://doi.org/10.48550/ARXIV.1805.08974

Krupenevich, R. L., Funk, C. J., & Franz, J. R. (2021). Automated analysis of medial gastrocnemius muscle-tendon junction displacements in heathy young adults during isolated contractions and walking using deep neural networks. *Computer Methods and Programs in Biomedicine*, *206*, 106120. https://doi.org/10.1016/j.cmpb.2021.106120

Labuguen, R., Matsumoto, J., Negrete, S. B., Nishimaru, H., Nishijo, H., Takada, M., Go, Y., Inoue, K. I., & Shibata, T. (2021). MacaquePose: A Novel "In the Wild" Macaque Monkey Pose Dataset for Markerless Motion Capture. *Frontiers in Behavioral Neuroscience*, *14*. https://doi.org/10.3389/fnbeh.2020.581154

Langer, S. (1967). *Mind: An essay on human feeling*. John Hopkins Press.

Liang, Y.-C., Wijaya, I., Yang, M.-T., Cuevas Juarez, J. R., & Chang, H.-T. (2022). Deep Learning for Infant Cry Recognition. *International Journal of Environmental Research and Public Health*, *19*(10). https://doi.org/10.3390/ijerph19106311

Maroulis, A.; Theuws, J.J.M. & Noldus, L.P.J.J. (2019). Validating Baby FaceReader to Analyze Infant Affective and Cognitive Responses. *Poster Presentation AIMS-2-TRIALS Annual Meeting, 11-13 June 2019.*

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*(9), 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

Mathis, M. W., & Mathis, A. (2020). Deep learning tools for the measurement of animal behavior in neuroscience. *Current Opinion in Neurobiology*, *60*, 1–11.

https://doi.org/10.1016/j.conb.2019.10.008

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, *14*(7), 2152–2176. https://doi.org/10.1038/s41596-019-0176-0

Oster, H. (2016). Baby FACS: Facial Action Coding System for infants and young children. *Unpublished monograph and coding manual.* New York University.

Peláez-Nogueras, M., Field, T. M., Hossain, Z., & Pickens, J. (1996). Depressed mothers' touching increases infants' positive affect and attention in still-face interactions. *Child Development*, *67*(4), 1780–1792.

Perez, D. L., Laudanska, Z., Radkowska, A., Babis, K., Koziol, A., & Tomalski, P. (2021). Do we need expensive equipment to quantify infants' movement? A cross-validation study between computer vision methods and sensor data. *IEEE International Conference on Development and Learning, ICDL 2021*. https://doi.org/10.1109/ICDL49984.2021.9515664

Pouw, W., Trujillo, J. P., & Dixon, J. A. (2019). The quantification of gesture–speech synchrony: A tutorial and validation of multimodal data acquisition using device-based and video-based motion tracking. *Behavior Research Methods*, *June*. https://doi.org/10.3758/s13428-019-01271-9

Solby, H., Radovanovic, M., & Sommerville, J. A. (2021). A New Look at Infant Problem-Solving: Using DeepLabCut to Investigate Exploratory Problem-Solving Approaches. *Frontiers in Psychology*, *12*(November), 1–18. https://doi.org/10.3389/fpsyg.2021.705108

Sperry, R.W. (1952). Neurology and the mind-brain problem. *American Scientist*, 40, 291-312.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of* *cognition and action*. MIT Press.

Trevarthen, C., & Delafield-Butt, J. T. (2013). Autism as a developmental disorder in intentional movement and affective engagement [Hypothesis & Theory]. *Frontiers in Integrative Neuroscience, 7*, 49. https://doi.org/10.3389/fnint.2013.00049

Ťupa, O., Procházka, A., Vyšata, O. et al. Motion tracking and gait feature estimation for recognising Parkinson's disease using MS Kinect. *BioMed Eng OnLi*ne 14, 97 (2015). https://doi.org/10.1186/s12938-015-0092-7

Wagner, J., Schiller, D., Seiderer, A., & André, E. (2018). Deep learning in paralinguistic recognition tasks: Are hand-crafted features still relevant? *Proceedings of the Annual Conference of the International Speech Communication Association,* 147–151. https://doi.org/10.21437/Interspeech.2018-1238

Warlaumont, A. S., Oller, D. K., Buder, E. H., Dale, R., & Kozma, R. (2010). Data-driven automated acoustic analysis of human infant vocalizations using neural network tools. *The Journal of the Acoustical Society of America*, *127*(4), 2563–2577. https://doi.org/10.1121/1.3327460

Weinberg, M. K., & Tronick, E. Z. (1996). Infant affective reactions to the resumption of maternal interaction after the still-face. *Child Development*, *67*(3), 905–914.

Zhan, W., Zou, Y., He, Z., & Zhang, Z. (2021). Key Points Tracking and Grooming Behavior Recognition of Bactrocera minax (Diptera: Trypetidae) via DeepLabCut. *Mathematical Problems in Engineering*, *2021*. https://doi.org/10.1155/2021/1392362

Figures and Tables

Figure 4.1. A young toddler has her movements tracked whilst she pushes her toy truck is a solo project encouraged by her father (who is out of camera shot). Images 1-3 capture the toddler as she pushes her truck, whilst images 4-6 capture her pausing, steadying herself, and then applauding her own efforts. A tracking model was created using DLC and trained on 200 images to provide positional and tracking data for the feet, eyes, nose, and hands of the toddler during a short recording lasting 15 seconds. Each coloured dot represents a tracking point for which positional data is collected in each frame of the video. Due to the video only being captured from a single camera perspective, only 2D coordinated are provided.

Chapter 5 - Quantifying Narrative: Overcoming Biases in Human Interpretation

1.0 Utilising Pose Estimation and Associated Challenges

Where footage allows for the additional metrics outlined above to be measured, the data can be folded in with pose estimation for analysis. However, the pipeline below will be based on the proposed analysis of the interactions that formed the basis of our longitudinal study in Chapter 3, and the nature of these recordings prevent the automation of these aspects of our adult-infant engagements. Instead, and following the logic that led to the use of pose data to classify affective states (Chang et al., 2022), we propose the use of pose estimation and motion data in order to ascertain the presence of narrative phases. Both supervised and unsupervised neural networks can be implemented achieve this aim. A supervised approach would enable automated detection of narrative phases based on human labelled training data. This approach is contingent on human coding, but still enables the discovery of novel features within the interactions from its data-driven, computational approach. An unsupervised approach enables automated detection of narrative phases, but with the addition that its pattern recognition is not contingent on any human-coded ground truth. Given our aim to remove human bias from the identification of narrative, below we will focus on exploring unsupervised approaches in greater detail and propose a potential pipeline for such an analysis.

The attraction of implementing such an approach stems from the weaknesses inherent to human rater coding of video data. Human annotation is often time and labour intensive whilst also being subject to the biases of the human raters. The changes between narrative phases can be a subtle and nuanced, and so by their nature will be open to interpretation. Reliable, automatic classification of interactions is therefore preferable as it removes these drawbacks. The ability to implement such a procedure though has come about only relatively recently. If focusing on pose and motion, it requires the automatic tracking of key points of interest (as outlined above) and the use of novel machine learning classification/clustering approaches. The tracking of pose and motion using computer vision based markerless motion tracking has been focused upon human adult models. The creation of tailorable computer vision models (such as DLC) and advances in machine learning clustering/classification approaches has now opened the potential for the automative identification of narrative phases.

1.1 Challenges

Before discussing potential approaches and outlining a proposed pipeline, it is first necessary to highlight some of the challenges associated with this enterprise. Firstly, challenges associated with the video footage itself. Some of these issues can be mitigated through the collection of carefully organised recordings of adult-infant interactions, but they are harder to avoid when utilising secondary data recorded without the purpose of computer vision motion tracking in mind. For example, the resolution of footage can inhibit the accurate tracking of data points (although DLC is robust in this regard) and is a particular issue when attempting to identify facial expression, whilst occlusion of a PoI (either by a dyadic partner or object in the environment) can remove valuable data points and be prohibitive for further analysis if it is a regular occurrence.

Finally, there is an issue regarding depth. Where only a single camera perspective is available, DLC cannot create an estimate of pose in 3D space. Instead, a 2D estimate is provided which can limit our understanding of the interaction that is taking place. There are two possible options in this regard. Firstly, 2D data can be used and features extracted which best allow for the classification the movements that are hypothesised to compose a narrative. However, this may be a suboptimal solution as key components of the interaction that take place within the third, depth axis may be lost or mis-interpreted in 2D space. An alternative, is the estimation of depth *post hoc*.

1.2 Moving Beyond Two-Dimensional Spatial Positioning

This limitation of flat 2D video-based motion capture is due to the single perspective tracking obtained from most readily accessible single perspective videos (such as those typically used in infant studies or home videos). When planned from the beginning, this can be overcome through the use of synchronised multi-camera perspectives, with toolkits such as "Anipose" specifically designed to work in tandem with DLC to facilitate the integration of three or more camera perspectives. To do this they contain robust inbuilt calibration modules and refinement filters, tailored for 3D tracking, to produce reliable 3D motion capture (Karashchuk et al., 2021). However, such additional toolkits are cumbersome and may be unnecessary. For example, DLC offers a simpler in-built facility with the ability to produce 3D spatial positioning by calibrating two camera perspectives within a tracking model, and triangulating the 2D positioning data to produce coordinates in three-dimensional space (Nath et al., 2019).

From a single camera perspective, motion tracking across three dimensions becomes considerably more difficult due to the obvious lack of available depth data. However, CNNs have been trained to offer frameworks capable of learning discriminative image features from which 3D spatial positioning can be estimated (Kosourikhina et al., 2022; Park et al., 2016). Park and colleagues (2016) proposed using a combination of features specific for the estimation of 3D spatial position, and information from 2D spatial pose estimation. For example, on the basis of reliable 2D pose estimation data, 3D pose estimations that are impossible or highly unlikely to have occurred can be ignored (e.g., if joints are found to have taken unnatural positions in 3D pose estimation on the basis on 2D positional data). On this basis, Park et al., (2016) suggest utilising 2D joint classification results as input features for the estimation of 3D poses, and using relative positioning between multiple joints to inform 3D positional data to improve data provided by baseline CNNs.

An alternative approach is to utilise existing 3D skeletal data or exemplars to lift 2D poses into 3D space (Tome et al., 2017; Andriluka et al., 2010). Some existing CNNs have additional layers embedded within their architecture that lift 2D co-ordinates into 3D using probabilistic 3D models that ensure the resulting pose estimations are physically plausible (Tome et al., 2017), This has the additional benefit of improving 2D estimation by ensuring it aligns with the physically plausible anatomical limits of the 3D model. Despite the emergence of these approaches in recent years, it remains a considerable challenge to transpose 2D pose estimations into 3D space. As such, the pipeline described below will focus on the use of 2D pose estimation data, whilst retaining the potential for future amendment for the incorporation of 3D pose estimation data.

2.0 Unsupervised Clustering According to Narrative Phase

2.1 Unsupervised Learning

To cluster or classify pose estimation data (and derivative motion data) there are multiple possible approaches. For example, supervised learning approaches leverage a set of pre-labelled data, where data points (in our case post estimates within a frame of video footage) are tagged as belonging to a specific narrative phase. This labelled data serves as the foundation for training a machine learning model to automatically recognize and classify similar patterns in unseen data. In many respects, a supervised approach represents a clear way to help ensure our model is focused upon specific features most likely to identify patterns of arousal and intensity in an interaction. However, it also introduces the potential for some human rater bias in the creation of training data.

Given our primary aim is to address the criticism that the identification of narrative by human beings is an inherently flawed exercise (as human beings would be biased to seeing narratives in the world regardless of whether narrative exists) we instead favour the creation of a fully unsupervised analysis pipeline. Unsupervised learning operates on unlabelled data, where the model has no prior knowledge of the specific patterns present. Instead, it seeks to discover hidden structures and relationships within the data itself. This approach removes the need for pre-labelled training data, and provides a fully automative approach for narrative identification.

2.2 Existing Application in Infant Research

Unsupervised learning approaches to temporal analysis have a growing base in existing research with temporal clustering algorithms demonstrating promise in identifying human action. For example, in Zhou et al. (2013), combined spectral clustering and dynamic time warping was employed to group time series data (effectively building taxonomies of facial expressions), in Si et al. (2011) event co-occurrence clustering was used to construct action dictionaries, and in Wang et al. (2006) the authors explored deformable template matching to identify action classes within static images. Additionally, Turaga et al. (2009) utilized switching linear dynamical systems for video-based action clustering.

More pertinently, unsupervised learning has also been employed in infant development research. One example of this is the work of Chu et al. (2017) who implemented an unsupervised algorithm to detect synchronized affective responses between parents and infants during face-to-face interactions. This algorithm, utilizing facial features, recognized shared areas of action between the infant and mother in sections of video (specifically synchronized smiles), signifying positive emotional communication. Similarly, Prabhakar et al. (2010) utilised unsupervised learning to identify playful parent-child interactions in

YouTube videos by analysing recurring spatiotemporal patterns. These studies showcase the potential of unsupervised learning to reveal new insights into early emotional interactions, paving the way for further exploration in this domain.

3.0 Procedure

Our objective through this proposed analysis pipeline is to uncover patterns within sequences of video frames that align with the narrative structure proposed by Bruner (1993). Due to the secondary nature of our data (the recordings and engagements that formed the basis of our analysis in Chapter 3) we restricted this analysis to estimated body pose skeletons as well as derivative motion based features. However, this choice does not limit the scalability of the analysis pipeline, and future work (discussed below and in greater detail in the Chapter 4) could incorporate additional modalities such as those mentioned above. Each feature representation will enable multiframe-level clustering. To validate our model a quantitative comparison is proposed to ascertain the alignment between our own manually human rated data from Chapter 2, and this novel machine learning temporal clustering.

Before outlining our pipeline, it is necessary to highlight that this analysis will be focused on 2D pose estimation data rather 3D pose estimation data. As discussed above, utilising 3D pose data holds numerous advantages for our project. Primarily, the inclusion of depth information would have offered enhanced accuracy, detail, and nuance to our proximity information, aiding in distinguishing situations that appear ambiguous from a 2D perspective. This, in turn, would have impacted our choices regarding normalisation and handling of proximity data. The richer pose representations would also likely have enriched the accuracy of our clustering pipeline. However, achieving robust 3D pose estimation presents significant challenges, particularly when applied to secondary data with only a single camera perspective. This difficulty is further exacerbated in adult-infant recordings due to the inherent scale differences between parents and children, requiring additional considerations for accurate 3D pose estimation.

3.1 Processing Pipeline for Pose Estimation Data

3.1.1 Data Preparation and Pre-processing

This section outlines the various processing steps and filters applied to our raw pose data (obtained from the markerless pose estimation model described in Chapter 4). Our goal is to transform this raw data into a format suitable for our clustering algorithm. Figure 5.1 provides a visual representation of this workflow.

Prior to extracting features or clustering, it is necessary to clean and process our data to determine which recordings and frames are suitable for analysis. For example, for our machine learning pipeline it is important video frames are selected to ensure minimal occlusion of tracking points. For illustrative purposes we propose utilising the same recordings as used in Chapter 2 as this will provide the greatest accuracy when validating the model through comparison against human rater coding.

These data cleaning steps are relatively straightforward approaches to remove potentially unreliable data and outlier frames. The precise nature of the pre-processing filters to be applied is individual to the dataset and the specific research aims of a project as there is a trade-off between data accuracy and data availability (many of the outliers in our own data stem from limitations within the pose estimation pipeline itself, often caused by inherent issues in the recordings such as occlusions or truncated body parts). It is necessary to strike a balance between pose estimation accuracy (which will directly affect the accuracy of our clustering approach) and data availability. We believe the below process should minimise the impact of noise (found within our pose estimation data) upon our final clustering process.

3.1.2 Pose Estimate Video Confidence Filter

We first propose an automatic filtering of frames and videos based on the confidence scores of tracking points (the confidence score represents the likelihood that the point of interest was actually visible in the relevant frame). The precise nature of this filter is individual to the recordings under scrutiny and the specific research aims of a project (a higher confidence inclusion cut-off score will result in a larger cull of fames for analysis, whilst a lower inclusion cut-off score can impact the accuracy of the clustering process). For our own project we have 54 recordings each with 5 minutes of interaction footage. We propose excluding videos where more than 30% of the frames have an average confidence score across all PoI below 60% (this would result in the exclusion of 2 videos from our analysis). We propose this balance as even if a smaller data set is created it helps ensure reliable pose estimates for the creation of robust and informative clusters. Many frames with lower confidence scores were likely caused by occlusions or situations where the parent and child were in very close proximity, which can significantly challenge the pose estimation model and lead to suboptimal pose estimation.

3.1.3 PoI Examination and Exclusion

Understanding the utility of each PoI within the context of our analysis, and the performance of DLC in tracking each PoI, can help inform whether a PoI should be taken forward into the final stages of this pipeline for clustering. If any PoI is found to have suboptimal tracking it would likely be beneficial to remove these from analysis, especially in cases where their theoretical benefit is less well founded. Examination of the pose estimated confidence scores from our tracking (rated on a scale from 0 to 1) revealed some PoI exhibited significantly lower average scores and higher variability than their peers (see Table 5.1). Specifically, the infant hip PoI yielded average confidence scores below 0.5. Such low scores make these PoI problematic for inclusion in analysis. Furthermore, within the contextual framework of understanding the temporal architecture of adult-infant interactions, these PoI hold relatively little significance due to the seated position of the infant. As such, in order to prevent the undue influence of the PoI on clustering outcomes, we would exclude the two infant hip PoI (left and right hip) from subsequent processing stages and our final analysis.

3.1.4 Addressing False Estimations Through High Motion Frame Filtering

The experimental set-up in our recordings (described in Chapter)3 was initially designed to allow for human raters to code behaviours and observations (for example the affective states of both infant and adult). To facilitate this, mirrors were placed strategically in the camera shot so that mother facial affect could be identified (the mother had their back to the camera - see figure 3.2, Chapter 3). Unfortunately, a consequence of this is that there are rare times when DLC confuses the reflection in the mirror with the authentic individual. This is also sometimes true for the infant who in later infancy often turns to face and investigate to their own reflection. Another example of this kind of error occurs in the rare instances of another adult entering the room during a recording (DLC can mistakenly attribute PoI to the novel adult). Such changes result in dramatic and unrealistic jumps from one frame to another for specific PoI. By filtering these we can eliminate frames with inaccurate detections or estimations.

This filter works by computing motion data as outlined in section 3.1.8. If, for a given frame, the motion data surpasses a predefined threshold, across a minimum number of PoI coordinates (it would likely be too strict and potentially result in too large a data loss to remove frames because of a false estimate relating to a single PoI), the frame is excluded. By eliminating such frames, we ensure the clustering process relies on reliable motion data, ultimately leading to the formation of meaningful clusters.

3.1.5 Assessment of Missing Frames

A serious draw back of the pre-processing steps described above is that they have the potential to deplete the stock of available data for clustering, and therefore impact the quality of the clustering process. Depending on the number of frames removed by the above filters, it may be beneficial to interpolate missing frames based on surrounding data. This is possible to a reasonable level of certainty if the surrounding data is robust and reliable. As such, the rationale behind this procedure is twofold: firstly, to expand the scope of processable data, and secondly, to bolster the coherence of sequential data. Employing a linear interpolation method, with a threshold of 12 frames as the maximum permissible gap for interpolation, frames can be interpolated assuming a linear transition between the coordinate values of two valid frames. Subsequently, the coordinates for the interpolated frames are determined based on the values they would assume along this linear trajectory. This process does not in any way guarantee that interpolated frames will be an accurate representation of reality (real world co-ordinate changes do not necessarily adhere to a linear model and are subject to the internal states of the individual), but we would anticipate that motion occurring within a 12-frame interval would generally exhibit limited variation. A drawback to this approach arises in cases where individuals are exhibiting variable levels of jerk, effecting the linear relationship between co-ordinates from one frame to another and thus undermining assumptions that can be made regarding missing data points. However, the 12-frame threshold should be short enough to minimise any impact of this.

3.1.6 Data Normalisation

In order to analyse the pose estimation and motion data of multiple individuals, either within or between recording sessions, it is often necessary to address variations arising from differences in body size, perspective, recording space, and time. A common strategy for mitigating such variations involves the application of one or more normalisation techniques. Generally speaking, normalising data before training models yields substantially improved performance making it an integral step in a preprocessing pipeline (Cao et al., 2016). Additionally, the practice holds particular significance for methodologies employing distance measures, such as nearest neighbour classification and clustering algorithms (Cao et al., 2016).

Poppe and colleagues (2014) outline possible approaches that can be deployed depending on the form of normalisation required. For example, where it is not possible to temporally synchronise multiple perspectives at the time of recording, it may be necessary to synchronise videos by trimming footage to ensure identical start and end points. Or where there is frame rate misalignment it may be necessary to resample measurements in each sequence to ensure uniform temporal spacing (thus aligning the data to a standardized rate). Spatial normalisation is also often required, particularly in cases where pose estimates and postures are being compared between individuals as discrepancies in global position can distort the comparisons. To address this issue, poses estimates can be normalised with respect to position by centring all pose estimates around the root of the body. For example, the pelvis or torso can serve as a centre point with its positional co-ordinates translated to (0, 0, 0). This normalisation process involves mean centring the data by subtracting the original position of the central tracking point from the position of each of the other tracking PoI (Poppe et al., 2014). Such an approach could be utilised in our own normalisation process when analysing the pose estimation data from either mother or child independently (discussed below).

However, as these pose normalisation techniques focus on coordinates belonging to individuals they neglect the inherent interdependent relationships that exist between poses in dyadic interaction. One method of normalising post estimates whist also taking account of these relationships involves individual normalisation in geographic space (as outlined above) followed by the computation of additional proximity features. However, this strategy necessitates establishing a weighting scheme to balance the influence of the original pose features and the newly introduced proximity features. Determining an optimal weighting function can be challenging and application-specific.

An alternative method leverages the z-score normalisation technique (e.g. Cao et al., 2016; Pei et al., 2023). Z-score normalisation standardises each data point by subtracting the mean value and dividing by the standard deviation of the entire dataset. This transformation results in a normalised representation with a zero mean and unit standard deviation. Therefore, the normalised coordinates of a specific PoI are not solely determined by its own original value but are also impacted by the positions of other tracking points in the frame. Consequently, this approach preserves the relative distances between tracking points within the normalised coordinates and therefore captures the inherent relational structure of the pose data.

3.1.7 Intra-personal Normalisation

Whilst the primary purpose of the pipeline described here is to utilise both parent and child pose estimation and motion data in order to map the temporal architecture of their interaction, mother and infant pose estimates could also be informative when considered independent of one another. Given our precise aim is the identification and clustering of narrative phases, it could even be the case that the pose estimates from child or mother could be more effective independently than combined, therefore requiring an intra-personal normalisation process.

The intra-personal normalisation of pose estimation data would normalise PoI relative to one another on an intra-personal basis. We therefore propose two possible approaches in this regard. The first approach normalises all the PoI on an individual around a central or root point on the body such as the torso or hip (as outlined above in Section 3.1.6). The second approach is similar to the combined approach utilising the z-score method but independently for each person within the dataset. Both approaches would allow mother and child pose data to be independently used for feature extraction and clustering.

3.1.8 From Pose Estimation to Motion

As well as raw pose estimation data, we propose utilising motion data (derived from pose estimates) in our clustering process. Motion data, even in the raw form calculated as described below, is indicative of the states of arousal and intensity that exist both within and between members of a dyadic partnership. For our purposes, motion data across frames is characterised by the differential displacement between the current frame and its immediate predecessor. This displacement is computed for the x and y coordinates for each PoI independently. This involves subtracting from coordinates of a current frame the coordinates of the previous frame. To ensure data integrity, motion values are only calculated for frames where both the current frame and the preceding frame possess valid PoI coordinates (the data interpolation step described above will minimise the occurrence of missing frames, maximising the amount of available motion data).

3.2 Feature Extraction, Selection and Representation

Before discussing the specific approach for clustering our data according to narrative phase, it is important to outline the process of feature extraction and selection. The effectiveness of clustering algorithms relies heavily on the quality of features extracted from the raw data. Feature extraction serves to transform raw data into a set of informative features that encapsulate relevant characteristics of the underlying temporal pattern. These features form the variables that the final model uses to cluster our pose estimation data, making their selection an essential component of our machine learning workflow. Features are determined strategically from the dataset so as to reduce model complexity (resulting in higher computational efficiency) and the risk of overfitting (where the model performs well against training/test data but does not generalise effectively to novel data) whilst also promoting the most effective and accurate clustering.

The feature selection process is additionally significant due to its role in promoting transparency and interpretability in the final model. The importance of explainability in feature selection and wider machine learning has become a broader topic of interest amongst growing media and government attention focused on the development of artificial intelligence (e.g. European Commission, 2019). Whilst deep learning models often surpass methods reliant on manually-crafted features, they can also be opaque and make it difficult to understand how they arrive at their predictions. This lack of transparency, akin to a black box, can be problematic when the model exploits unintended relationships in the data to achieve high performance, potentially leading to flawed reasoning (Fraile et al., 2021). To address this black box problem, the field of explainable artificial intelligence has emerged and gained significant traction in recent years (Adadi & Berrada, 2018). Moreover, the European Union has recognized explainability as a fundamental principle for trustworthy AI (European Commission, 2019).

3.2.1 Feature Extraction

When extracting and selecting features, pose estimation offers an excellent body of data from which informative features can be drawn as it provides an informative visual modality for exploring human interactions (Sun et al., 2023). It provides insights into non-verbal communication, movements and gestures and allows the calculation of movement kinematics. These allow a view as to the internal levels of arousal of an individual. For our own pipeline we propose utilising the tsfeatures package (Hyndman et al., 2023) in R (R Core Team, 2024) to extract features from our pose and motion data.

The tsfeatures package provides a set of functions designed to capture specific characteristics of time series data. These include statistical features (including measures such as mean, variance, skewness, and kurtosis), structural features (such as trend, seasonality, periodicity, entropy and fractal dimension), and frequency domain features derived from spectral analysis techniques that characterize the frequency components present in the data (for example spectral entropy, dominant frequency, and spectral energy). These facilitate the analysis of underlying temporal dynamics and subsequent machine learning tasks such as clustering.

3.2.2 Feature Selection

Feature extraction will be followed by feature selection with the aim of identifying a subset of the most informative features that contribute significantly to the discriminatory power of the model while discarding redundant or irrelevant features. Two possible approaches to feature selection include principal component analysis (PCA) and mutual information. PCA identifies a set of orthogonal linear transformations, known as principal components, which capture the greatest variance within the data. By selecting a subset of principal components that explain a significant portion of the total variance, PCA effectively reduces the dimensionality of the feature space while retaining the most informative features. This compressed representation offers enhanced interpretability by facilitating a clearer understanding of the underlying structure within the data whilst also improving computational efficiency. Mutual information can be employed to assess the relevance and discriminative power of individual features with respect to the clustering task. By ranking the features based on their importance scores, less informative features can be pruned from the feature space, resulting in a more refined set of features that encapsulate the underlying structure and variability of the data set. For the purposes of our pipeline both of these approaches are suitable and effective options.

3.2.3 Feature Representations

Our raw pose data can be represented and clustered according to different combinations, and a necessary step in a workflow such as this is discerning the best possible representations. Whilst we would envisage that mother-child combined motion data and mother-child combined pose data would provide the best representations for narrative phase clustering, it is also possible that separating these data will present a more accurate representation of the temporal changes within the interaction. For example, infant pose and motion data could represent a more unfiltered characterisation of the levels of arousal and intensity in the dyadic interaction. On this basis we propose exploring potential combinations of these representations in our analysis so as to combine complementary insights into the temporal architecture of the interaction. To this end, we suggest examining six distinct representations: parent pose, child pose, parent motion, child motion, combined pose (comprising both parent and child poses), combined motion (incorporating both parent and child motion).

3.3 Clustering

At the core of this unsupervised approach, we propose the clustering of data based on levels of similarity. The choice of a suitable clustering algorithm is crucial for effective data analysis and hinges on the specific characteristics of the data and the intended goals of the analysis. Clustering algorithms can be grouped into distinct types, with a prime example being partitional algorithms such as K-means (Reddy & Vinzamuri, 2018). Normally, a key limitation of K-means and other partitional methods is the requirement for a pre-defined number of clusters, which can be challenging to determine if a fully exploratory analysis is being undertaken. However, in the pipeline proposed here we know a priori the seed clusters as they will be defined according to the four phases of narrative.

3.3.1 K-means Clustering

K-means aims to group similar data points into K pre-defined clusters based on their Euclidean distances. The algorithm iteratively assigns each data point to the nearest cluster, recalculating the cluster centres after each iteration. This process continues until the clusters stabilise, revealing groups of data points with similar features, representing potential interaction patterns.

K means clustering has an extensive existing record in unsupervised clustering. For example, in the field of unsupervised action segmentation (which aims to identify meaningful segments of actions within videos without requiring prior knowledge of the specific actions present), Wang et al. (2022) utilised a "co-occurrence action parsing" algorithm. Thito estimate the temporal order of sub-actions within a video by analysing their co-occurrence and recurrence across the entire dataset. The process firstly involved self-supervised feature generation which used a neural network to learn informative representations directly from video frames. From these features, video frames were clustered using a K-means clustering approach with a pre-defined number of clusters. Several matrices were then generated from the clusters, including a video score matrix (representing the probability of a frame belonging to a cluster), a co-occurrence matrix (capturing the likelihood of observing two specific clusters within a video), and a cluster temporal location histogram (indicating the typical temporal locations of each cluster). These matrices, along with the self-supervised features, were used to estimate the temporal sequence of sub-actions within each video. Also in the field of unsupervised action segmentation, Kukleva et al. (2019) proposed a method that first learns a continuous temporal embedding of frame-based features. These features are then clustered using K-means, with the optimal number of clusters (K) determined based on ground truth labels. When ground truth is unavailable, the entire dataset is used for embedding and clustering, followed by representing each video as a "bag-of-words" based on the identified clusters. Finally, videos are clustered again to identify groups with similar sub-actions.

3.3.3 Approach

The pose estimation and motion data obtained from our video footage and pre-processing can be employed for each of the six feature representations listed above. This data can then be flattened and subjected to clustering using the K-means algorithm. K-means was chosen for its computational efficiency, and the fact we can define the precise number of clusters we wish to use (the selection of four clusters being based on the four narrative phases [for a broader exploratory analysis that seeks to cluster to an undefined number of groups, hierarchical clustering would have many benefits]). Moreover, K-means utilisation for clustering pose estimation data has a proven track record (e.g. Johnson & Everingham, 2010; Zhang et al., 2019).

3.4 Comparative Validation

To validate the results of our models we envisage a quantitative comparison using Cohen's Kappa to check alignment between our own human coded data from Chapter 2, and this novel machine learning temporal clustering performed by our K-means algorithm. As in Chapter 2, it would be necessary to calculate both an event-based kappa and a time-based kappa. Using both forms of Kappa will provide a more nuanced indication of inter-rater agreement between the different sets of labels. Time-based Kappa will assess the overall consistency in labelling over time. The Kappa contingency table for a time-base Kappa will contain rows that represent time intervals, and columns which represent the labels assigned by our human coder and the k-means algorithm. Observed and expected agreement is then calculated based on the temporal consistency between values. The event-based Kappa is akin to a classic Cohen's Kappa whereby agreement between the labels given to specific interaction events by algorithm and human are compared to create a Kappa score. The results of this inter-rater reliability analysis will allow us to ascertain the overall validity of our models, and the utility of our different feature representations.

4.0 Wider Application

The application of machine learning approaches to the analysis and interpretation of adult-infant interaction remains in its infancy. Our approach represents the first attempt to apply this to the field of infant narrative, but it's potential to facilitate the exploration of narrative and wider infant development is substantial. Moreover, approaches such as ours have the capacity for wider application and in combination with the features we identify could provide insights into both typical and atypical child development. For example, in examining the temporal structure of adult-infant interaction when children are at increased risk of developing ASD. This would help provide empirical evidence to support

Bruner and Feldman (1993), who in their narrative deficit hypothesis suggest that ASD arises from an infant's inability to engage in joint, non-verbal narrative meaning-making. This hypothesis suggests ASD may stem from a deficit in narrative communication skills and the inability to organise personal experiences into a coherent narrative framework. By analysing the overarching framework in which ASD communication take place it would be possible to ascertain how and whether this differs from the narrative temporal structure found in typically developing children. This could help inform not only early interventions but also the very nature and structure of teaching delivery in educational settings.

Exploring the temporal structure within which learning takes place is itself another possible application for our pipeline. For example, as well as analysing the structure of teaching delivery (similar to how multimodal analysis has been applied to recorded lectures to investigate the use of various modalities to communicate complex information to students [Alviar et al., 2019]), student engagement and response to lesson delivery could be analysed through a multimodal lens to ascertain the level of entrainment to narratively structured learning. This emerging field of inquiry holds considerable promise across diverse domains in research and practice.

5.0 Conclusion

Motion tracking offers an avenue to examine core aspects of agency, social interaction, and interaction with the environment through an exploration of the structures that underpin human experience. Whilst different motion tracking approaches offers benefits that make them well suited to specific experimental paradigms, video motion tracking provides the opportunity to measure and quantify the motion of key points of interest in pre-recorded videos. Its tailorable nature allows it to be applied to populations where optical motion tracking and device-based tracking encounter significant challenges (which can interfere with the movements performed by a subject). DLC's existing record of motion tracking in such populations demonstrates it versability whilst comparisons to more established device based tracking evidence it's accuracy.

Our analysis pipeline for the pose estimation produced by DLC represents a novel approach to the analysis of narrative in human interactions. The appeal of adopting such an approach arises from the inherent limitations associated with human rater coding of video data. Human annotation is frequently characterized by being time and labour intensive, and is also susceptibility to the biases of individual raters. Given that transitions between narrative phases can often be subtle and nuanced, they inherently lend themselves to varying interpretations. Consequently, reliable automatic temporal clustering of interactions is an attractive option, as it mitigates the drawbacks inherent in the current identification of narrative.

In its current form, this pipeline is relatively specific to the secondary data we used in Chapter 2 but could easily be generalised to other data sets. Indeed, one of the key weaknesses of the proposed analysis stems from the nature of the video footage. Although possible to estimate depth on the basis of 2D pose estimation (as discussed above), this approach is far less accurate than 3D pose estimation created using multi-camera perspectives (S. Zhang et al., 2022). The adoption of a 3D pose estimation approach using DLC multi-camera pose estimation (Nath et al., 2019) would address these limitations. It would provide the capability to capture depth and proximity information, as well as overcoming the challenges posed by adult-infant occlusions in our current 2D representations. However, to do this would require the collection of primary data and the recording of interactions within an experimental paradigm specifically designed with this purpose in mind. This was not possible within the constraints of this thesis (as outlined in the general introduction). However, the enriched data this would provide would have significant advantages in detecting semantically meaningful visual cues and situations. The inclusion of depth and proximity data would allow for a more nuanced understanding of an engagement, capturing artifacts that might otherwise be obscured in a 2D representation. All this would be highly beneficial in the identification of narrative.

The pipeline suggested here could also be extended through the inclusion and integration of additional modalities. As discussed extensively in Chapter 1, adult-infant interactions are by their nature multi-modal, and the narrative meaning making that is born within these is composed of a rich repertoire of multi-modal communication (including but not limited to vocalisation, facial expression, gaze, and touch). The inclusion these additional modalities builds a more complex picture of the interactions being analysed, and creates a more informed clustering mechanism. The inclusion of these modalities in such an approach would in itself represent an important advancement in the automated analysis of adult-infant interaction.

Beyond the interaction, there is also potential for the incorporation of participant demographic and characteristic information. This data could include socio-economic status, cultural background, age, and even Bailey scores (as we included in some of the GLMM utilised in Chapter 2). Integrating such data could not only aid in contextualising the output of our models, but also allow for an exploration of the wider factors in infant development that impact the creation of joint narratives.

References

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). OpenFace 2.0: Facial Behavior Analysis Toolkit. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 59–66. https://doi.org/10.1109/FG.2018.00019

Cao, X. H., Stojkovic, I., & Obradovic, Z. (2016). A robust data scaling algorithm to improve classification accuracies in biomedical data. *BMC Bioinformatics*, *17*(1), 1–10. https://doi.org/10.1186/s12859-016-1236-x

Chang, A., Klein, L., Rosales, M. R., Deng, W., Smith, B. A., & Mataric, M. J. (2022). Evaluating Temporal Patterns in Applied Infant Affect Recognition. *2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 1–8. https://doi.org/10.1109/ACII55700.2022.9953842

Chu, W.-S., De la Torre, F., Cohn, J. F., & Messinger, D. S. (2017). A Branch-and-Bound Framework for Unsupervised Common Event Discovery. *International Journal of Computer Vision*, *123*(3), 372–391. https://doi.org/10.1007/s11263-017-0989-7

European Commission. (2019, April 9). High-level expert group on artificial intelligence: Ethics guidelines for trustworthy AI. [Report]. *AI HLEG*. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Fraile, M., Lindblad, J., Fawcett, C., Sladoje, N., & Castellano, G. (2021). Automatic analysis of infant engagement during play: An end-to-end learning and Explainable AI pilot experiment. *Companion Publication of the 2021 International Conference on Multimodal Interaction*, 403–407. https://doi.org/10.1145/3461615.3485443

Hyndman R, Kang Y, Montero-Manso P, O'Hara-Wild M, Talagala T, Wang E, Yang Y

(2023). *tsfeatures: Time Series Feature Extraction*. https://github.com/robjhyndman/tsfeatures.

Johnson, S., & Everingham, M. (2010). Clustered pose and nonlinear appearance models for human pose estimation. *British Machine Vision Conference, BMVC 2010 - Proceedings*, *ii*, 1–11. https://doi.org/10.5244/C.24.12

Kukleva, A., Kuehne, H., Sener, F., & Gall, J. (2019). Unsupervised Learning of Action Classes With Continuous Temporal Embedding. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12058–12066. https://doi.org/10.1109/CVPR.2019.01234

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, *14*(7), 2152–2176. https://doi.org/10.1038/s41596-019-0176-0

Pei, X., Zhao, Y. hong, Chen, L., Guo, Q., Duan, Z., Pan, Y., & Hou, H. (2023). Robustness of machine learning to color, size change, normalisation, and image enhancement on micrograph datasets with large sample differences. *Materials and Design*, *232*, 112086. https://doi.org/10.1016/j.matdes.2023.112086

Poppe, R., Van Der Zee, S., Heylen, D. K. J., & Taylor, P. J. (2014). AMAB: Automated measurement and analysis of body motion. *Behavior Research Methods*, *46*(3), 625–633. https://doi.org/10.3758/s13428-013-0398-y

Prabhakar, K., Oh, S., Wang, P., Abowd, G. D., & Rehg, J. M. (2010). Temporal causality for the analysis of visual events. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1967–1974. https://doi.org/10.1109/CVPR.2010.5539871

R Core Team. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Reddy, C. K., & Vinzamuri, B. (2018). A survey of partitional and hierarchical clustering algorithms. In C. C. Aggarwal & C. K. Reddy (Eds.), *Data clustering* (pp. 87–110). Chapman and Hall/CRC.

Si, Z., Pei, M., Yao, B., & Zhu, S.-C. (2011). Unsupervised learning of event AND-OR grammar and semantics from video. *2011 International Conference on Computer Vision*, 41–48. https://doi.org/10.1109/ICCV.2011.6126223

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2023). Human Action Recognition from Various Data Modalities: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(3), 3200-3225. https://doi.org/10.1109/TPAMI.2022.3183112

Tomè, D., Russell, C., & Agapito, L.D. (2017). Lifting from the Deep: Convolutional 3D Pose Estimation from a Single Image. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5689-5698.

Turaga, P., Veeraraghavan, A., & Chellappa, R. (2009). Unsupervised view and rate invariant clustering of video sequences. *Computer Vision and Image Understanding*, *113*(3), 353–371. https://doi.org/https://doi.org/10.1016/j.cviu.2008.08.009

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

Vehlen, A., Spenthof, I., Tönsing, D., Heinrichs, M., & Domes, G. (2021). Evaluation of an eye tracking setup for studying visual attention in face-to-face conversations. *Scientific Reports*, *11*(1), 2661. https://doi.org/10.1038/s41598-021-81987-x

Wang, Y., Jiang, H., Drew, M. S., Li, Z.-N., & Mori, G. (2006). Unsupervised Discovery of Action Classes. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, *2*, 1654–1661. https://doi.org/10.1109/CVPR.2006.321

Wang, Z., Chen, H., Li, X., Liu, C., Xiong, Y., Tighe, J., & Fowlkes, C. (2022). SSCAP: Self-supervised Co-occurrence Action Parsing for Unsupervised Temporal Action Segmentation. *Proceedings - 2022 IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2022*, 175–184. https://doi.org/10.1109/WACV51458.2022.00025

Zhang, S. H., Li, R., Dong, X., Rosin, P., Cai, Z., Han, X., Yang, D., Huang, H., & Hu, S. M. (2019). Pose2Seg: Detection free human instance segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 889–898. https://doi.org/10.1109/CVPR.2019.00098

Zhang, S., Wang, C., Dong, W., & Fan, B. (2022). A Survey on Depth Ambiguity of 3D Human Pose Estimation. In *Applied Sciences* (Vol. 12, Issue 20). https://doi.org/10.3390/app122010591

Zhou, F., Torre, F. D. la, & Hodgins, J. K. (2013). Hierarchical Aligned Cluster Analysis for Temporal Clustering of Human Motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(3), 582–596. https://doi.org/10.1109/TPAMI.2012.137

Figures and Tables

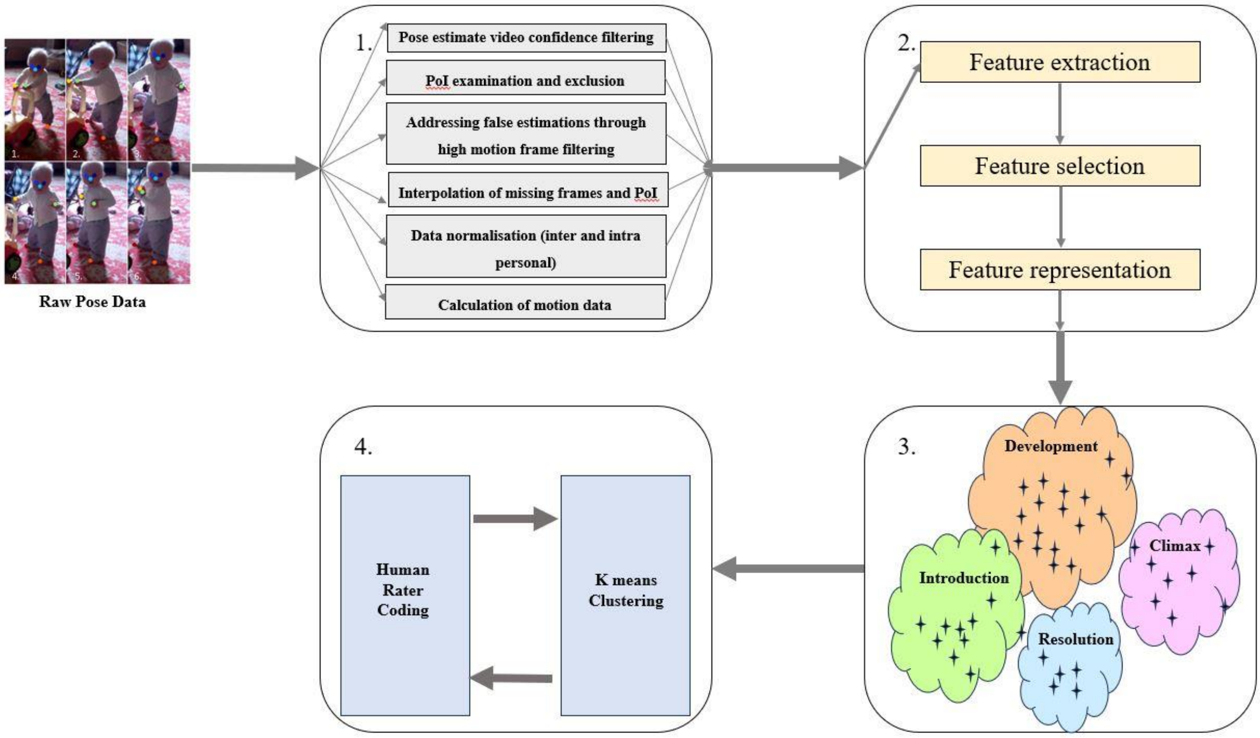


Figure 5.1. Machine learning workflow. Raw pose estimation data is **1)** pre-processed to remove unreliable videos, frames and PoI; interpolate missing data; normalise data; and finally calculate motion based on pose estimates. This processed data is then **2)** fed into the tsfeatures feature extraction package (Hyndman et al., 2023) from which the most discriminating features can be selected using approaches such as principal component analysis or mutual information. Features pertaining to the target representation can then be **3)** fed into a clustering algorithm (such as k-means) where data is clustered accordingly. The results of this unsupervised clustering can then be **4)** compared and validated against our own human coded data (see Chapter 2).

Table 5.1 Mean Confidence Scores Based on DLC Pose Estimates of PoI for Mother and Infant

Point of Interest	Mean Confidence Score
Infant	
Head (centre)	0.94
Shoulder (left)	0.74
Shoulder (right)	0.78
Elbow (left)	0.81
Elbow (right)	0.80
Wrist (left)	0.88
Wrist (right)	0.83
Hip (left)	0.46
Hip (right)	0.39
Knee (left)	0.79
Knee (right)	0.74
Ankle (left)	0.58
Ankle (right)	0.61
Adult	
Head (centre)	0.91
Shoulder (left)	0.84
Shoulder (right)	0.78
Elbow (left)	0.69
Elbow (right)	0.84
Wrist (left)	0.68
Wrist (right)	0.77
Hip (left)	0.71
Hip (right)	0.86

General Discussion

1.0 Summary of Thesis Aims

In this thesis we set out to further our understanding of the temporal architecture of adult - infant interaction and identify common underlying temporal structures within which intersubjective engagement could arise. In particular we considered whether there exists a common narrative temporal framework underpinning such engagement. This work extends our understanding of narratives in infancy (considering their form, function, development, and identification) and makes theoretical, empirical and methodologically contributions to the field. In Chapter 1 we sought to understand the current state of research and literature with regard to narrative and adult-infant interaction. In Chapter 2 we began to address a gap in infant research with regard to the application of narrative theory. We applied narrative theory to the neonatal imitation paradigm, proposing that neonatal propensity to imitate is tied to the temporal architecture of the experimental interaction. In Chapter 3 we undertook the first longitudinal mapping of narrative in mother-infant interaction, using human raters to identify periods and duration of narrative engagement, and explore the relationship between narrative and positive affect. In Chapters 4 and 5 of the thesis we aimed to advance current methodological approaches for the identification of narrative within adult-infant interactions. Considering movement as a core expression of vitality, intent and affect, we aimed to establish the optimal approach for tracking motion in footage of adult - infant interaction. We then aimed to develop a machine learning approach for the automatic identification of narrative, removing any potential for human bias in narrative recognition.

2.0 Summary and Discussion of the Work Presented

2.1 Part 1

In Part 1 of this thesis, we presented our theoretical and literature based contribution. In Chapter 1 we reviewed evidence of the infant's role in the creation of narratives and corresponding meaning-making, and the function this has as the foundation for learning patterns and their temporal organisation. Furthermore, we considered the downstream consequences of narrative co-regulation being thwarted, and explored theoretical positions that propose deficits in the creation and sharing of narratives may underlie ASD symptomatology. Perhaps most importantly though, we identified a gap in the current state of infant interaction and co-regulation research in that recent work largely fails to consider an underlying narrative dimension. Additionally, we showed that much of our current understanding of the form and function of narrative is derived from and focused upon human based research. To combat this bias, we highlighted emerging evidence of narrative patterning in chimpanzee pant hooting and proposed building on this through a new field of study - 'Comparative Narrative'.

In Chapter 2 we began to address the 'narrative gap,' identified in Chapter 1, with a theoretical contribution to the neonatal imitation debate. We considered neonatal imitation as inherently communicative in nature, but also further explored the potential role of narrative in framing the imitative exchange within which neonatal imitation unfolds. We argued neonatal imitation is more than mere reaction to external stimuli, but rather facilitates purposeful interactions with adults in the earliest stages of infancy. Such interactions may evoke affective responses, fostering a shared emotional space between infants and adults. Viewing neonatal imitation as inherently dialogical offers insights into its fundamental importance in human development, and we looked at evidence from studies from infant macaques that underscore the link between neonatal imitation and future social development. Ultimately, we proposed that successful instances of neonatal imitation are underpinned by a narrative framework, akin to adult-infant proto-conversations. This narrative architecture, expressed multimodally, requires a comprehensive analysis to identify whether a narrative framework underpins the neonatal imitation phenomenon. Below (Section 3.1) we expand on this suggestion in a more extensive manner.

2.2 Part 2

Part 2, Chapter 3 of this thesis presents our empirical contribution to the field of infant development and narrative. We proposed that the narrative structure underlying interactions develops and alters throughout early childhood. To investigate this, we analysed mothers and infants engaged in naturalistic interactions at three time points - 4 months, 7 months, and 10 months. Each interaction spanned 10 minutes, with the first 5 minutes subject to detailed analysis.

This analysis involved coding and time-stamping infant eye gaze, affective states, and the narrative phase reached during engagement within the 5-minute period. We hypothesised that the duration of positive affect would be a function of the narrative phase reached, with older infants engaging in longer interactions and reaching more advanced narrative phases. Our results broadly supported these predictions, except for interaction durations, which unexpectedly decreased with infant age. This decline might be attributed to methodological limitations (such as distraction caused by motion tracking markers) or increased interest in the wider experimental environment on the part of the infant. Despite this, our findings indicate that the narrative structure of engagements changes significantly in early infancy. These results align with the primary intersubjective thesis, suggesting that human unique sociality emerges early in infancy. Additionally, narrative cycles reaching climax and resolution phases are associated with increased positive affect, indicating the close link between narrative and affect. Somewhat surprisingly, interactions terminating at the climax phase resulted in significantly increased duration of positive infant affect similar to the resolution phase, suggesting the potential for "partial-narratives" to benefit infant development. Finally, our findings provide additional evidence that narrative formation occurs in the absence of language, so highlighting the foundational role of narrative structure in early development.

2.3 Part 3

Finally, in Part 3 of the thesis we presented our methodological contribution to the field. In Chapter 4 we proposed utilising movement and motion tracking as a core approach in the automatic identification of narrative phases. To this end we trained a deep learning pose estimation model, using open source software DeepLabCut, and successfully produced a pose estimation data set derived from the video footage that formed the basis of our empirical study in Chapter 3.

In Chapter 5 we outlined a machine learning analysis pipeline for this pose estimation data, which represents a novel approach to studying narrative in human interactions. Our motivation for this was based on weaknesses inherent to the human annotation of video data; a process that can be prohibitively time-consuming, labour intensive and prone to biases. Automatic clustering of narrative phases within interactions presented an attractive alternative, addressing these limitations. While our pipeline is tailored to the secondary data used in Chapter 2, it can be adapted for other datasets. In the chapter we outlined the pre-processing and data cleaning steps required to prepare the data for narrative clustering, the feature extraction, selection and representation steps that provide the most discriminative features, and the k-means algorithm for data clustering into which these can ultimately be fed. Despite the strengths of this analysis, our pipeline currently relies on 2D pose estimation, which is less accurate than 3D pose estimation. Utilising 3D pose estimation from multi-camera perspectives (sadly not possible using the video footage used in Chapter 2) would enhance depth and proximity information, help overcome occlusion challenges and provide richer data for narrative identification. Furthermore, expanding the pipeline to include additional modalities such as vocalization, facial expression, gaze, and touch would further enrich the analysis, capturing the multi-modal nature of adult-infant interactions. It could even be possible to integrate participant demographic and characteristic information to help contextualise model outputs and explore broader factors influencing joint narrative creation.

While our approach is a pioneering effort in applying machine learning to infant narrative analysis, its potential extends beyond this domain. It could offer insights into typical and atypical child development, including conditions like ASD and postnatal depression. This interdisciplinary field holds promise for diverse areas of research and practice.

3.0 Future Directions in Research: Building on the Theory, Findings and Methods of this Thesis

In the immediate term, we foresee this doctoral work developing across three key strands. These will examine the temporal architecture of interaction and social engagement in new populations and experimental paradigms, using state-of-the-art machine learning methodology. Each of these strands is of considerable importance in the field of early years education, with potential to deepen our knowledge of human learning, including in some of our societies most disadvantaged young people. The first strand focuses on neonatal imitation, which is of fundamental concern for understanding the nature of human communication, its evolution, and development. The second strand crosses the disciplines of primatology, developmental psychology and education, exploring the underlying temporal structure of action and communication in multiple populations (important in expanding our understanding of the roots of human communication and learning, especially in relation to the developmental pathogenesis of ASD in children). The third strand will examine the temporal structure of communication in individuals diagnosed with ASD, and how this might differ from other populations. It will build on the neurodiversity model, and consider alternative temporal structures as a difference rather than a deficit.

A key component of this future work is the methodology we explore in Part 3 of this thesis - the markerless tracking of movement using deep learning based videography, and the machine learning clustering of time series data. This method of motion tracking will be deployed across all three research strands, alongside more traditional methodologies (such as acoustic recording and analysis to ascertain the pitch, timbre and intensity of acoustic expression, and human rater video coding).

3.1 Strand 1 - Neonatal Imitation

This project will build on the theory and hypothesis work described in Chapter 2. We propose investigating the temporal nature of imitation among human neonates, and ascertain whether a narrative structure is present. Thus, this work would test the hypothesis that neonatal imitation sits within a particular temporal architecture of communication - narrative - that permits meaning to unfold between adults and infants. The work would aim to (1) map infant expressive temporal patterning through both the human coding and machine learning procedures, (2) test if this pattern follows the four phases that identifies a narrative pattern, and (3) investigate if the identified pattern relates to imitative engagement.

To achieve this end a multimodal approach that considers vocalisations, movement, body and head posture, eye gaze, and facial expression could be employed. As well as deep learning based markerless motion tracking, we would also use audio analysis software to capture the pitch, timbre, volume, and general intensity of acoustic expression within recordings. The data obtained will provide a comprehensive account of the expressive actions of the infant and adult model, mapped across time during each experimental trial. This map of the temporal structure of the imitative interaction provides the raw data from which further analysis will be conducted.

In order to test for the presence of a narrative structure within movement and audio data, both supervised and unsupervised machine learning approaches could be deployed to cluster or classify data according to pertinent features that best discriminate its temporal patterning. However, the unsupervised approach outlined in Chapter 5 could be easily tailored for the needs of this analysis. The results of this can be compared with whether imitative responses were identified from the original studies, and the relationship between temporal structure and response type analysed.

This analysis will create a unique data set that will allow for the first time a view of how neonatal imitation is organised, and will potentially add considerable weight to theories proposing neonatal imitation to be inherently interactional and perhaps the first example of primary intersubjectivity (Kugiumutzakis & Trevarthen, 2015; Nagy, 2006; Nagy & Molnar, 2004).

3.2 Strand 2 - Comparative Narrative

The exploration of comparative narrative will be conducted via two avenues of study. The first will be an analysis of the temporal architecture of naturalistic interactions between different chimpanzee pairings (adult-infant, adult-adult, close quarter grooming) as well as long distance calling (pant hooting) and buttress drumming. The second will involve the analysis of action projects within a controlled experimental paradigm.

3.2.1 Narrative as the Temporal Architecture of Naturalistic Interaction in Chimpanzees

To date, the experimental examination of narrative theory has been applied exclusively to human beings. As part of strand two, we propose the application of narrative theory into non-human primates, specifically chimpanzees.

Dautenhahn (2001) proposed the Narrative Intelligence Hypothesis (NIH) (described in detail in Chapter 1 Section 6.2), which provides a theoretical foundation upon which comparative narrative study can be conducted. We propose mapping the temporal structure of naturalistic interactions between chimpanzees in a variety of contexts. Footage of chimpanzee interaction can be analysed to address the following goals: (1) map the temporal structure of chimpanzee expressive action and vocalisation, (2) test if this pattern follows the four phases that correspond to a narrative pattern, and (3) investigate the characteristic differences and similarities between chimpanzee interactional temporal patterning, and the narrative temporal patterning identified in human beings.

Using deep learning based markerless motion tracking, acoustic analysis and human rater video coding, we will map vocalisations, movement, body and head posture, and eye gaze of chimpanzees to create a data set unique in primatology. This can be subsequently analysed using unsupervised or supervised machine learning approaches to identify common temporal patterning across interactions.

3.2.2 Narrative as Macrostructure of Movement Projects in Chimpanzees and Children with and without ASD.

In addition to naturalistic interaction, the temporal architecture of actions for a shared purpose in a controlled experimental paradigm can be investigated. Actions by human beings create a generative experience in which we are able to engage with the others and the world;

they convey vital forms of intent from the actor (Stern, 2010). These experiences, and the movements and actions which come together to build them, are structured by the spatiotemporal properties of the body and our future orientated, purposeful nature, that is drawn to find meaning through interaction with the environment (Delafield-Butt & Gangopadhyay, 2013). Small, simple action units (e.g. an extension of the arm) can be pinned together to form action chains (e.g. extending and reaching the arm to grasp an object) which can be combined to build more complex projects that extend further into the future (e.g. to make coffee or to eat dinner). Human action is therefore always organised prospectively, even from the beginning of life (Delafield-Butt et al., 2018), with narrative temporal organisation being key in bringing structure to activities and engagements. For human beings, narrative parcels our engagement with the world into embodied experiences of purposeful events. Movement projects structured and organised in this way, can be shared with another attentive participant, where the expressive movements of one partner are mirrored in the actions of the other, enabling reciprocal turn-taking in a protoconversational narrative envelope (see figure 1, Chapter 1).

An analysis of the subsecond microstructure (the precise kinematics of a movement) of joint movement projects in chimpanzees, typically developing children, and children diagnosed with ASD could be conducted to gain a better understanding of differences in movement macrostructure, and the role this plays in the development of joint projects, and the ability of individuals to repair and continue broken collaborative projects with other individuals.

To do this we propose (1) developing touch-screen cognitive paradigms that elicit goal-directed actions in both solitary and social situations for chimpanzees, and humans with and without autism; (2) utilising deep learning based markerless motion capture to quantify the micro-motor control (e.g., subsecond control of velocity, acceleration, and jerk) of participants, and (3) revealing the evolutionary and developmental trajectory of micro-kinematics and its relation to narrative structure and social cognition.

This work would open a new research field in comparative psychology with precise subsecond action kinematic analysis available with comparable, high-precision psychomotor paradigms between primate species. This will advance scientific knowledge of the role of prospective neuromotor control in the evolution and development of cognition, social cognition and social understanding, especially in relation to the developmental pathogenesis of autism spectrum disorder in children. These movement data will inform autism theory, advancing with computational precision an embodied psychological perspective that attends to the role of movement in psychological development to enable improved care and learning.

3.3 Strand 3 - Narrative and ASD

Deficits in social communication represent one of the core aspects of ASD and can have a broad impact on the lives of those diagnosed. Individuals with ASD tend to have poorer social support networks, difficulty maintaining employment, suffer from loneliness and have poorer quality of life (Bauminger & Kasari, 2000; Billstedt et al., 2005; Howlin et al., 2013; Shattuck et al., 2012). Bruner and Feldman (2003) proposed that a narrative deficit underlies the range of symptoms typical of ASD.

Strand Three will explore the Bruner and Feldman (1993) narrative deficit hypothesis of autism, which proposes that ASD develops due to difficulties in creating and partaking in narratively structured, co-operative, non-verbal meaning-making. Such a deficit has downstream impacts on communication, co-regulation, the organisation and structure of joint projects, learning, the transmission of information, and can lead to the negative life outcomes described above.

Whilst the majority of research to date has considered ASD from a deficit perspective, this project could build on the neurodiversity model and recent findings by Crompton et al (2020) to explore the temporal architecture of autistic interaction, social projects, and information transmission. Crompton and colleagues used a cultural learning paradigm to examine the transmission of information between pairs of non-autistic individuals, autistic individuals, and mixed pairs of those with and without autism. The authors found that those diagnosed with ASD could transfer information as effectively as those without ASD when the transfer was taking place within an exclusively ASD pairing, but when the pairs were mixed information transmission broke down. Positive feelings of rapport were also reported within non-mixed pairings.

Utilising the tasks and paradigm described by Crompton et al (2020), we propose applying markerless motion tracking, audio analysis and human coded microanalysis to map the temporal structure of information transmission and interaction. The aim of this would be to 1) ascertain if there is a common temporal architecture structuring information transmission and interaction amongst individuals diagnosed with ASD, and 2) identify if this temporal structure differs from those individuals without ASD. Based on the Bruner and Feldman (1993) narrative deficit of autism we would expect the temporal structure of information transmission and interaction in autistic individuals to depart from the typical narrative structure.

Understanding the temporal structure of information transmission and interaction between individuals with ASD will have important implications in education and health care. Viewing this structure as an effective alternative temporal architecture for information processing and interaction could provide a route to novel interventions in the classroom and beyond. For the autism community, understanding autistic social behaviours as a difference rather than deficit could have a significant impact on mental health, anxiety, and experiences within schooling and wider society (individuals with ASD are at a significantly increased risk of anxiety [Mattila et al., 2010] and depression [Rai et al., 2018]).

4.0 Why does any of this matter?

Narrative temporal architecture forms the foundation of human learning. It allows us to organise knowledge (Bruner, 1986), and forms the basic framework within which children can structure and make sense of their experiences (Kearney, 2002). Narrative forms the macrostucture around which young children can engage with their learning, allowing them to organise small units of learning into larger projects within a clearly delineated narrative framework (Delafield-butt & Adie, 2016). The importance of narrative in human learning, culture and communication extends throughout human life, and as evidenced in this thesis, from the first months and dialogues of infancy.

Infant involvement in narrative creation and meaning-making forms the foundational learning patterns that will be utilised in later schooling. In infancy though, they facilitate the development of complex embodied practices, the schemas of which can be held in memory (Cobley, 2013; Delafield-butt & Adie, 2016; Delafield-Butt & Trevarthen, 2015; Gratier & Trevarthen, 2008; Negayama et al., 2015). Initial motor behaviours, regulated by brainstem processes, form the basis of these schemas before higher forms of linguistic and conceptual learning can occur. Consequently, disruption in narrative shared meaning-making and co-regulation can have profound downstream effects (e.g. Bruner and Feldman, 1993), which serves to further emphasise the importance of understanding the characteristics of narrative in early infancy.

By demonstrating and beginning to address the narrative gap in infant co-regulation and interaction research, we also demonstrate the need for a renewed emphasis on this dimension of early infant scholarship. This has the potential to extend our understanding of the role of narrative in both typical and atypical development, thus increasing the potential for effective intervention and learning strategies tailored to differences in temporal organisation. This has the potential for significant impact in infant care and development whilst also influencing disability and special needs provision in education. Professionals working with pre-verbal children recognise movement and behaviour as forms of communication (Bunning et al., 2013). Viewing and understanding narrative as the underlying structure of this expression provides practitioners and parents with a framework for structured engagement (Delafield-Butt et al., 2020), aiding in the regulation of internal states and facilitating learning for these young individuals.

References

Bauminger, N., & Kasari, C. (2000). Loneliness and friendship in high-functioning children with autism. *Child Development*, *71*(2), 447–456. https://doi.org/10.1111/1467-8624.00156

Billstedt, E., Gillberg, C., & Gillberg, C. (2005). Autism after adolescence: Population-based 13- to 22-year follow-up study of 120 individuals with autism diagnosed in childhood. *Journal of Autism and Developmental Disorders*, *35*(3), 351–360. https://doi.org/10.1007/s10803-005-3302-5

Bruner, J. S. (Jerome S. (1986). Actual minds, possible worlds. Harvard University Press.

Bruner, J. S. & Feldman, C. (1993). Theories of mind and the problem of autism. In S. Baron-Cohen, H. Tager-Flusberg, D. J. Cohen (Eds.), *Understanding other Minds: Perspectives from Autism.* Oxford: Oxford University Press.

Bunning, K., Smith, C., Kennedy, P., & Greenham, C. (2013). Examination of the communication interface between students with severe to profound and multiple intellectual disability and educational staff during structured teaching sessions. *Journal of Intellectual Disability Research*, *57*(1), 39–52. https://doi.org/https://doi.org/10.1111/j.1365-2788.2011.01513.x

Cobley, P. (2013). *Narrative* (2nd ed..). New York : Routledge.

Crompton, C. J., Ropar, D., Evans-Williams, C. V. M., Flynn, E. G., & Fletcher-Watson, S. (2020). Autistic peer-to-peer information transfer is highly effective. *Autism*, *24*(7), 1704–1712. https://doi.org/10.1177/1362361320919286

Dautenhahn, K. (2001). The Narrative Intelligence Hypothesis: In Search of the Transactional Format of Narratives in Humans and Other Social Animals. In M. Beynon, C. L. Nehaniv, & K. Dautenhahn (Eds.), *Cognitive Technology: Instruments of Mind* (pp. 248–266). Springer Berlin Heidelberg.

Delafield-butt, J. T., & Adie, J. (2016). of Learning : Nurture in School. *Mind, Brain and Education*, *10*(2), 117–131.

Delafield-Butt, J. T., Freer, Y., Perkins, J., Skulina, D., Schögler, B., & Lee, D. N. (2018). Prospective organization of neonatal arm movements: A motor foundation of embodied agency, disrupted in premature birth. *Developmental Science*, *21*(6), 1–16. https://doi.org/10.1111/desc.12693

Delafield-Butt, J. T., & Gangopadhyay, N. (2013). Sensorimotor intentionality: The origins of intentionality in prospective agent action. *Developmental Review*, *33*(4), 399–425. https://doi.org/10.1016/j.dr.2013.09.001

Delafield-Butt, J. T., & Trevarthen, C. (2015). The ontogenesis of narrative: from moving to meaning. *Frontiers in Psychology*, *6*(September), 1–16. https://doi.org/10.3389/fpsyg.2015.01157

Delafield-Butt, J. T., Zeedyk, M. S., Harder, S., Vaever, M. S., & Caldwell, P. (2020). Making Meaning Together: Embodied Narratives in a Case of Severe Autism. *Psychopathology*, *53*(2), 60–73. https://doi.org/10.1159/000506648

Gratier, M., & Trevarthen, C. (2008). Musical narrative and motives for culture in mother-infant vocal interaction. *Journal of Consciousness Studies*, *15*(10–11), 122–158.

Howlin, P., Moss, P., Savage, S., & Rutter, M. (2013). Social outcomes in mid- to later adulthood among individuals diagnosed with autism and average nonverbal IQ as children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *52*(6), 572-581.e1. https://doi.org/10.1016/j.jaac.2013.02.017

Kearney, R. (2002).*On stories*. London, UK: Routledge.

Kugiumutzakis, G., & Trevarthen, C. (2015). Neonatal Imitation. In *International Encyclopedia of the Social & Behavioral Sciences: Second Edition* (pp. 481–488). https://doi.org/10.1016/B978-0-08-097086-8.23160-7

Mattila, M. L., Hurtig, T., Haapsamo, H., Jussila, K., Kuusikko-Gauffin, S., Kielinen, M., Linna, S. L., Ebeling, H., Bloigu, R., Joskitt, L., Pauls, D. L., & Moilanen, I. (2010). Comorbid psychiatric disorders associated with asperger syndrome/high-functioning autism: A community- and clinic-based study. *Journal of Autism and Developmental Disorders*, *40*(9), 1080–1093. https://doi.org/10.1007/s10803-010-0958-2

Nagy, E. (2006). From imitation to conversation: The first dialogues with human neonates. *Infant and Child Development*, *15*, 223–232. https://doi.org/10.1002/icd.460

Nagy, E., & Molnar, P. (2004). Homo imitants or homo provocans? Human imprinting model of neonatal imitations. *Infant Behavior and Development*, *27*(1), 54–63. https://doi.org/10.1016/j.infbeh.2003.06.004

Negayama, K., Delafield-Butt, J. T., Momose, K., Ishijima, K., Kawahara, N., Lux, E. J., Murphy, A., & Kaliarntas, K. (2015). Embodied intersubjective engagement in mother-infant tactile communication: A cross-cultural study of Japanese and Scottish mother-infant behaviors during infant pick-up. *Frontiers in Psychology*, *6*(FEB), 1–13. https://doi.org/10.3389/fpsyg.2015.00066

Rai, D., Heuvelman, H., Dalman, C., Culpin, I., Lundberg, M., Carpenter, P., & Magnusson, C. (2018). Association between Autism Spectrum Disorders with or without Intellectual Disability and Depression in Young Adulthood. *JAMA Network Open*, *1*(4), 1–11. https://doi.org/10.1001/jamanetworkopen.2018.1465

Shattuck, P. T., Narendorf, S. C., Cooper, B., Sterzing, P. R., Wagner, M., & Taylor, J. L. (2012). Postsecondary education and employment among youth with an autism spectrum disorder. *Pediatrics*, *129*(6), 1042–1049. https://doi.org/10.1542/peds.2011-2864

Stern, D. N. (2010). *Forms of vitality : exploring dynamic experience in psychology, the arts, psychotherapy, and development*. Oxford University Press.

Appendix A

Supplementary Materials from this Thesis

Key terms published as part of Chapter 1

Embodied narrative – Narrative is not exclusively a creation of language, nor is it only expressed through vocalization. Rather, narrative is a framework that structures human activity from its most basic level to its most advanced, both solitary and social. It structures thought and movement, cognition and consciousness. Its structure is developed from the serial organisation of purposeful body movement to give units of self and other regulation with prospective knowledge, the affective, intentional and autonomic systems of which form a deep, invariant embodied structure in which meaning is contextualised.

Co-created narrative – A narrative can be a self-generated in solitary projects, but they can also be created in conjunction with another individual. These social narratives are made with active participation from both sides, and are refered to as co-created narrative.

Narrative projects – Projects refer to the purposeful organisation of actions to achieve a task. Arranged within an envelope with the four-part structure of introduction, development, climax, and resolution, these projects fulfill the criteria of narrative.

Co-created narrative projects – Narrative projects can be created and regulated in partnership with another individual. In such cases the projects are co-created through social engagement and co-operative participation with an attentive and engaged other, from stacking building blocks in the accomplishment of a desire, to the formal discourse of speech in adult-adult dialogue.

Table S.1

Total Number of Interactions Analysed in Adult-Infant Dyads According to Age and Narrative Phase in Chapter 2

	4 months	7 months	10 months	Total
Introduction	58	46	77	208
Development	48	46	71	165
Climax	19	24	35	78
Resolution	6	10	22	38
Total	121	126	205	

Appendix B

Contributions and Collaborations Outwith the Thesis

Observing and Interpreting Embodied Interactions: Interpreting Voice from Birth to 3 Years

Lorna Arnott, **Timothy McGowan** and Jonathan Delafield-Butt

This chapter has been published in its current form:

Arnott, L., McGowan, T., & Delafield-Butt, J. (2021). Observing and interpreting embodied interactions: interpreting voice from birth to 3 years. In L. Arnott, & K. Wall (Eds.), *Research Through Play: Participatory Methods in Early Childhood* SAGE Publications Ltd.

The majority of the examples in this book support research through play with children from 3 years and older. Of course, the chapters in Parts 1 and 2 that are foundational, theoretical and conceptual are relevant to children under 3 years old but in this Part the under-3s have yet to be explored. That is not because under-3s are less important or incapable; it simply reflects the novelty in this work. It is important to recognise that children under 3 still have valuable stories to share, and therefore this chapter offers a unique contribution to how we might draw on play and key scientific approaches to analysis of observational data to support our understanding of children's experiences from birth to 3.

- We consider the place of consulting with and understanding the perspectives of children from birth to 3 years old, in a broad sense.
- The chapter is positioned within the perspective that 'voice' should not necessarily privilege verbal voice, articulating that all children also communicate in embodied ways that are non-verbal. This is especially important in pre-verbal children, and those who do not develop language.

- Linking with research in neuropsychology, we articulate how the embodied voice, especially evident in young children and babies' pre-verbal behaviour, communicates intentional and affective interests.
- From the above theoretical perspective, we demonstrate example ways to tune into children's intentional actions to interpret their meaning in research.

In this chapter we take a unique approach in framing consultation with babies and very young children. We offer an interdisciplinary dialogue, as we marry perspectives on pedagogies of listening in education (Bath, 2013), as part of the overall discussion of play in this book, with the neuropsychology of infant minds (Trevarthen and Delafield-Butt, 2013; Delafield-Butt and Trevarthen, 2013). We draw on educational knowledge and high-quality early childhood pedagogies about how best to listen to children by following their actions and interests, their 'non-verbal voice' that allows access to their minds (Reddy, 2008). This technique, we will show, is routine in knowledge from neuropsychology and paediatrics (Brazelton, 2006) where we understand that even before birth children's movements and behaviours offer an insight into the very young child's intentions and desires (Delafield-Butt and Gangopadhyay, 2013; Quintero and De Jaegher, 2020).

In Chapter 2, Murray details the rights of children to be listened to – all children regardless of age-related competences. Yet much of the work that is taking place is still happening with older children, and the knowledge of how to consult with very young children and babies is still considered challenging. This is largely because methods for consultation with children, even play-based methods, can privilege verbal voices, or means of communication which ultimately showcase children's perspectives on adults' agendas. For example, in the case described in Chapter 12 of this book, Mertala engages children in designing and building objectives related to digital play. A discrete output was sought. While this was appropriate for the more mature age group, for babies and younger children, such explicit contributions are unrealistic.

With babies and young children, it is much more challenging to start with an adult agenda and then to extract children's ideas and perspectives on matters that affect them (UN Commission on Human Rights, 1990). A shift in perspective of what we mean by listening is required to better understand babies' participation in research and design. Listening must include much more than linguistic, verbal utterances and instead take into account the infant's or non-verbal child's perspectives, understood by following their lead and reading their feelings, intentions and desires. Children communicate with one hundred languages (Malaguzzi, 1996), and in this chapter, we explore the bodily language of movement and expressive gesture. We will demonstrate through world-leading neuropsychological research that every movement made by children holds a degree of intentionality that demonstrates their motivations and feelings in expressions of body and (non-verbal) voice (Delafield-Butt and Trevarthen, 2020; Malloch et al., 2019).

Understanding that babies' movements are not arbitrary opens possibilities to engage in fine-grained observations of babies' play to understand their perspectives, which offers a meaningful route to involving even small babies in the collection of research data about their interests and feelings of participation. This requires a particular skill in how we observe babies' movements – a process of watching and re-watching is often required. Learning this technique from how psychologists conduct observations of babies offers a route to data collection that is systematic, allowing for the collection of more rigorous and articulated evidence. Findings can then be interpreted with the help of educational frameworks to successfully interpret babies' voices in early childhood practice. Furthermore, we can learn from similar lessons from clinical perspectives that attend to the mental health and well-being of infants and young children (Douglas, 2007; Zeanah, 2009). Yet before we can understand these observational techniques, we must first understand the theory behind babies' intentional actions.

What We Can Learn from Neuropsychology

Infant psychology now recognises the expressive movements and gestures of young babies as intentional from birth (Delafield-Butt and Gangopadhyay, 2013; Delafield-Butt and Trevarthen, 2020; Delafield-Butt et al., 2018). They express the feelings, interests and desires of the infant through their form and direction, soliciting attention and interest in a reciprocal dialogue with others (Stern, 2000). These patterns of interaction start to form small stories that begin, develop, and very often reach a point of maximal tension in a climax, before receding to a quiet state again and concluding (Delafield-Butt and Trevarthen, 2015). Altogether, these projects are experienced as a story, or narrative event, and their experience is held in memory.

Infant narratives are made through experience and shared experience with others. They are pre-verbal and embodied, rather than being linguistic and structured by ideas and concepts, as they are in adult storytelling (Delafield-Butt, 2018). Rather, infant stories are directly experienced in body action and interaction. But despite the fact these are not told in the language that adult stories are, they hold the same temporal pattern and four-part structure of (i) introduction, (ii) development, (iii) climax, and (iv) resolution (Figure 9.1). This is a structure of human narrative meaning-making invariant across the lifespan, and it is used to share the drama of life events from birth, right though to technical sharing of ideas in adult industry (Bruner, 1987; Bruner, 1990). Narratives structure the way we experience the world and remember it in shared stories with others, or as we reflect and make plans for the future. And its lived, shared experiences can form a bedrock of meaning-making in school (Delafield-Butt and Adie, 2016).

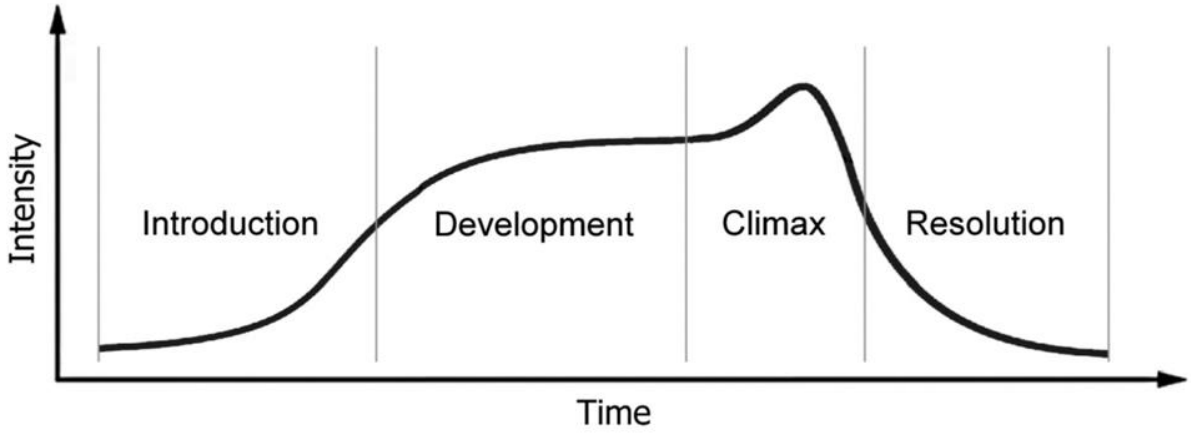


Figure 9.1 The intensity contour of a narrative over its four phases

Source: Trevarthen, C. and Delafield-Butt, J. T. (2013)

(i) ‘interest’ in the narrative begins at a low intensity in the introduction, which ‘invites’ participation in purposefulness; (ii) the coordination of the actions and interests of real and imagined agents intensifies over the development, as the ‘plan’ or ‘project’ is developed; (iii) a peak of excitation with achievement of a goal in mutual intention is reached at the climax; after which (iv) the intensity reduces as the purposes of the participants share a resolution, and those who were closely engaged separate (Trevarthen and Delafield-Butt, 2013).

Even the pre-verbal speech of newborn infants made in proto-conversation – vocal and movement-based interactions – communicates the emotions, intentions, and interests of the infant, and is important for professional understanding (Trevarthen and Delafield-Butt, 2015). A musicality of shared tempo, reciprocal to-and-fro of timbre, pitch, and intensity of vocalisation altogether communicate and share reciprocal feelings and interest, and the mutual pleasure in sensitive care and support (Malloch, 1999 Malloch and Trevarthen, 2009). Interestingly, these vocal interactions follow the rules of jazz music: partners can withhold their contribution and wait for a split second, increasing dramatic tension in the expectant partner with uncertainty. Or a partner can interrupt early, disrupting the flow and creating new patterns of interaction (Gratier and Apter-Danon, 2009). What is clear from these studies is that infants and their adult partners follow the same formal patterns of engagement, sharing feelings and interest, and structuring their discourse in shared time with each other as we do as adults. And this sharing of time and experience is deeply rooted in our common neurobiological rhythms and expectations (Trevarthen, 2009; Trevarthen et al., 2014). This 'communicative musicality' that exists between individuals allows humans (including infants) to share time meaningfully, and collaborate not only in interaction but also in the completion of tasks (Malloch and Trevarthen, 2009). It is important to note that the definition of 'musicality' here differs somewhat from what would be generally understood by the term – it refers to the basic and foundational human ability for the production and appreciation of music (Blacking, 1969/1995), alongside other temporal endeavours such as dance or theatre. Malloch and Trevarthen (2009) link musicality with the attraction to cultural learning exhibited by humans, and the affinity and appreciation we have for moving in coordination with others (particularly within the narrative structures found in the temporal arts of music, poetry, dance and drama). Narrative structuring forms a key component of communicative musicality, bringing structure to interpersonal interactions that serves to frame knowledge and enhance the feeling within social relationships and individual experience.

The earliest form of proto-conversations, or the 'first dialogue', is evident in the imitative interactions of the neonate (newborn baby) with an adult partner (Nagy and Molnar, 2004). Neonates are sensitivite to the actions and intentions of another human being, and may imitatate them from as early as the first hours after birth (Trevarthen, 2001). Imitation is one of multiple ways neonates can express themselves to join a shared, social behavioural repertoire (Trevarthen, 1979). The importance of these imitative interactions in building attachment and developing cultural intelligence has been supported through recent findings in neuroscience (Ammaniti and Gallese, 2014). Adults can use imitation as an effective method of engagement with infants and children yet to develop language skills. Its function runs deeper than being purely reactionary, it enables engagement for meaningful and purposeful interaction without language for infants and children of all ages, and it can be especially important in reaching hard-to-reach children averse to social engagement, such as those with autism spectrum disorder (Delafield-Butt et al., 2020; Nadel, 2014). Neonatal imitation appears to be a foundation stone for primary intersubjectivity (a shared understanding), which goes on to develop into the full blown proto-conversations of later infancy. It is a useful tool for professionals.

Children learn and explore the world in narrative cycles of interaction all of the time, whether in solitary play in engagement with objects or in social, intersubjective interaction with others. These narrative ‘projects’ underpin learning and structure communication with others. Recognising these moments of learning and communication allow researchers and practitioners to make sense of the experiences of the infant, recognising their body actions of interest or disinterest, aroused excitement or under-aroused weariness, and expressive of inner feelings of confidence or uncertainty, or pride and shame in shared performance with others (Trevarthen and Delafield-Butt, 2017). Tuning in to these embodied feelings expressed in short narratives allows the researcher or practitioner to recognise those movements as purposeful engagements that give access to the child’s perspective, and that are displays – in body actions and vocalisations – of the infant’s voice.

Methods of Observation: Translating Babies’ Communications

Once tuned into the child’s attempts at embodied communication, the practitioner researcher can find ways to observe and follow the child’s interests and feeling to learn the story the child is creating as he or she moves step-by-step, movement-by-movement into his or her self-generated future. In this way, the infant creates his or her own world, constructing its experience and structuring it in body action that expresses his or her feelings of interest and care. The practitioner researcher must watch and observe sensitively to put one's own 'eye' in the perspective of the child. What is the infant or child attending to and experiencing now? What feelings accompany these interests and actions? Is the infant or child seeking engagement with an object, or a person of interest? Or is he or she seeking to disengage and withdraw? Why and how is he or she doing this? Through simple observation, the observer can place his or her mind into the mind of the infant or child. And the interests and feelings of the infant or child will resonate with the observer, because we have evolved as a social species to intuit the feelings and intentions of another without needing to think reflectively (mentalise) about them (Ammaniti and Gallese, 2014). By doing so, one can gain insight into the lived experiences of the young child to understand those experiences, understand where they are enjoyed and successful in their engagement with the world, and where they require support and assistance to help the infant realise his or her intentions.

In this section we demonstrate examples of what these observations may look like when conducted in a systematic way. We present a case study of a baby's attempts to communicate and how this observation is framed. After each example, we re-draw your attention to an educational perspective by analysing the case studies using key educational analytical tools.

Case Study

Example 1: Accomplishing a Task, and Sharing the Joy of Success

This example is one narrative episode taken from a chain of narratively structured projects in which a toddler walks with the support of a toy push-truck, as her father watches, films and encourages. The narrative we are interested in here follows immediately from the conclusion of another. The toddler's attention and gaze are firmly focused on her supporting partner (her father), who has been making positive vocalisations in joint celebration with her previous accomplishment. She holds her hands together in a moment of apparent contemplation lasting 1.5 seconds, closing the previous narrative with a resolution phase. Her gaze and attention remain focused on her father. Immediately following this brief pause, the toddler's attention returns to her toy truck. Her gaze shifts from her father to the handle bars of the truck and she moves her hands so as to grasp them – therefore beginning a new narrative and bringing us into its introduction phase. As soon as her hands have grasped the handle bars the toddler's left leg makes a step forward, her gaze and focus entirely on her hands and the truck. She moves forward, placing one foot in front of the other, her gaze remaining fixed on her hands and the truck as she travels forward. Continuing to travel forward her gaze shifts to her father, and we enter the development phase of the project. Her hands remain firmly placed on the toy but her stance and weight distribution shift awkwardly as she navigates the multiple layers of this action chain. Her gaze shifts to the direction of travel as she continues to move towards her goal, her velocity increasing as the narrative progresses. Briefly, the toddler's gaze shifts to her table (containing other colourful items of interest) but she continues to travel forward. At this point, approximately 4 seconds since the beginning of the narrative, peak velocity is reached and the toddler begins to slow. The development stage of this narrative project is ending and the climatic segment is beginning. The toddler comes to a standstill, and returns her focus and gaze to her father. She releases the truck, first with her left hand, and then (after a short delay) with her right hand. As she does so, she redistributes her weight in order to maintain balance and then vocalises triumphantly, her gaze all the while remaining fixed on her supporting partner. She has successfully completed her goal, and the climax of the project and narrative has been reached. The toddler applauds herself for her accomplishment whilst maintaining eye contact with her father, encouraging him to join her in celebrating the achievement (this he does as he vocalises positively, congratulating his daughter). As the narrative begins to enter the resolution stage, the toddler stops applauding but continues to take pleasure in the completion of her goal and pride in sharing these with her caring parent. In the final acts of the project the toddler disengages from her father, shifting her gaze back to her toy and her hands. She vocalises again to herself, gesturing and whirling with her hands as she does so. With this the narrative is concluded and the toddler starts the process of beginning a new project.

Figure 9.2 A toddler performing a short narrative project, walking with the help of a push-truck (Delafield-Butt, 2018).

The project (a) initiates with attention to the task and (b–d) develops over repeated steps, one in front of the other, maintaining difficult shifts in balance with support from the truck until (e–g) final conclusion, accomplishment shared with delight and applause with her father filming.

Structuring Observations and Analysis

What we have described above, from a neuropsychology perspective, is an observation of a young child's play and learning (about how to move her body, and value moments of accomplishment with others) shared in companionship. The narrative perspective adds structure to that observation and helps unpick the meaning associated with fine-grained movements and expressions. What is really interesting when applying this perspective – that babies' movements have meaning and intentions – in an educational setting is that these fine-grained observations of children's play highlight what skilled practitioners do as part of their daily professional work with children. Detailed observation can give explicit framing of the actions and interactions in everyday settings, and taking the time to observe young children's movements and infer meaning from them adds insight and value to that work (Palaiologou, 2019). These observations form the basis for key pedagogical documentation (Fleet et al., 2017) and records of children's achievements through approaches like Learning Stories (Carr and Lee, 2019).

What we are suggesting here is that everyday observation of play and babies' and toddlers' experiences can be refined to produce data for research projects or to inform practice and to offer a unique insight into babies' perspectives. In order to do so in a way that builds on the neuropsychological knowledge of babies' intentional movements (Delafield-Butt and Trevarthen, 2020), you impose some structure to the analysis process, whether that be the focus on moment-by-moment interpretations of the contour of a narrative (See Figure 9.1) or something broader like Broadhead's Social Play Continuum, where children's reciprocal actions and behaviours are mapped out over multiple episodes of time to interpret the social nature of the play (Broadhead, 2006). Or indeed, any other framework which helps structure how you interpret meaning from data, such as Laevers' well-being scale, as is discussed later in this chapter. This does not necessarily mean the observation changes. If using video to support the analysis process, for example, the play episode remains the focus, but the interpretation of that observation changes depending upon the structure imposed. The case study above in Figure 9.2 could be analysed with a different frame and the insight would be different.

This process of imposing some structure into the observation analysis offers a supporting tool or guide to help you draw conclusions. For example, in the following snippet of analysis (see Figure 9.3), a 7-week-old baby interacts with her mother, practising her first smile. Her mother playfully interacts with her making different sounds to maintain the baby's engagement, focus and to encourage a smile in response. The play interaction does not need to change to generate research data, instead the episode only becomes data in the way that the observation is analysed and interpreted, and this is facilitated by video footage to allow the baby's experience to remain as naturalistic as possible.

In the example below the process is supported by identifying 'critical moments' in the child's playful experience; moments that represent a turning point in the play or are of some significance to the narrative. In relation to a teacher context, Coultras (2015: 32) describes these as 'the moments [that] helped them [teacher] to learn something'. The process builds on traditional approaches from disciplines like Industrial Psychology where the Critical Incident Technique was employed (Corbally, 1956) and was later re-imagined for research in educational contexts. As it became embedded in education the identification of these critical moments offered a starting point for deep self-reflection and was often a precursor for a process of change (Labercane et al., 2006).

In the example below, we draw on this notion of moments of significance and turning points in the play to instigate that enhanced self-reflection. At each point that Rosie appears to shift in her expression and movement, we are afforded the opportunity to interpret the narrative of her story and infer meaning. In this case, the reflection is guided by the previous knowledge of children's narratives of communication, set out in the previous section, and is contextualised by 'storying' the video screen shots to provide necessary richness to the example.

Critical Moment Sequence in Playful Interaction	Video Frame and Narrative Description	Contour of Narrative
1	 Mum begins talking to Rosie, making a silly noise 'lalalalalalala'. Rosie notices the sound and sits for a moment of silence.	Introduction Development
2	 Rosie responds with a smile and mum cheers 'yey, that's a lovely smile, is that funny, mummy doing that noise?'	Climax
3	 Mum attempts to draw more smiles. Rosie sits quietly looking.	Resolution
4	 Rosie becomes frustrated. Her body tenses and she pushes back. Mum asks what's wrong. She says 'oh, ready?'	Distraction
6	 Rosie's attention is diverted and she spots the camera mum is holding. She stares for a few seconds. Mum explains it's a camera.	
7	 Mum draws Rosie's attention back 'look at mummy'. 'Look, shall we make funny noises?'	Introduction
8	 Mum makes the same funny noise 'lalalalalala'. Rosie looks and begins to contemplate the action. Half smiles. Mum encourages 'yey, that's a nice cute smile.	Development
9	 Mum makes the silly noise 'lalalala'. Rosie smiles a big smile. Mum cheers, 'yeah that's a cute smile'.	Climax

Figure 9.3 Rosie's smile and contour of narrative

In this example, the structure imposed to help analyse the playful experience is facilitated by three components:

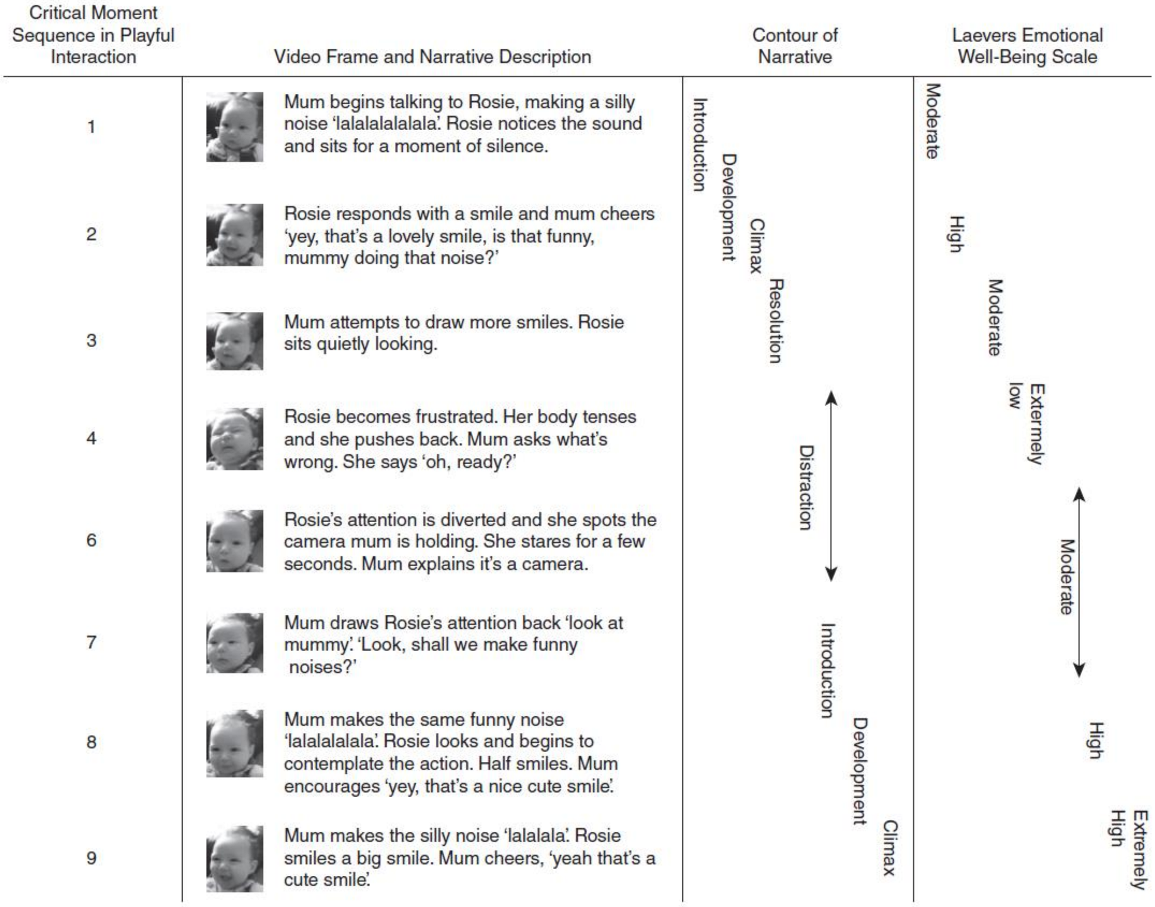
1. The recognition of critical moments in the play where specific elements of the baby's expression and embodied movements have significance and tell another snippet of the child's story.
2. A focus on narrative descriptions of the play experience, to contextualise the data.
3. The neuropsychological model of babies' communication processes, laid out in the previous section, helps understand the babies' attempts at communication.

Drawing conclusions from this simple and very short exploration of data (the video footage for this analysis lasted 1.08 minutes), allows us to unpick Rosie's perspective. You can see:

- The cycles of interaction emerging and the period of time that Rosie can remain focused and engaged. Her desire to interact, or not, becomes clear. We can also see how that process of re-introducing an interaction again after a period of distraction can extend the play.
- Instances of developmental milestones like first smiles or responding to her name.
- Moments of discomfort and frustration from Rosie's perspective.
- Moments of joy and her ability to express that joy.

In this example, we take what first appears to be rather complex interpretations of babies' communication, set out in neuropsychology and described in the previous sections, and applied the systematic process of observation to a simple and short extract about a playful encounter. Although the theory about babies' communication may seem complex, when applied with an educational lens, we show that the process is important, relevant and applicable in early childhood settings for practitioner research. We then begin to understand that neuropsychology gives us a grounding to govern our analysis in rigorous ways and with the confidence that we are understanding the child's story.

Such analysis could be extended with alternative frameworks to help support how you interrogate the data and draw meaning. For example, you could re-analyse the data from the perspective of key pedagogical tools, such as Ferre Laevers' emotional well-being scale (2017). Laevers' process may be better suited to longitudinal interpretations of a particular child, but nevertheless the various points on the scale can still help structure interpretations of a snapshot in time. In doing so you gain a new interpretation of the data. You can even combine two approaches to analyse across frameworks to see similarities or differences (Figure 9.4).



Critical Moment Sequence in Playful Interaction	Video Frame and Narrative Description	Contour of Narrative	Laevers Emotional Well-Being Scale
1	Mum begins talking to Rosie, making a silly noise 'lalalalalalala'. Rosie notices the sound and sits for a moment of silence.	Introduction Development	Moderate
2	Rosie responds with a smile and mum cheers 'yey, that's a lovely smile, is that funny, mummy doing that noise?'	Climax Resolution	High
3	Mum attempts to draw more smiles. Rosie sits quietly looking.		Moderate
4	Rosie becomes frustrated. Her body tenses and she pushes back. Mum asks what's wrong. She says 'oh, ready?'	Distraction	Extermely low
6	Rosie's attention is diverted and she spots the camera mum is holding. She stares for a few seconds. Mum explains it's a camera.		Moderate
7	Mum draws Rosie's attention back 'look at mummy'. 'Look, shall we make funny noises?'	Introduction	
8	Mum makes the same funny noise 'lalalalalala'. Rosie looks and begins to contemplate the action. Half smiles. Mum encourages 'yey, that's a nice cute smile'.	Development	High
9	Mum makes the silly noise 'lalalala'. Rosie smiles a big smile. Mum cheers, 'yeah that's a cute smile'.	Climax	Extremely High

Figure 9.4 Extended analysis

What was key in this episode, when comparing the neuropsychological framework to the Laevers' well-being scale (2017), was that it became clear that the two frameworks somewhat align. The contours of the narrative of communication tied in with varying levels of emotional well-being. As you see the conversation reaching climax, so too is the child's emotional well-being according to Laever's scale. This drops as the communication narrative is interrupted by a period of distraction, and then increases again as the communication narrative is re-introduced.

By adopting this approach you make reasoned judgements and choices about what to analyse and from what perspective. The episode can be interpreted differently depending upon your conceptual or theoretical stance. This process of adopting a particular stance was described in Part 2 and is also reaffirmed in Chapter 11 where Marilyn Fleer demonstrates the influence of a Vygotskian perspective in shaping her findings. As Nutbrown points out in Chapter 6, high-quality research makes this process of analysis explicit. Thus, we are not suggesting here that this is the only way to understand babies' voices in practice, but rather that for us drawing on the neuropsychological knowledge of children's intentional movements provides us with a framework to guide our interpretation and offers us a degree of confidence in the findings.

Ethical Considerations

One practical consideration which must be deliberated in this chapter is the complexity of ensuring assent or consent with babies to participate in research. As the UNCRC (UN Commission on Human Rights, 1990) continues to guide the Children's Right's movement, and countries likes Scotland begin to enshrine the convention in Scottish Law, now more than ever we must make an attempt to be respectful of children's rights to participate (or not) in research (see Chapters 2, 3 and 7 for more on this). While more challenging when working with babies, it should not be disregarded as impossible. Palaiologou, for example, has developed an Ethical Helix as a set of reflective questions which researchers and practitioners may ask themselves when conducting research with babies to better evaluate the child's desire to assent to, or dissent from participation (Arnott et al., 2020; Palaiologou, 2015). Consent or assent should not privilege verbal voice because we know when considering children's participation that 'training in "listening" skills should take appropriate account of the range of non-verbal "cues" which children deploy when expressing themselves' (Lundy, 2007: 937). What we have described in this chapter is that one form of 'listening' to children requires attuning to their experiences and listening to their interests and feelings by attending to their embodied forms of non-verbal communication, and the structure of those actions through time in narratives of solitary play or shared engagement. Thus, the same principle of interpretation and inference of non-verbal expression for understanding communication can be extended to issues of consent with children under 3 years old and those with developmental delay or speech and language disorder.

While we do not have space here to debate this process in full, we advocate the need to make ethical choices, engage in iterative critical reflection and attempt to interpret babies' expressions in considerate ways when involving them in research.

Conclusion

This chapter combines knowledge from neuropsychology about children's communication capabilities with knowledge of pedagogies when working with babies/toddlers to arrive at some practical examples of how children's communication techniques can be utilised in research data to help understand the child's perspective. The chapter extends new knowledge that babies' actions are intentional from birth, and so research data at this stage doesn't need to be overly complex but instead should find ways to document and observe children's patterns of communication to understand meaning.

Importantly, this chapter demonstrates techniques which relate to pedagogy:

- For practitioners – you already listen to children, but this chapter articulates how this information can be translated to serve research or develop practice.
- For researchers – it documents the need to look at pedagogy and children's ways of communicating to pinpoint what to focus on in your data collection.

By amalgamating interdisciplinary perspectives from neuropsychology with pedagogies relating to children’s play and listening in early childhood education, this chapter offers a useful approach to help solidify what observations of embodied actions, engagements and interactions may look like in practice and in research in order to articulate infant, young children’s and non-verbal children’s perspectives.

References

Ammaniti, M. and Gallese, V. (2014). *The Birth of Intersubjectivity: Psychodynamics, Neurobiology, and the Self*. New York: Norton.

Arnott, L., Martinez-Lejarreta, L., Wall, K., Blaisdell, C. and Palaiologou, I. (2020). Reflecting on three creative approaches to informed consent with children under six. *British Educational Research Journal*, 46 (4): 786–810.

Bath, C. (2013). Conceptualising listening to young children as an ethic of care in early childhood education and care. *Children & Society*, 27 (5): 361–71.

Blacking, J. (1969/1995). The value of music in human experience. The 1969 Yearbook of the International Folk Music Council. Chicago, IL: University of Chicago Press. (Republished as Chapter 1: Expressing human experience through music. In P. Bohlman and B. Nettl (Eds.) (1995), Music, Culture and Experience: Selected Papers of John Blacking. Chicago, IL: University of Chicago Press.

Brazelton, T. B. (2006). *Touchpoints: Birth to Three*. Cambridge, MA: Da Capo Press.

Broadhead, P. (2006). Developing an understanding of young children's learning through play: The place of observation, interaction and reflection. *British Journal of Educational Research*, 32 (2): 191–207.

Bruner, J. S. (1987). *Actual Minds, Possible Worlds*. Cambridge, MA: Harvard University Press.

Bruner, J. S. (1990). *Acts of Meaning*. Cambridge, MA: Harvard University Press.

Carr, M. and Lee, W. (2019). *Learning Stories in Practice*. London: Sage.

Corbally Jr, J. E. (1956). The critical incident technique and educational research. Educational Research Bulletin, 35 (3): 57–62.

Coultas, V. (2015) Case studies of teachers' understandings of the pedagogy of classroom talk: some critical moments explored, Literacy, 50: 32–39. DOI: 10.1111/lit.12065.

Delafield-Butt, J. (2018). The emotional and embodied nature of human understanding: Sharing narratives of meaning. In C. Trevarthen, J. Delafield-Butt and A.-W. Dunlop (Eds.), *The Child's Curriculum: Working with the Natural Voices of Young Children*. Oxford: Oxford University Press.

Delafield-Butt, J. and Adie, J. (2016). The embodied narrative nature of learning: Nurture in school. *Mind Brain & Education*, 10 (2): 14.

Delafield-Butt, J. T. and Gangopadhyay, N. (2013). Sensorimotor intentionality: The origins of intentionality in prospective agent action. *Developmental Review*, 33 (4): 399–425.

Delafield-Butt, J. T. and Trevarthen, C. (2013). Theories of the development of human communication. *Theories and Models of Communication*. DOI: https://doi.org/10.1515/9783110240450.199.

Delafield-Butt, J. T. and Trevarthen, C. (2015). The ontogenesis of narrative: From moving to meaning. *Frontiers in Psychology*. DOI: https://doi.org/10.3389/fpsyg.2015.01157.

Delafield-Butt, J. and Trevarthen, C. (2020). Infant Intentions: The role of agency in learning with affectionate companions. In M. Peter (Ed.), *Encyclopedia of Teacher Education*. Singapore: Springer Nature.

Delafield-Butt, J. T., Freer, Y., Perkins, J., Skulina, D., Schögler, B. and Lee, D. N. (2018). Prospective organization of neonatal arm movements: A motor foundation of embodied agency, disrupted in premature birth. *Developmental Science*, 21 (6): e12693.

Delafield-Butt, J. T., Zeedyk, M. S., Harder, S., Vaever, M. S. and Caldwell, P. (2020). Making meaning together: Embodied narratives in a case of severe autism. *Psychopathology,* 53 (2): 60–73.

Douglas, H. (2007). *Containment and Reciprocity: Integrating Psychoanalytic Theory and Child Development Research for Work with Children*. London and New York: Routledge.

Fleet, A., Patterson, C., Robertson, J. and Robertson, J. (2017). *Pedagogical Documentation in Early Years Practice: Seeing through Multiple Perspectives*. London: Sage.

Gratier, M. and Apter-Danon, G. (2009). The musicality of belonging: Repitition and variation in mother-infant vocal interaction. In S. Malloch and C. Trevarthen (Eds.), *Communicative Musicality*. Oxford: Oxford University Press.

Laevers, F. (2017). How are children doing in ECEC? Monitoring quality within a process-oriented approach. In German Youth Institute (Ed.), *Monitoring Quality in Early Childhood Education and Care*. pp. 178–200. Available at: www.dji.de/fileadmin/user_upload/bibs2017/Monitoring_Sammelband_E_final.pdf (accessed 15 November 2020).

Lundy, L. (2007). ‘Voice’ is not enough: Conceptualising Article 12 of the United Nations Convention on the Rights of the Child. *British Educational Research Journal,* 33 (6): 927–42.

Malaguzzi, L. (1996). *The Hundred Languages of Children: The Reggio Emilia Approach to Early Childhood Education*. Norwood, NJ: Ablex Publishing Corporation.

Malloch, S. (1999). Mothers and infants and communicative musicality. *Musicae Scientiae,* Special Issue Rhythms, Musical Narrative, and the Origins of Human Communication, 29-57.

Malloch, S., & Trevarthen, C. (2009). Musicality: Communicating the vitality and interests of life. In S. Malloch & C. Trevarthen (Eds.), *Communicative Musicality: Exploring the basis of human companionship* (pp. 1-12). Oxford: Oxford University Press.

Malloch, S., Delafield-Butt, J. and Trevarthen, C. (2019). Embodied musicality of infant intersubjectivity in learning and teaching. In M. A. Peters (Ed.), *Encyclopedia of Teacher Education*. Singapore: Springer Singapore. pp. 1–5.

Nadel, J. (2014). *How Imitation Boosts Development in Infancy and Autism Spectrum Disorder*. Oxford: Oxford University Press.

Nagy, E. and Molnar, P. (2004). Homo imitants or homo provocans? Human imprinting model of neonatal imitations. *Infant Behavior and Development*, 27 (1): 54–63.

Palaiologou, I. (2015). Ethical issues associated with educational research. In I. Palaiologou, M. Needham and T. Male (Eds.), *Doing Research in Education: Theory and Practice*. London: Sage.

Palaiologou, I. (2019). *Child-Observation: A Guide for Students of Early Childhood*. Sage: London.

Quintero, A. M. and De Jaegher, H. (2020). Pregnant agencies: Movement and participation in maternal–fetal interactions. *Frontiers in Psychology*. DOI: https://doi.org/10.3389/fpsyg.2020.01977.

Reddy, V. (2008). *How Infants Know Minds*. Cambridge, MA: Harvard University Press.

Stern, D. N. (2000). *The interpersonal world of the infant : a view from psychoanalysis and developmental psychology / Daniel N. Stern.* (1st pbk. ed.). Basic Books.

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before Speech: The Beginning of Human Communication*. London: Cambridge University Press. pp. 321–47.

Trevarthen, C. (2001). The neurobiology of early communication: Intersubjective regulations in human brain development. In A. F. Kalverboer and A. Gramsbergen (Eds.), *Handbook on Brain and Behavior in Human Development* (pp. 841–82). Dordrecht: Kluwer.

Trevarthen, C. (2009). Human biochronology: On the source and functions of 'musicality'. In R. Haas and V. Brandes (Eds.), *Music that Works*. Vienna: Springer. pp. 221–65.

Trevarthen, C. and Delafield-Butt, J. T. (2013). Biology of shared meaning and language development: Regulating the life of narratives. In M. Legerstee, D. Haley and M. Bornstein (Eds.), *The Infant Mind: Origins of the Social Brain*. New York: Guildford Press. pp. 167–99.

Trevarthen, C. and Delafield-Butt, J. T. (2015). The infant's creative vitality, in projects of self-discovery and shared meaning: How they anticipate school, and make it fruitful. In S. Robson and S. F. Quinn (Eds.), *International Handbook of Young Children's Thinking and Understanding*. Abingdon and New York: Routledge. pp. 3–18.

Trevarthen, C. and Delafield-Butt, J. T. (2017). Intersubjectivity in the imagination and feelings of the infant: Implications for education in the early years. In E. J. White and C. Dalli (Eds.), *Under-Three Year Olds in Policy and Practice*. New York: Springer. pp. 17–39.

Trevarthen, C., Gratier, M. and Osborne, N. (2014). The human nature of culture and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5 (2): 173–92.

UN Commission on Human Rights (1990). *Convention on the Rights of the Child* (E/CN.4/RES/1990/74), 7 March. Available at: www.refworld.org/docid/3b00f03d30.html (accessed 5 September 2020).

Zeanah, C. H. (2009). *Handbook of Infant Mental Health* (3rd edn). New York: The Guilford Press.

Appendix C

Contributions and Collaborations Outwith the Thesis

Motor Signature Differences Between Autism Spectrum Disorder and Developmental Coordination Disorder, and Their Neural Mechanisms

Christiana Butera, Jonathan Delafield-Butt, Szu-Ching Lu, Krzysztof Sobota,
Timothy McGowan, Laura Harrison, Emily Kilroy, Aditya Jayashankar, Lisa Aziz-Zadeh

This chapter has been published in its current form:

Butera, C., Delafield-Butt, J., Lu, S.-C., Sobota, K., McGowan, T., Harrison, L., Kilroy, E., Jayashankar, A., & Aziz-Zadeh, L. (2023). Motor signature differences between autism spectrum disorder and developmental coordination disorder, and their neural mechanisms. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-023-06171-8

Abstract

Here we use machine learning to classify autism spectrum disorder (ASD) from developmental coordination disorder (DCD) and typically developing (TD) children using kinematic features obtained during a 5-minute coloring game on a smart tablet (ages 8-17; 18 ASD, 16 DCD, 20 TD). A subset of participants (16 ASD, 15 DCD, 19 TD) additionally completed functional magnetic resonance imaging (fMRI) during action execution and imitation tasks. While more traditional motor assessments could not significantly differentiate the two clinical groups, machine learning analysis provided predictive discrimination between: TD and ASD (76% accuracy), TD and DCD (78% accuracy), and ASD and DCD

(71% accuracy). Two kinematic markers which strongly drove categorization were correlated with neural activity in cerebellar regions. Taken together, machine learning on kinematic data during a tablet-based coloring game significantly differentiated ASD from other similar motor developmental disorders and differences may be related to cerebellar function.

Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by restricted or repetitive behaviors and interests, and impairments in social communication and social interaction (American Psychiatric Association, 2013). Although not included in the diagnostic criteria, about 80% of individuals with ASD present with noticeable differences in motor functioning as assessed by current instrumentation (Bhat et al., 2011; Chukoskie et al., 2013; Edwards, 2014; Green et al., 2009; Hilton et al., 2012; Williams et al., 2004, Kilroy et al., 2022). Motor differences have been observed in the neonatal period, suggesting prenatal neurodevelopmental origins (Delafield-Butt & Trevarthen, 2017; Trevarthen & Delafield-Butt, 2013; Teitelbaum et al., 1998). Similarly, individuals with Developmental Coordination Disorder (DCD) experience impairments in fine motor skill, gross motor skill, dexterity, limb speed, and visual-motor integration (Blank et al., 2019; Chang & Yu, 2016; Kaplan et al., 2006, Kilroy et al., 2022) that present in early development. Unlike ASD, a diagnosis of DCD is not defined by impairments in social functioning, although some secondary social differences may occur as a result of reduced opportunities to engage in sports teams or other social activities (Cermak & May-Benson, 2019). Both children with ASD and DCD display motor disruption in basic and postural motor control, and purposeful movement (Dewey et al., 2007; Mostofsky et al., 2006; Paquet et al., 2019; Roley et al., 2015). The early presentation of motor disturbances in both populations suggests an opportunity for the development of early identification tools long before other more perceivable behavioral symptoms arise in ASD.

While assessments of motor skills can be extremely valuable for identifying individuals at high risk for motor delays, more subtle but significant aspects of coordination and timing may be harder to capture (Campbell & Hedeker, 2001). Assessments of motor skills are frequently performed using evaluation measures that are sensitive to human error, time consuming, and measure duration, frequency, and speed of actions involving fine motor, gross motor, and balance tasks (Harris et al., 2015). Other methods, like optical motion capture, can demonstrate disruptions to more subtle kinematic differences, but are expensive, time consuming, and require technical expertise. To mitigate these problems, there has been enthusiasm for utilizing motor game play coupled with machine learning to contribute to diagnosis and understanding of motor dysfunction. Previous literature has shown that it is possible to use machine learning on data from smart tablet motor games or wearable devices to significantly distinguish ASD from TD (e.g. Anzulewicz et al., 2016; Tunçgenç et al., 2020). However, until these digital serious game assessments can provide differential identification between similar childhood disorders (e.g., ASD and DCD), their precise contribution to psychological insight into differences in neurodevelopmental disorders will be limited, as will their potential clinical impact. Therefore, here we aim to identify differences between subtle kinematic motor markers when comparing ASD to DCD using a smart tablet game. Additionally, we use functional Magnetic Resonance Imaging (fMRI) tasks to elucidate underlying neural mechanisms of these kinematic differences in the two clinical groups.

ASD and DCD kinematics

In ASD, motor coordination deficits are pervasive (~80%, for review, see Fournier et al., 2010), and specific impairments are seen in kinematics for prospective goal-directed movement (Cavallo et al., 2021; Chua et al., 2022; Trevarthen & Delafield-Butt, 2013). Motor differences in ASD are demonstrated in horizontal arm movements (Cook et al., 2013), touch screen use (Anzulewicz et al., 2016; Dowd et al., 2012; Lu et al., 2022), handwriting (Kushki et al., 2011), postural changes for bimanual load-lifting (Schmitz et al., 2003), reaching and grasping (Sacrey et al., 2014; Stoit, et al., 2013), gait (Rinehart et al., 2006), as well as ball skills, balance, and manual dexterity (Craig et al. 2018). Synthesis of findings suggest that general sensorimotor integration for the prospective organization of movement is disrupted, and predictive feedforward and feedback mechanisms are consistently impaired (Chua et al., 2021; David et al., 2012; Gowen & Hamilton, 2013; Sinha et al., 2014; Trevarthen & Delafield-Butt, 2013). Deficits in praxis have also been observed in ASD, including poor imitation, gesture to command, and tool use skills (Kilroy et al. 2022; Mostofsky et al., 2006; Roley et al., 2015).

Children with DCD commonly display a generalized pattern of deficits in internal modeling, rhythmic coordination, interlimb coordination, gait and postural control, catching, sensoriperceptual function, and discontinuous movements (Wilson et al., 2013). In addition, children with DCD display deficits in praxis skills compared to typical participants, including deficits in imitation, gesture to command, and tool use (Paquet et al., 2019, Kilroy et al., 2022). In terms of writing and hand control, children with DCD have difficulties with control in manipulation tasks (Oliveira et al., 2006), hand posture, pen grip force, pen pressure, speed, fluctuations in velocity, and oversized movements (Biotteau et al., 2019).

ASD vs. DCD kinematic motor differences

The results for motor skills such as balance, aiming and catching, and manual dexterity, are mixed, with some studies showing poorer skills in the ASD group (Dewey et al., 2007; Wisdom et al., 2007), while other studies showing the opposite (Paquet et al., 2019), or mixed results, depending on the motor assessment (Green et al., 2002). In general, motor performance tests (balance, manual dexterity, ball skills) do not commonly show significant differences between children with ASD and DCD, but the two groups may differ in ability to imitate meaningful gestures and other gesture skills (ASD performing worse than DCD; Kilroy et al., 2022; Dewey 2007, Green et al., 2002; Paquet et al., 2019). Further, there may be differences in the underlying neurobiological basis of motor deficits between the two groups. A previous fMRI study from our group comparing ASD, DCD, and TD youth groups during action imitation, observation, and mentalizing tasks found ASD vs. DCD differences in a number of regions associated with motor planning, sensorimotor functioning, and action understanding (Kilroy et al., 2021). Specific differences in the cerebellum in particular are discussed below.

Cerebellum in ASD and DCD

Prior studies have shown functional and structural differences in both ASD and DCD groups in the cerebellum (Fatemi et al., 2012; Heijden et al., 2021). The cerebellum has been associated with skills of oculomotor control, motor speech, grip, control of voluntary movement, timing, sensorimotor coordination, and perception of hand movement (Manto et al., 2012). In addition to motor and balance skills, the cerebellum is also involved in working memory, executive and social functioning, and language processing (Levisohn et al., 2000; Riva & Giorgi, 2000). Cerebellar alterations may be associated with a number of behaviors seen in individuals with ASD (Sivaswamy et al., 2010), including difficulties with affect processing, executive function, prosody, social skills, eye contact, and repetitive behaviors (Riva & Giorgi, 2000). Studies have demonstrated in individuals with ASD, alterations in cerebellar gray matter structure (D'Mello et al., 2015; Stoodley, 2014), disruption of white matter tracts to and from the cerebellum (Catani et al., 2008; Di et al., 2018; Sivaswamy et al., 2010), altered functional connectivity between the cerebellum and the cerebral cortex (Khan et al., 2015; Noonan et al., 2009), and abnormal functional activity in the cerebellum during simple motor tasks (Allen et al., 2004).

Children with DCD also show cerebellar abnormalities, including decreased gray matter volume in cerebellar sensorimotor regions (lobule VIIIa; lobule IX), and an increased gray matter volume in cerebellar regions associated with motor behavior and cognition (lobule VI; crus I and crus II), compared to typically developing children (Gill et al., 2018). Compared to controls, children with DCD show reduced activation of the cerebellum during motor tasks of manual dexterity (Fuelscher et al., 2018 for review), predictive motor timing (Debrabant et al., 2013), finger sequencing (Licari et al., 2015), and visuomotor drawing (Pangelinan et al., 2013; Zwicker et al., 2011) in several fMRI and EEG studies. Further, in individuals with DCD, a recent study showed gray matter volume increase in the right crus II, left IX, and bilateral VIIIa following an intervention targeted for improving motor performance (Gill et al., 2022). Taken together, there is ample data suggesting that cerebellar regions may be involved in the interplay between sensorimotor and cognitive processing, and may be relevant to both ASD and DCD symptomology. In the current study, we focus in particular on the crus I and crus II as regions of interest in our analyses.

Motor Games

While prior studies using standard motor assessments (e.g. Motor Assessment Battery for Children) do not find significant gross motor, fine motor, and balance differences between

ASD and DCD groups, this finding may be reflective of measurement issues. One way to overcome potential measurement issues is to use machine learning analysis on more subtle motor information collected during digital motor games using smart tablets or other motion capture hardware. Machine learning has previously been used to analyze children's movements with an iPad serious game (Anzulewicz et al., 2016), a Kinect dance imitation game (Tunçgenç et al., 2020), and kinematic and eye movement features (Vabalas et al., 2019) to distinguish between ASD and TD children and adults. These studies classified individuals to their respective groups with between 73%-93% accuracy. Thus, previous findings support the use of machine learning with kinematic data for classifying ASD and TD individuals. However, to make this technology useful, it is essential to be able to distinguish ASD from other neurodevelopmental motor disorders. Thus, here we aim to attempt a similar method to distinguish ASD from another major group of children with developmental motor deficits, those with DCD.

Methods

Design: The current study was part of a larger study where youths (aged 8-17; Ns = 30 ASD, 23 DCD, 33 TD) participated in one day of behavioral testing and a second day of brain imaging (Kilroy et al., 2021). A subset of those youths participated in the current study (aged 8-17; Ns = 18 ASD, 16 DCD, 20 TD), performing on an iPad serious game following their scan session. Between-group comparisons of the full data set, including behavioral and brain imaging data, can be found in Kilroy et al. (2021), Harrison et al. (2021), Kilroy et al. (2022), and Ringold et al. (2022). Here, for brain imaging data, we include analysis from only the subset that completed smart-tablet games and those with usable brain imaging data (N=50; 19 TD, 16 ASD, 15 DCD).

Participants: Exclusion criteria for all groups included (a) IQ < 80 (in the clinical groups, cases where the full-scale IQ was less than 80, participants were included if their verbal IQ score or perceptual reasoning IQ score were greater than 80 as assessed by the Wechsler Abbreviated Scale of Intelligence 2nd edition (WASI-2; Wechsler, 2011); (b) history of loss of consciousness greater than five minutes; (c) left-handedness by self-report or as assessed by a version of the (Crovitz & Zener, 1962) questionnaire; (d) not fluent in English or parent without English proficiency; (e) born before 36 weeks of gestation. All participants were screened for MRI compatibility.

TD controls were additionally excluded if they had any psychological or neurological disorder. Additional exclusionary criteria included: scores below the twenty-fifth percentile on the Movement Assessment Battery for Children (MABC-2; Henderson et al., 2007), suspected DCD based on the Developmental Coordination Disorder Questionnaire (DCDQ; Wilson et al., 2007), and a Social Responsiveness Scale, Second Edition (SRS-2; Constantino & Gruber, 2012) score of T>60, indicating a risk of social impairment. Additionally, a T> 65 on the Conners 3AI-Parent report (Conners, 2008), indicating a risk for attention deficit and hyperactivity disorder (ADHD), was exclusionary for the TD group.

ASD group eligibility included a previous diagnosis through clinical diagnostic interview or diagnostic assessment as well as current clinical symptoms assessed using the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord et al., 2000), or previous symptoms using the Autism Diagnostic Interview-Revised (ADI-R; Lord et al., 1994). Individuals were excluded if they had a diagnosis of other neurological or psychological disorders with the exception of attention deficit disorder or generalized anxiety disorder. Eight ASD participants were on previously prescribed psychotropic medication at the time of data collection.

Probable DCD group eligibility criteria additionally included: (a) performance at or below the 16th percentile on the MABC-2; (b) no first degree relatives with ASD; (c) no concerns about an ASD diagnosis. The ADOS-2 was administered to participants whose SRS-2 scores were in the "severe risk" category of T= 65-74 (N=3), and participants who met ASD criteria on the ADOS-2 were excluded from the probable DCD group (none were excluded). Four DCD participants were on previously prescribed psychotropic medication at the time of data collection.

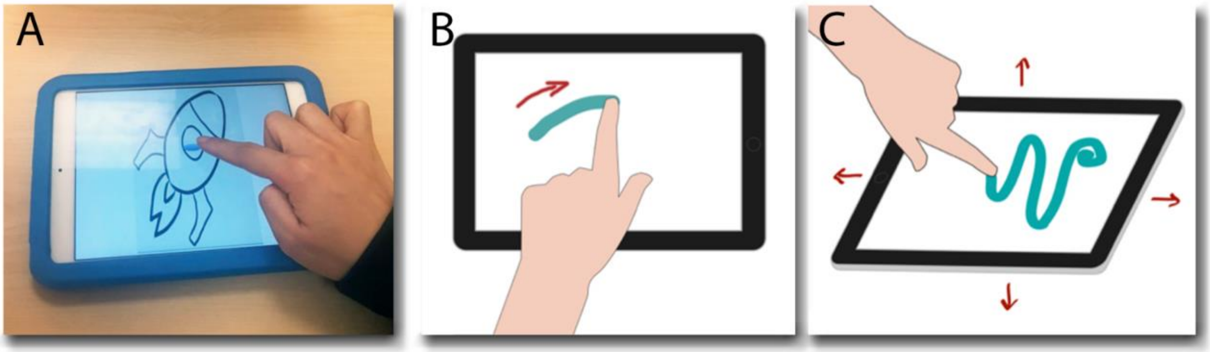
Participants were recruited through flyers and advertisements posted in community centres, on social media, website postings, clinics in the greater Los Angeles healthcare system, and local schools. Participants and parents were evaluated for their capacity to give informed consent, the study details were relayed in accordance with the protocols approved by the University of Southern California's Institutional Review Board, and written child assent and parental consent were obtained.

Behavioral Measures

Smart-tablet Game: A previously-tested naturalistic colouring game (Anzulewicz et al., 2016) was played by participants on an iPad mini (iPad mini 4; iOS 13.3.1). After a two minute trial session to become familiar with available tools and pictures, children completed the coloring game using their dominant hand for five minutes and these spontaneous action patterns were captured by inertial sensor data and touch-screen data from the iPad (Figure 1).

Figure 1.

Movement data acquisition.



(A) A child engages freely with the smart-tablet game, played on an iPad mini 4. The tablet is protected and made slip-resistant by a bumper and placed firmly on a table. Movement data are acquired from (B) the touch screen and (C) the inertial movement unit (IMU) sensor that detect the touch trajectories and accelerations along with the change of orientation of the iPad resulting from a gesture, respectively. Adapted from Anzulewicz et al. (2016).

Motor skills: Motor skills were assessed using the MABC-2 (Henderson et al., 2007), an administered assessment that evaluates motor skill performance in the domains of manual dexterity, gross-motor aiming and catching skills, and balance ability. Subtest and total scores were calculated using the normative samples from the second (ages 7–10) and the third (ages 11–16) age bands.

Praxis: A version of the The Florida Apraxia Battery (Rothi et al., 2003), modified for children (FAB-M, Mostofsky et al., 2006), was used to identify praxis ability. The assessment includes the completion of gestures prompted first by verbal command, then by action imitation, and last by use of physical tools. Seventeen gestures are transitive (e.g. stir cocoa with a spoon), seven items are intransitive (e.g. wave goodbye), and 9 items include unfamiliar nonrepresentational movements. Scoring followed the protocol developed by Mostofsky and colleagues (2006).

Behavioral Gameplay Coding Analysis from Video Data

Recordings from a minimum of 55 participants (see Table S1 for full breakdown), made during the five minutes in which kinematic data was captured from tablet gameplay, were coded for posture, engagement, and gameplay behavior. Participants' activity was coded across 13 characteristics (summarized in Table S1). Where a participant was not included for analysis in a particular characteristic, this was because it was not possible to code this characteristic from their video footage. Two researchers coded the recordings, with fourteen percent of videos being coded by both researchers. From this, Cohen's Kappa was used as a chance corrected measure to assess inter-rater agreement between researcher's coding of videos across 12 of the characteristics, whilst the Intraclass Correlation Coefficient (ICC) was calculated for the thirteenth characteristic ('number of pictures colored'). Coding results for 'number of pictures colored' were analyzed using a within-subjects one-way ANOVA, whilst differences between groups within the remaining 12 characteristics were analyzed using Fisher's exact test. Statistical significance was set to $p<.05$.

Machine Learning Analysis of Drawing Patterns

Gameplay and Feature Extraction. Data were collected for analysis from the 5-minute assessment phase. Gameplay data were collected by two sets of sensors within the smart-tablet: (1) the touch screen sensor recorded the Cartesian coordinates of each touch, with its displacement across the screen recorded with a variable sampling rate of ca. 60 times per second as the finger traveled across the screen; and (2) a triaxial accelerometer and gyroscope inertial movement unit (IMU) sensor that detected the small accelerations and rotations of the smart-tablet device as the child's fingers impact on the screen, creating small displacements to give subtle, but significant displacive forces during a gesture, with a regular sampling rate of ca. 20 times per second. Movement 'features' were then calculated from these raw sensor signals to build a computational characterization of each child's gameplay. These included, for example for the touch screen sensors, the duration of a gesture, its maximum velocity, deviation from a straight line, its peak acceleration, and the variance of these parameters across a gameplay session. Features from the raw IMU sensors were similarly extracted by calculating, for example, peak acceleration and rotation for each axis, their mean values and SDs.

Dataset Preparation and Feature Preprocessing. Two methods were used to extract features for two different forms of machine learning classification. Before employing the machine learning algorithms (namely neural networks), data were normalized in the standard feature-wise fashion. It was ensured that data were normalized only with usage of training data.

First, following previous work, Anzulewicz et al., 2016, we extracted 269 features from the data by simple computations of the raw sensor signals. These features were obtained from both the touch screen (105 features) and IMU (164 features) data. Features were computed using two approaches, either by calculating the variables for each individual feature or by calculating the variables across the gameplay session. Both were performed using a dedicated, bespoke engine. Touch data for each gaming session were aggregated and split into atomic gestures based on the start and end of any particular gestures. For every gesture, sets of variables were calculated. These features can be split into two major groups: (i) features of movements' kinematics, for e.g., velocity and acceleration; and (ii) tap-based features, (e.g., the number of taps in a game). Inertial sensor values were computed across the game session irrespective of the touch data. The values for each feature for each game were then reduced to its mean and used as input for machine learning.

A second approach chosen to preprocess the touch data required using singular movements. This included investigation of the differences in velocity and acceleration profiles of the movements. In previous work, this approach led to observation of significant differences between ASD and TD groups. Given that each particular movement consisted of a different number of points due to differing gesture durations, each movement was time normalized (start 0.0 and end 1.0) and split into n time bins, to ensure comparability of the movements. After assignment of a bin for each data point, values for each bin were calculated. Simple features were extracted for each bin: mean, median, standard deviation, 25% percentile (Q1), and 75% percentile (Q3). Next, each movement was treated separately and the prediction for this method required averaging the result. Probabilistic output from the model was generated for each movement and saved in the array. The final prediction was calculated as an average of the predictions generated for each movement.

Machine learning algorithms and methods of validation. Given the size of the dataset and imbalance between groups (ratio [1.1 - 1.25] between groups) several methods were employed to ensure stability of the results and the best possible generalization properties. First, imbalance between groups was mitigated by automatically computing an offset for each classified pair by weights (e.g., the bigger groups were multiplied by a factor less than 0). Next, cross-validation methods were employed with addition of repetitions for k-fold related methods. In the first run given the small sample size, one-leave-out cross validation was employed. Leave-one-out works in a way that there are n iterations (n - sample size) of the training. In each iteration, the model is trained on the n-1 samples and prediction is done on one sample that has been chosen in a random manner before the training. This ensures that every sample is treated once as the test set, and for the rest of the n-1 iterations, it is included in a training set. Data were shuffled before each step. Finally, methods using k-folding were used, both with use of standard and nested approaches (inner loop of cross validation) and to additionally randomize choice of the fold for validation, the whole process was repeated 10 times. Given the small sample size, this method was treated as an experimental approach. The number of folds was treated as a hyper-parameter. The number of folds that produced the best classification accuracy was four.

Two classes of models were tested. For the combined touch and sensor feature sets, neural networks were primary target models for testing. A typical scheme included usage of rather shallow networks (two layers deep, with 20-40 units per layer). Adam optimiser [2] (Beta 1 = 0.95, Beta 2 = 0.99) with learning rate (0.01 - 0.001) was used, with addition to invert scaling learning rate (with power = 0.5). Activation function employed: ReLU[3]. For the kinematic feature set, a gradient boosting machine [4] algorithm was used. A typical scheme employed consisted of 500-1000 trees, with a learning rate of 0.1, a subsample of 0.8, and with low depth (2) and log2 (feature count) features. Parameters for each model were optimized with a grid search method and the best results selected.

iPad feature analyses

First, the normality tests of a total of 283 variables were tested, including 269 features and 14 non-feature measures (i.e. age, IQ, MABC-2, and FAB-M scores). The normality of each variable was tested across all participants and within each group (TD, ASD, and DCD) to decide if a parametric or non-parametric test should be used for further analyses. One-sample Kolmogorov-Smirnov test was performed using customized MATLAB scripts. A p-value <0.05 was returned in most of the tests, which means the test rejected the null hypothesis that the data come from a standard normal distribution. As all the non-feature variables were not normally distributed across all participants nor within each group, non-parametric analyses were performed to determine the correlations between each feature and other non-feature measures. Kendall’s tau was chosen because it would be less affected by extreme values. Customized MATLAB scripts were used to perform the correlation analyses.

iPad feature analysis I. Group comparisons. A total of 269 features of each participant were calculated from their smart-tablet gameplay session, including 164 features calculated using the inertial measurement unit (IMU) data and 105 features calculated using the touch trajectories on the screen. The purpose of this analysis is to identify the smart-tablet features statistically different among or between groups for further correlation analyses with the brain imaging data. For those with normal distributions, one-way ANOVA was used to compare the feature values among the TD, ASD, and DCD groups, and an independent-t test was used to compare the feature values between two groups. Except for some testing conditions of the 4 IMU features listed in Table S3, non-parametric tests were performed for comparisons. Kruskal-Wallis test was used to compare the feature values among the TD, ASD, and DCD groups, and Mann-Whitney U test was used to compare the feature values between two groups. These statistical analyses were performed using SPSS.

Machine Learning Differentiation Between ASD, DCD, and TD Groups

Two classes of models were tested. For the combined touch and sensor feature sets, neural networks were primary target models for testing. A typical scheme included usage of shallow networks (two layers deep, with 20-40 units per layer). Adam optimiser [2] (Beta 1 = 0.95, Beta 2 = 0.99) with learning rate (0.01 - 0.001) was used, with addition to invert scaling learning rate (with power = 0.5). Activation function employed: ReLU[3]. For the kinematic feature set, a gradient boosting machine learning [4] algorithm was used. A typical scheme employed 500-1000 trees with a learning rate of 0.1, a subsample of 0.8, and with low depth (2) and log2 (feature count) features. Parameters for each model were optimized with a grid search method and the best results selected.

Functional Brain Imaging Acquisition and Analysis

Complete information on stimuli, imaging acquisition and analysis can be found in Kilroy et al., 2021, as well as in the Supplemental Materials. Two tasks from that prior study -- action execution and imitation -- presented in an 8-min run each, were used for current data analysis (see Kilroy et al., 2021). Subjects practiced all tasks in a mock scanner prior to scanning. They were also filmed and monitored in the MRI in order to confirm task completion. In the current study, we either used only the hand condition (since it is most similar to the iPad task) or took a mean of activity across conditions (all condition; in order to maximize number of trials).

Execution. Still stimuli were used as cues to execute a pre-trained action. In each 15-second block, three stimuli (5-sec each) were presented from one of three categories: emotional facial actions, non-emotional facial actions, and hand actions. The emotional facial actions were cued in the following way: a photograph of a dead plant cued making a sad expression, a photograph of moldy bread cued making a disgust expression, and a poison bottle cued making a fear expression. The non-emotional facial actions were cued in the following way: a spot of whip cream on one of three points around the mouth of a face with a neutral expression cued bringing the tongue to lip in that spot. Bimanual hand actions were cued by: a photograph of a xylophone cuing pantomime playing xylophone, a photograph of grapes cuing pantomiming pulling grapes from stems, and a photograph of a game controller cuing pantomiming pushing buttons on the game controller as if playing a video game. Participants were instructed to perform the cued action for the entire time that the stimulus was presented (5 s).

Imitation. The Imitation task used videos depicting three categories of actions: (a) emotional face actions (e.g., smiling); (b) nonemotional face actions (e.g., tongue to upper lip); (c) bimanual hand actions (e.g., hands playing xylophone; face not shown). Stimuli were presented for 3.75 s in a block design consisting of three stimuli per block with a 1.25-s black screen as a transition between each video/still followed by a 15-s rest block. A black crosshair in the middle of a white screen was shown for the rest block. Excluding an initial junk block, five blocks of each stimulus condition were alternated with rest in a pseudo-random sequence creating a total of 15 different videos for each category per run. Seven different Caucasian adult actors were used to create the stimuli. No stimulus was repeated in the same run and no block contained more than two same sex actors. No more than two videos per block contained the same valenced emotion (i.e., two high valenced emotions and one low valenced emotion). Participants were instructed to copy what they saw for the full duration of each video.

fMRI data acquisition. fMRI data were acquired on a 3 Tesla MAGNETOM Prisma (Siemens, Erlangen, Germany) with a 20-channel head coil. Each functional scan consisted of an echo-planar imaging (EPI; 150 whole brain volumes) acquired with the following parameters: TR = 2s, TE = 30 ms, flip angle = 90˚, 64x64 matrix, in-plane resolution 3x3mm, and 41 transverse slices, each 1.5mm thick, covering the whole brain with a multiband factor of three. Spin Echo EPI field mapping data was also acquired in AP and PA directions with identical geometry to the EPI data for EPI off-resonance distortion correction (TR = 1020 ms, TE1 = 10 ms, TE2 = 12.46 ms, flip angle = 90°, FOV = 224 × 224 × 191 mm3, voxel size = 1.5 × 1.5 × 1.5 mm). A structural T1-weighted MPRAGE was acquired for each subject (TR = 1950 ms, TE = 3.09 ms, flip angle = 10˚, 256 x 256 matrix, 176 sagittal slices, 1 mm isotropic resolution). Total scan time was 5 minutes.

Within-subject analyses. Subject level functional imaging analyses were completed using FSL. The following preprocessing steps were taken: brain extraction for non-brain removal; spatial smoothing using a Gaussian kernel of FWHM 5mm; B0 unwarping in the y-direction, standard ICA-AROMA (Pruim et al., 2015) to remove motion-related noise, high pass filter with a cutoff period of 90 seconds, and subject-specific motion correction parameters were entered as nuisance regressors. Functional images were registered to the high-resolution anatomical image using a 7-degrees of freedom linear transformation. Anatomical images were registered to the MNI-152 atlas using a 12-degree of freedom affine transformation, and further refined using FNIRT for nonlinear registration. Experimental conditions were each modeled with a separate regressor derived from a convolution of the task design and a double gamma function to represent the hemodynamic response and temporal derivative.

Group Analysis and Main Effects. All three groups were entered into the multivariate linear regression model. Age, sex, and full scale IQ were centered across groups and entered as covariates. For group analysis, image registration was performed using FSL's FLIRT (Jenkinson et al., 2002; Jenkinson & Smith, 2001). Each individual's statistical images were entered into a higher level mixed-effects analysis using FSL's FLAME algorithm. Three stimulus conditions (emotional face, non-emotional face, and hand actions) were collapsed to determine the main effect of the task compared to a resting baseline. Resulting group level images were thresholded using FSL's cluster probability algorithm, with $Z>3.1$ and a corrected cluster size probability of $p = .05$, FDR.

Whole brain activation related to iPad features. To determine whether smart-tablet features were correlated with BOLD response to the action execution task, three separate regression analyses were performed with the mean-centered features. These comparisons were also thresholded at $Z>3.1$, FDR. Parameter estimates for significant clusters were extracted from each participant and plotted in a graph to rule out the presence of outliers. Any individual who had a mean percent signal change over 3 box lengths (length between first and third quartiles) from the median was removed from feat query analyses, and R and p values for the whole group and within-group correlations were calculated.

ROIs. We focused on anatomical ROIs defined using the Harvard-Oxford dictionary and including the cerebellar crus I and cerebellar crus II. Parameter estimates for ROIs were extracted from two conditions in the execution and imitation tasks. Independent-samples t-tests were performed to identify group differences in ROI activation. Pearson correlation was performed across groups and within groups between each parameter estimate with the 3 selected iPad features that demonstrated the group differences in earlier analyses.

Results

Behavioural Gameplay Coding

Cohen's Kappa demonstrated high inter-rater reliability between researchers with κ ranging from .579 to 1.00 with a mean of .808 across all characteristics excluding 'number of pictures colored'; for this characteristic there was perfect agreement between researchers (ICC = 1.00). There were no significant differences between groups (F(2,52) = 0.142, p = 0.868) in the number of pictures colored. Fisher's exact tests performed for the remaining 12 characteristics showed no significant difference between groups (p > .08 for all characteristics). Results demonstrate a lack of evidence for visible behavioral, postural or engagement differences during gameplay between participants with ASD, DCD, or TD youths.

Group differences

All means and standard deviations are reported below in Table 1. Children with ASD and DCD did not differ on IQ, motor performance, or praxis measures, though both groups differed on motor performance and praxis as compared to TD participants (Table 1).

Table 1.

Descriptives and Group Comparisons

	TD	ASD	DCD	TD, ASD, DCD	TD, ASD	TD, DCD	ASD, DCD
	Mean ± SD	Mean ± SD	Mean ± SD	*p*	*p*	*p*	*p*
Age	12.09 ± 2.58	12.36 ± 1.99	12.03 ± 2.25	0.755			
WASI-II VCI	114.85 ± 12.73	109.72 ± 20.27	115.44 ± 17.54	0.339			
WASI-II PRI	111.95 ± 12.42	108.22 ± 20.82	107.56 ± 23.13	0.710			
WASI-II FSIQ-4	115.30 ± 11.22	109.56 ± 19.35	112.38 ± 19.88	0.583			
WASI-II FSIQ-2	115.35 ± 11.07	109.00 ± 17.77	114.44 ± 19.01	0.481			
MABC-2 MD	10.00 ± 2.38	4.56 ± 1.98	5.25 ± 2.60	0.000**	0.000**	0.000**	1.000
MABC-2 AC	11.15 ± 2.96	6.22 ± 3.69	6.44 ± 2.76	0.000**	0.000**	0.001**	1.000
MABC-2 Balance	10.40 ± 2.50	6.78 ± 3.46	5.31 ± 2.15	0.000**	0.003**	0.000**	0.328
MABC-2 Total	10.40 ± 1.67	4.67 ± 2.45	4.44 ± 1.79	0.000**	0.000**	0.000**	1.000
FAB-M GTC	0.69 ± 0.14	0.57 ± 0.13	0.63 ± 0.15	0.035*	0.029*	0.724	0.575
FAB-M IMI	0.65 ± 0.13	0.46 ± 0.16	0.51 ± 0.17	0.001**	0.001**	0.034*	1.000
FAB-M TU	0.80 ± 0.09	0.59 ± 0.15	0.65 ± 0.14	0.000**	0.000**	0.012*	0.675
FAB-M IMI ML	0.58 ± 0.21	0.40 ± 0.22	0.44 ± 0.20	0.027*	0.032*	0.160	1.000
FAB-M IMI MF	0.70 ± 0.13	0.51 ± 0.17	0.58 ± 0.17	0.002**	0.002**	0.054	1.000

Note. Kruskal-Wallis was first used to test the differences among the three groups. If the distribution was not the same across the three groups ($p<0.05$), Dunn's pairwise tests were carried out for the three pairs of groups. *$p < 0.05$, **$p < 0.01$; *** $p < 0.001$. Abbreviations: WASI-II=Wechsler Abbreviated Scale of Intelligence, Second Edition; VCI= Verbal Comprehension Index; PRI: Perceptual Reasoning Index; FSIQ-4: Full Scale IQ; WASI-II: FSIQ-2= Two Factor IQ; MABC-2=Movement Battery for Children, Second Edition; MD=Manual Dexterity; AC=Aiming and Catching; FAB-M=Florida Apraxia Battery-Modified; GTC=gesture to command; IMI=imitation; TU=tool use; ML=meaningless gestures; MF meaningful gestures; SD= standard deviation

Machine Learning Differentiation Between ASD, DCD, and TD Groups

Despite no human-observed behavioral differences between groups during gameplay (above), and no motor differences detecting between ASD and DCD using conventional human-rated instrument (MABC-2, FAB-M), machine learning analytics of the smart-tablet sensor data were successful in differentiating ASD from DCD motor patterns with 71% accuracy, as well both ASD and DCD as from TD motor patterns with 76% and 78% accuracy, respectively (Table 2). Moreover, when classified in a 3-way paradigm, i.e. where each subject could fit into any one of the three classes, which is a more challenging classification task, overall accuracy yield was 57%, or 73% above chance (Table 3). These results reveal the presence of different motor signatures specific to the specific neurodevelopmental disorders, and demonstrate computational analysis – even in a simple tablet game – is significantly more sensitive to detecting motor differences than human-rated instruments. Further, this computational assessment can be specific to clinical groups, identifying reliable pattern differences between ASD and DCD, as well as between ASD and TD or DCD and TD.

Table 2.

Machine learning classification performance between sets of two groups.

Pair	Feature set	Accuracy	Sensitivity	Specificity
TD, ASD	touch, sensors	0.763	0.85	0.67
TD, ASD	kinematic	0.634	0.65	0.60
TD, DCD	touch, sensors	0.778	0.75	0.81
TD, DCD	kinematic	0.666	0.7	0.63
ASD, DCD	touch, sensors	0.706	0.67	0.75
ASD, DCD	kinematic	0.694	0.67	0.80

Table 3.

Machine learning classification matrix for all three groups

		Clinical group		
		TD	ASD	DCD
Machine learning classification	TD	14	2	4
	ASD	5	11	2
	DCD	5	5	6

Note: Table is showing overall accuracy yield of 57%, or 73% above chance.

Comparisons Between ASD, DCD, and TD Groups

Statistical results indicated no 3-way significant differences in the feature distributions across all groups. Features with 3-way differences of $p<0.1$ between all groups, and significant differences ($p<0.05$) between pairwise groups are reported (Table 4). Based on these data, four features were initially selected for further correlation analyses with brain imaging data: Attitude Variance, Gesture Area Variance, Minimum Gesture Acceleration, and Gesture Directness Variance. Attitude Variance was then removed from the imaging analysis due to the presence of outliers, and its potential confounding with protocol deviations of the smart-tablet being moved from the testing spot during assessment. However, as machine learning analysis was not as sensitive to this type of outlier interference, it was retained as a feature in the previous sections. The three remaining features computed (i) the variance in the area of a gesture of all swipes during a trial, where area was calculated by placing a minimal polygon around the swipe and its area calculated (Gesture Area Variance); (ii) the minimum acceleration of a swipe (Minimum Gesture Acceleration); and (iii) the variance in the smoothness of a gesture during its final data points across a gameplay trial (Gesture Directness Variance).

Neuroimaging Results

In order to try to better understand the neurological mechanisms behind our behavioural smart-tablet results, we correlated the previously mentioned three features that distinguished groups (Gesture Area Variance, Minimum Gesture Acceleration, Gesture Directness Variance) with levels of change in neural activation during action execution in the scanner. Both whole brain and ROI analyses are reported.

Table 4.

Feature Data; significant differences between groups

Feature name	Type	TD, ASD, DCD	TD, ASD	TD, DCD	ASD, DCD
		p	*p*	*p*	*p*
Attitude Variance (x-axis)*	IMU	0.058	0.038	---	0.042
Gesture Area Variance*	Touch	0.058	---	---	0.030
Minimum Gesture Acceleration*	Touch	0.073	0.030	---	---
Rate of Change of Acceleration Direction (y-axis)	IMU	0.076	---	---	0.046
Rotation Variance (x-axis)	IMU	0.076	0.033	---	---
Attitude Variance (y-axis)	IMU	0.094	0.035	---	---
Gesture Directness Variance*	Touch	0.098	---	0.036	---

Note. Features with both borderline significant differences ($p<0.1$) across TD, ASD, and DCD groups, and with significant differences ($p<0.05$) between groups are summarized in this table.

* starred features were initially selected for further correlation analyses with the brain imaging data while Attitude Variance was then removed due to the presence of outliers, and its potential confounding with protocol deviations of the smart-tablet being moved from the testing spot during data collection.

Main effects. During the execution and imitation tasks, across all groups and all conditions, there was widespread significant activation, including regions in the bilateral superior frontal cortex, precentral gyrus, postcentral gyrus, inferior frontal gyrus, premotor cortex, superior parietal cortex, cingulate cortex, lateral occipital regions, cerebellum, insular cortex, and STS (Figure 2A).

Whole Brain Group Differences. In the execution task, the only significant group difference identified was TD > DCD in the medial-frontal cortex during the all action condition (Figure 2B). During the imitation task, DCD demonstrated a decrease in cerebellar activation when compared to both TD (right cerebellar crus II) and ASD (left cerebellar crus I & II) groups during all conditions ($Z>3.1$; Figure 2B). During imitation of hand actions, the DCD group had reduced activation compared to the TD group in right lateral occipital cortex, and the angular gyrus during all conditions ($Z>3.1$; Figure 2B), and in the right lateral occipital cortex, and the left frontal pole ($Z>3.1$). During imitation of all actions, the ASD group had lower activation of the right postcentral gyrus compared to the DCD group, and of the right precuneus and right superior temporal gyrus compared to the TD group ($Z>3.1$; Figure 2B).

ROI group differences. Visuals of ROIs and group differences can be found in Figure 3. During execution of all actions, the ASD group had reduced activation in the right cerebellar crus I compared to the DCD group ($p= .033$). During action imitation, the DCD group had reduced activation compared to the TD group in left cerebellar crus I (HAND: $p = .015$; ALL: $p = .020$), the right cerebellar crus II (HAND: $p = .022$; ALL: $p = .022$), and left cerebellar crus II (ALL: $p = .012$). The DCD group also had reduced activation compared to the ASD group in left cerebellar crus II during imitation of all actions ($p = .009$).

Figure 2.

Whole brain main effects and group differences during all actions. STG: superior temporal gyrus; AG: angular gyrus; PCG: postcentral gyrus; MFC: medial-frontal cortex; R: right; L: left.

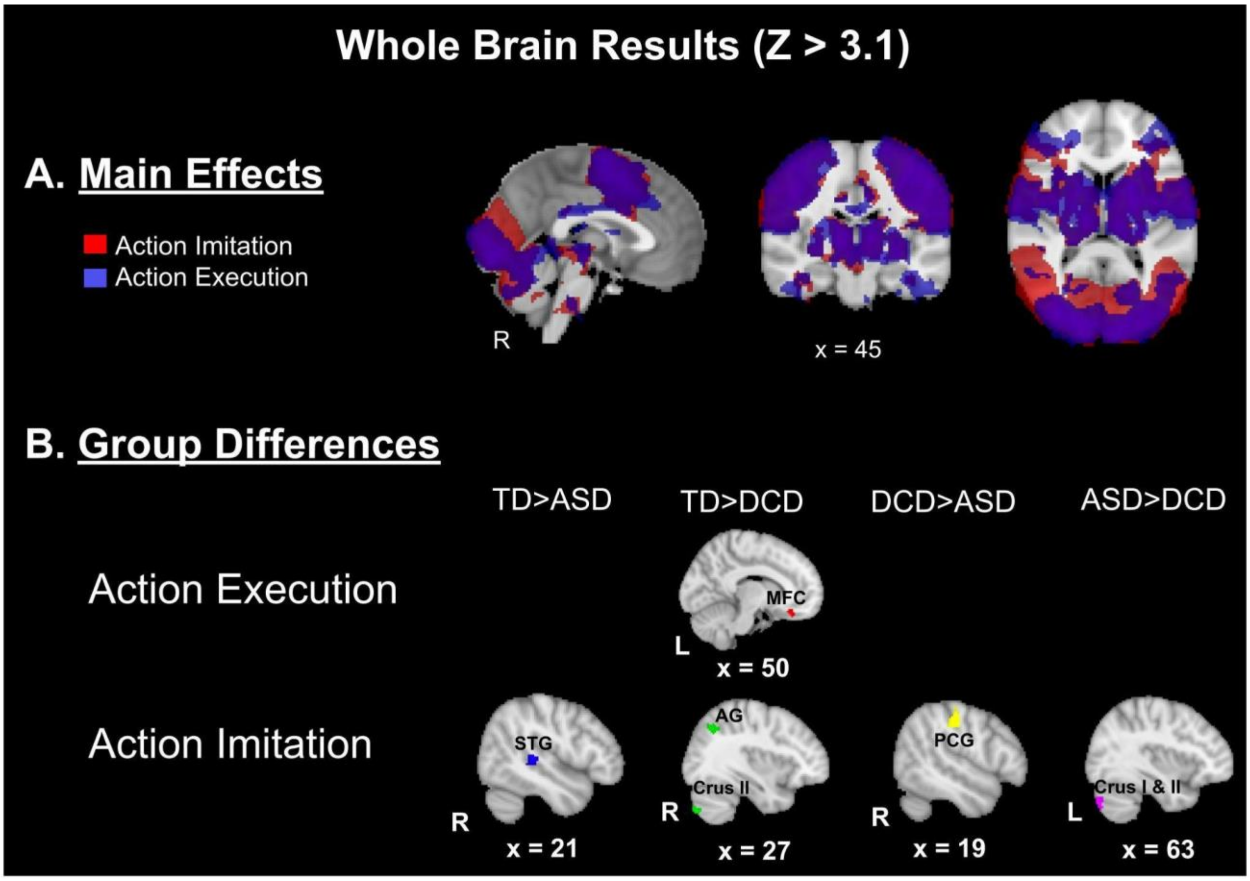
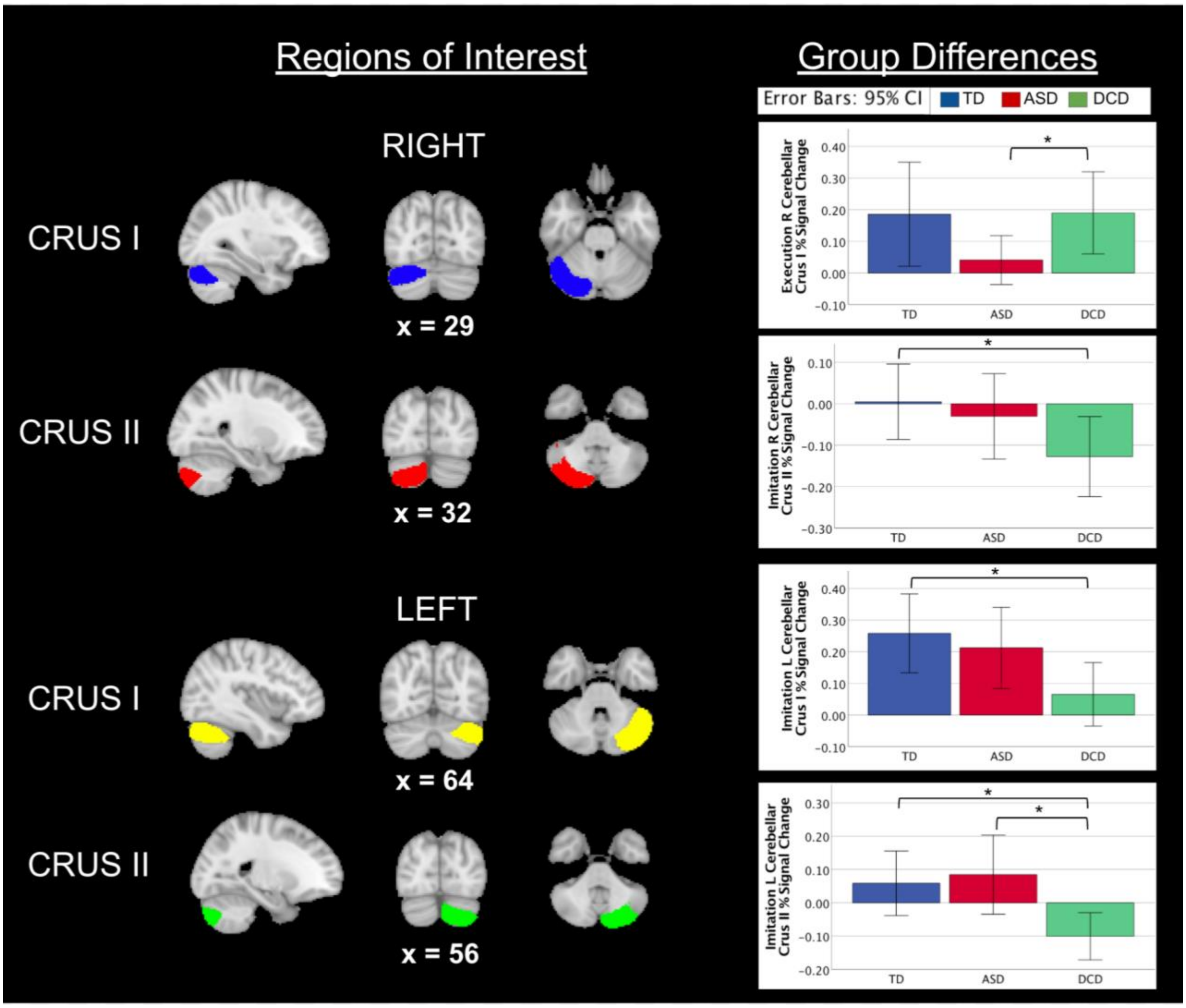


Figure 3

Regions of interest and group differences.



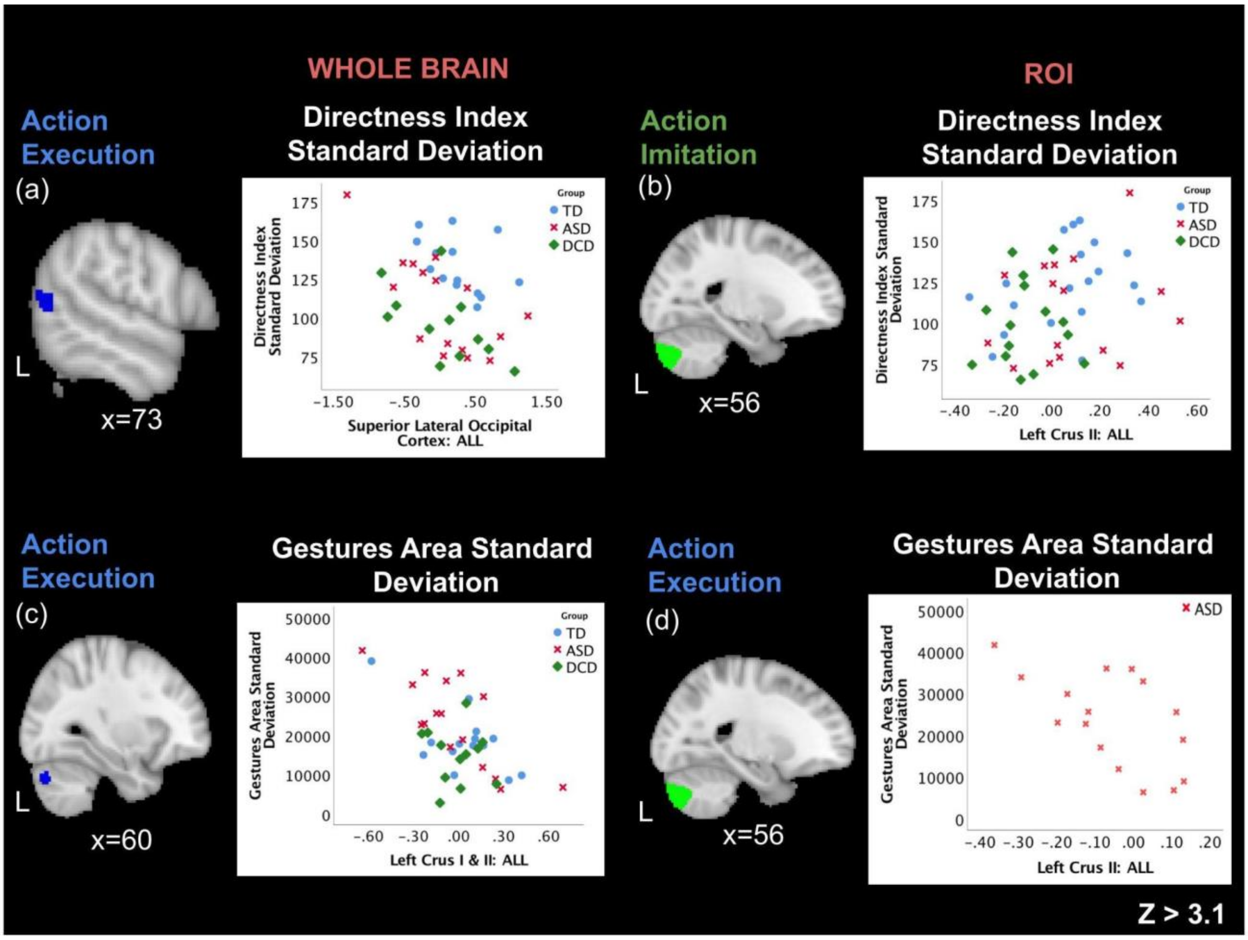
Note. Between group differences in cerebellar regions of interest during execution and imitation of all actions. TD = typically developing, ASD = autism spectrum disorder, DCD = developmental coordination disorder, R = right, L = left

Whole Brain Correlations. Across all participants during execution of all actions, Directness Variance was negatively correlated with activation in the left superior lateral occipital cortex, and Gesture Area Variance was negatively correlated with activation in the left cerebellar crus I & II (Figure 4). Significant correlations were also observed between Gesture Directness Variance and left superior lateral occipital cortex in the ASD (r= -.693, p=.003) and DCD groups (r= -.589, p=.044). Gesture Area Variance was negatively correlated with activation in the left cerebellar crus I & II, during execution of all conditions (Figure 4). This relationship was also significant in the ASD (r= -.765, p=.001) group. There were no whole brain correlations in the imitation task with any of the smart-tablet features.

ROI Correlations. Gesture Area Variance was one of the features that best separated DCD from ASD participants on the smart-tablet task. Across all participants, a significant negative correlation was observed between this feature and the right cerebellar crus II during execution of all actions (r= -.360, p=.019). This relationship was also observed in the ASD group where it trended toward a significant negative correlation with the right cerebellar crus II for all conditions (r= -.497, p=.050). Activity in the left cerebellar crus I was significantly negatively correlated with Gesture Area Variance (r= -.360, p=.019) during execution of hand actions across groups, though this relationship was not significant in any individual group. During imitation, for all actions across groups, the left cerebellar crus II was significantly positively correlated with Gesture Directness Variance (r= .284, p=.046), though this relationship was not significant in any individual group.

Figure 4.

Correlations with iPad features



Note. Correlations during action execution and action imitation across all participants during all action conditions.

Discussion

Here we demonstrate that using machine learning from a simple motor colouring game on a smart-tablet, we can significantly differentiate the gameplay of children with ASD, DCD, and TD. This is especially significant since in the current study, standard behavioural motor measures could not distinguish between ASD and DCD groups, nor could video coding analysis. We further show that the driving motor features that contributed greatly to differentiating clinical groups include measures that reflect control of movements and their degree of displacement. Finally, cerebellar regions previously associated with reduced activation in both ASD & DCD groups show significant relationships with kinematic features from the smart-tablet data. We further discuss each of these results below.

Classifying ASD/DCD/TD by game-play

By using a simple smart-tablet coloring game, one can compute gesture kinematics from the sensor data captured by the touchscreen, and record the forces put into the smart-tablet using data from the inertial measurement unit (Anzulewicz et al., 2016). Coupled with machine learning, these kinematics are able to categorize ASD from TD at 76%, ASD from DCD at 71%, and DCD from TD at 78% accuracy. To our knowledge, this is the first time serious game digital technology has been used to distinguish two similar motor developmental disorders – ASD from DCD. Given that visual behavioral analysis of video data nor standard motor assessments, such as the MABC-2 did not distinguish the two groups apart, this finding is especially remarkable, and gives important impetus to advance this scalable methodology to usefully contribute to clinical diagnosis, as well as better informing the particular underlying motor disturbances in each group. Refinements of this technique can be explored in future studies to increase between-group categorization accuracy, for example by additionally including social motor games (see below).

Motor Markers that Distinguish Groups

The kinematic markers that most contribute to differentiating between groups include the control of deceleration and variability in the distance, or area covered, of the motor gestures. On average, individuals with ASD were more variable in the size of the gesture area used on the smart-tablet than individuals with DCD for each motor gesture. This suggests that for an individual with ASD, there is more variability in gesture size, with some gestures made as very small and some as very big in the course of the coloring game, compared to individuals in the DCD group. Such large variation may be related to two contrasting types of gesture behavior within an ASD individual, large gestures driven by a reluctance to shift from the ongoing gesture once engaged with it, and very short gestures produced by rapid tapping. Either way, the underlying nature of ‘restricted and repetitive’ interests manifests in each type of motor behavior. Future work will need to investigate this to better understand individual action patterns and their distribution in ASD individuals.

Finally, we investigated neural regions (cerebellar crus I/II) previously associated with differences in ASD and DCD groups (Allen et al., 2004, Gill et al., 2018, Fuelscher et al., 2018, Debrabant et al., 2013, Licari et al., 2015, Pangelinan et al., 2013; Zwicker et al., 2011), and differences in imitation and praxis (Dapretto et al., 2006; Leslie et al., 2004). In ASD groups, crus I has previously been shown to be involved in control of hand movements, performance of precision grips (Vaillancourt et al., 2006; Neely et al., 2013), and force variability (McKinney et al., 2022) and related to repetitive behaviors in females (McKinney et al., 2022). Crus I and II together are involved in sensorimotor tasks, as well as working memory, attention, and social cognition (McKinney et al., 2022; Guell et al 2020; Van Overwalle 2020). Thus they may be particularly involved in the interplay between sensorimotor function and cognition. Notably, while difficulties with working memory, attention, and social cognition are common symptoms of ASD, individuals with DCD may fall between ASD and TD groups on all these behaviors (Kilroy et al., 2022, Ringold et al., 2022). Our data indicate that during motor imitation, the crus II is significantly hypoactive in DCD (Left: TD/ASD>DCD [ROI]; ASD>DCD [whole brain]; Right: TD>DCD [ROI & whole brain]). For the right crus I, during the execution task, we find the ASD group is hypoactive compared to the DCD group (DCD>ASD [ROI]). For the left crus I, during imitation, both clinical groups are hypoactive compared to TD, though the DCD group may show significantly more hypoactivity (TD>ASD/DCD [ROI] and ASD>DCD [whole group]). Taken together, these data indicate that during motor tasks, the right and left crus II are particularly hypoactive in DCD, while activity patterns in crus I may be more nuanced between groups. It is also possible that differences previously observed in DCD in imitation performance may be more related to cerebellar influences, than imitation differences previously observed in ASD, which may be more dependent on frontal cortical regions (Kilroy et al., 2021).

Interestingly, we find that during our fMRI motor tasks, activity in these cerebellar regions correlates with the iPad kinematic features that are the best at differentiating between specific pairwise groups. During motor tasks, we find activity in the left crus II correlates with Gesture Directness Variance across participants and with Gesture Area Variance in the ASD group. The latter pattern is also found for the left crus I and right crus II; during execution, activity in these areas correlates with Gesture Area Variance across groups and within the ASD group. These cerebellar regions may show differential activation patterns in ASD and DCD, and their activity may also be related to motor control of deceleration and measures of gesture size, which both clinical groups perform differentially, in alignment with prior studies (McKinney et al., 2022). Thus differential activity in these cerebellar regions may lead to behavioral motor differences between groups, allowing the use of kinematic patterns to distinguish between ASD, DCD, and TD groups. Interestingly, only features that were associated with classifying ASD vs DCD, and TD vs DCD differences were correlated with brain activity during our tasks, no features that were best associated with classifying TD vs ASD differences were correlated with brain activity, reiterating the idea that crus I & II may be more involved gesture execution and imitation deficits seen in the DCD group. Previous literature has shown hyperconnectivity (during resting state & motor tasks) between crus I & II and premotor & motor cortices (Jung et al., 2014; Verly et al.,2014), therefore domain specificity of cerebro-cerebellar connections might be abnormal in ASD, rather than cerebellar activation alone.

Limitations and Future Directions.

We note that future studies are needed with larger sample size and more diverse groups (e.g., more females; larger age range; left handers; wider range of IQ). Further, motor games with more social aspects (e.g., imitation, social interactions) may offer even better categorization accuracy between groups. We also acknowledge that the task in the scanner did not match the task outside the scanner, future studies should attempt to execute smart-tablet tasks during fMRI.

Conclusions

Here we show that kinematics from a simple motor smart-tablet game can be utilized to categorize ASD, DCD, and TD groups. We further show that two driving kinematic markers for this categorization are control of deceleration and variability in gesture size. These two kinematic markers are associated with neural activity in cerebellar regions during motor tasks across groups. These data may be important for the development of motor markers for screening and diagnosis of ASD and DCD and for development of individualized interventions.

References

Allen, G., Müller, R.-A., & Courchesne, E. (2004). Cerebellar function in autism: Functional magnetic resonance image activation during a simple motor task. *Biological Psychiatry*, *56*(4), 269–278. https://doi.org/10.1016/j.biopsych.2004.06.005

Anzulewicz, A., Sobota, K., & Delafield-Butt, J. T. (2016). Toward the Autism Motor Signature: Gesture patterns during smart tablet gameplay identify children with autism. *Scientific Reports*, *6*(1), 31107. https://doi.org/10.1038/srep31107

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: American Psychiatric Association.

Anjana N. Bhat, Rebecca J. Landa, James C. (Cole) Galloway, Current Perspectives on Motor Functioning in Infants, Children, and Adults With Autism Spectrum Disorders, Physical Therapy, Volume 91, Issue 7, 1 July 2011, Pages 1116–1129, https://doi.org/10.2522/ptj.20100294

Biotteau M, Danna J, Baudou É, Puyjarinet F, Velay JL, Albaret JM, Chaix Y. Developmental coordination disorder and dysgraphia: signs and symptoms, diagnosis, and rehabilitation. Neuropsychiatr Dis Treat. 2019;15:1873-1885 https://doi.org/10.2147/NDT.S120514

Blank, R., Barnett, A. L., Cairney, J., Green, D., Kirby, A., Polatajko, H., Rosenblum, S., Smits-Engelsman, B., Sugden, D., & Wilson, P. (2019). International clinical practice recommendations on the definition, diagnosis, assessment, intervention, and psychosocial aspects of developmental coordination disorder. *Developmental Medicine & Child Neurology*, *61*(3), 242–285. https://doi.org/10.1111/dmcn.14132

Campbell, S. K., & Hedeker, D. (2001). Validity of the Test of Infant Motor Performance for discriminating among infants with varying risk for poor motor outcome. *The Journal of Pediatrics*, *139*(4), 546–551. https://doi.org/10.1067/mpd.2001.117581

Catani, M., Jones, D. K., Daly, E., Embiricos, N., Deeley, Q., Pugliese, L., Curran, S., Robertson, D., & Murphy, D. G. M. (2008). Altered cerebellar feedback projections in Asperger syndrome. *NeuroImage*, *41*(4), 1184–1191. https://doi.org/10.1016/j.neuroimage.2008.03.041

Cavallo, A., Romeo, L., Ansuini, C., Battaglia, F., Nobili, L., Pontil, M., Panzeri, S., & Becchio, C. (2021). Identifying the signature of prospective motor control in children with autism. *Scientific Reports*, *11*(1), 3165. https://doi.org/10.1038/s41598-021-82374-2

Chang, S.-H., & Yu, N.-Y. (2016). Comparison of motor praxis and performance in children with varying levels of developmental coordination disorder. *Human Movement Science*, *48*, 7–14. https://doi.org/10.1016/j.humov.2016.04.001

Cermak, S. A., & May-Benson, T. A. (2020). Praxis and dyspraxia. In I. A. C. Bundy & S. Lane (Eds.), Sensory integration: Theory and practice (3rd ed., pp. 115–150). F.A. Davis.

Chua, Y. W., Lu, S.-C., Anzulewicz, A., Sobota, K., Tachtatzis, C., Andonovic, I., Rowe, P., & Delafield-Butt, J. (2022). Developmental differences in the prospective organisation of goal-directed movement between children with autism and typically developing children: A smart tablet serious game study. *Developmental Science*, *25*(3), e13195. https://doi.org/10.1111/desc.13195

Chukoskie, L., Townsend, J., & Westerfield, M. (2013). Chapter Seven - Motor Skill in Autism Spectrum Disorders: A Subcortical View. In G. Konopka (Ed.), *International*

Review of Neurobiology (Vol. 113, pp. 207–249). Academic Press. https://doi.org/10.1016/B978-0-12-418700-9.00007-1

Constantino, J. N., & Gruber, C. P. (2012). Social responsiveness scale: SRS-2 (p. 106). Torrance, CA: Western psychological services.

Conners, C. K. (2008). Conners 3rd edition: Manual (Vol. 14). Toronto, Ontario, Canada: Multi-Health Systems.

Cook, J. L., Blakemore, S.-J., & Press, C. (2013). Atypical basic movement kinematics in autism spectrum conditions. *Brain*, *136*(9), 2816–2824. https://doi.org/10.1093/brain/awt208

Craig, F., Lorenzo, A., Lucarelli, E., Russo, L., Fanizza, I., & Trabacca, A. (2018). Motor competency and social communication skills in preschool children with autism spectrum disorder. *Autism Research*, *11*(6), 893–902.

Crovitz, H. F., & Zener, K. (1962). A Group-Test for Assessing Hand- and Eye-Dominance. *The American Journal of Psychology*, *75*(2), 271–276. https://doi.org/10.2307/1419611

David, F., Baranek, G., Wiesen, C., Miao, A., & Thorpe, D. (2012). Coordination of precision grip in 2–6 years-old children with autism spectrum disorders compared to children developing typically and children with developmental disabilities. *Frontiers in Integrative Neuroscience*, *6*. https://www.frontiersin.org/article/10.3389/fnint.2012.00122

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature neuroscience*, *9*(1), 28-30.

Debrabant, J., Gheysen, F., Caeyenberghs, K., Van Waelvelde, H., & Vingerhoets, G. (2013). Neural underpinnings of impaired predictive motor timing in children with Developmental Coordination Disorder. *Research in Developmental Disabilities*, *34*(5), 1478–1487. https://doi.org/10.1016/j.ridd.2013.02.008

Dewey, D., Cantell, M., & Crawford, S. (2007). Motor and gestural performance in children with autism spectrum disorders, developmental coordination disorder, and/or attention deficit hyperactivity disorder. Journal of the International Neuropsychological Society, 13(2), 246-256. doi:10.1017/S1355617707070270

Di, X., Azeez, A., Li, X., Haque, E., & Biswal, B. B. (2018). Disrupted focal white matter integrity in autism spectrum disorder: A voxel-based meta-analysis of diffusion tensor imaging studies. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *82*, 242–248. https://doi.org/10.1016/j.pnpbp.2017.11.007

D’Mello, A. M., Crocetti, D., Mostofsky, S. H., & Stoodley, C. J. (2015). Cerebellar gray matter and lobular volumes correlate with core autism symptoms. *NeuroImage: Clinical*, *7*, 631–639. https://doi.org/10.1016/j.nicl.2015.02.007

Dowd, A. M., McGinley, J. L., Taffe, J. R., & Rinehart, N. J. (2012). Do Planning and Visual Integration Difficulties Underpin Motor Dysfunction in Autism? A Kinematic Study of Young Children with Autism. *Journal of Autism and Developmental Disorders*, *42*(8), 1539–1548. https://doi.org/10.1007/s10803-011-1385-8

Edwards, L. A. (2014). A Meta-Analysis of Imitation Abilities in Individuals With Autism Spectrum Disorders. *Autism Research*, *7*(3), 363–380. https://doi.org/10.1002/aur.1379

Fatemi, S. H., Aldinger, K. A., Ashwood, P., Bauman, M. L., Blaha, C. D., Blatt, G. J., Chauhan, A., Chauhan, V., Dager, S. R., Dickson, P. E., Estes, A. M., Goldowitz, D., Heck, D. H., Kemper, T. L., King, B. H., Martin, L. A., Millen, K. J., Mittleman, G., Mosconi, M. W., … Welsh, J. P. (2012). Consensus Paper: Pathological Role of the Cerebellum in Autism. *The Cerebellum*, *11*(3), 777–807. https://doi.org/10.1007/s12311-012-0355-9

Fournier, K. A., Hass, C. J., Naik, S. K., Lodha, N., & Cauraugh, J. H. (2010). Motor Coordination in Autism Spectrum Disorders: A Synthesis and Meta-Analysis. *Journal of Autism and Developmental Disorders*, *40*(10), 1227–1240. https://doi.org/10.1007/s10803-010-0981-3

Fuelscher, I., Caeyenberghs, K., Enticott, P. G., Williams, J., Lum, J., & Hyde, C. (2018). Differential activation of brain areas in children with developmental coordination disorder during tasks of manual dexterity: An ALE meta-analysis. *Neuroscience & Biobehavioral Reviews*, *86*, 77–84. https://doi.org/10.1016/j.neubiorev.2018.01.002

Gill, K.K.; Lang, D.; Zwicker, J.G. Cerebellar Differences after Rehabilitation in Children with Developmental Coordination Disorder. Brain Sci. 2022, 12, 856. https://doi.org/10.3390/brainsci12070856

Gowen, E., & Hamilton, A. (2013). Motor Abilities in Autism: A Review Using a Computational Context. *Journal of Autism and Developmental Disorders*, *43*(2), 323–344. https://doi.org/10.1007/s10803-012-1574-0

Green, D., Baird, G., Barnett, A. L., Henderson, L., Huber, J., & Henderson, S. E. (2002). The severity and nature of motor impairment in Asperger's syndrome: A comparison with Specific Developmental Disorder of Motor Function. *Journal of Child Psychology and Psychiatry*, *43*(5), 655–668. https://doi.org/10.1111/1469-7610.00054

Guell, X., & Schmahmann, J. (2020). Cerebellar Functional Anatomy: A Didactic Summary Based on Human fMRI Evidence. The Cerebellum, 19(1), 1–5. https://doi.org/10.1007/s12311-019-01083-9

Hansen, R. L., & Rogers, S. J. (2012). *Autism and Other Neurodevelopmental Disorders*. American Psychiatric Pub.

Harris, S. R., Mickelson, E. C. R., & Zwicker, J. G. (2015). Diagnosis and management of developmental coordination disorder. *Canadian Medical Association Journal*, *187*(9), 659–665. https://doi.org/10.1503/cmaj.140994

Harrison, L., Kats, A., Kilroy, E., Butera, C., Jayashankar, A., Keles, U., Aziz-Zadeh, L. (2021) Motor and sensory features successfully decode autism spectrum disorder and combine with the original RDoC framework to boost diagnostic classification. Scientific Reports, 11(1), 7839. doi:10.1038/s41598-021-87455-w

Heijden, M. E. van der, Gill, J. S., & Sillitoe, R. V. (2021). Abnormal Cerebellar Development in Autism Spectrum Disorders. *Developmental Neuroscience*, *43*(3–4), 181–190. https://doi.org/10.1159/000515189

Henderson, S. E, Sugden, D., & Barnett, A. L. (2007). Movement Assessment Battery for Children-2 [Database record]. APA PsycTests. https://doi.org/10.1037/t55281-000

Hilton, C. L., Zhang, Y., Whilte, M. R., Klohr, C. L., & Constantino, J. (2012). Motor impairment in sibling pairs concordant and discordant for autism spectrum disorders. *Autism*, *16*(4), 430–441. https://doi.org/10.1177/1362361311423018

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6

Jung, M., Kosaka, H., Saito, D.N. *et al.* Default mode network in young male adults with autism spectrum disorder: relationship with autism spectrum traits. *Molecular Autism* 5, 35 (2014). https://doi.org/10.1186/2040-2392-5-35

Kaplan, B., Crawford, S., Cantell, M., Kooistra, L., & Dewey, D. (2006). Comorbidity, co-occurrence, continuum: What's in a name? *Child: Care, Health and Development*, *32*(6), 723–731. https://doi.org/10.1111/j.1365-2214.2006.00689.x

Khan, S., Michmizos, K., Tommerdahl, M., Ganesan, S., Kitzbichler, M. G., Zetino, M., Garel, K.-L. A., Herbert, M. R., Hämäläinen, M. S., & Kenet, T. (2015). Somatosensory cortex functional connectivity abnormalities in autism show opposite trends, depending on direction and spatial scale. *Brain*, *138*(5), 1394–1409. https://doi.org/10.1093/brain/awv043

Kilroy, E., Harrison, L., Butera, C., Jayashankar, A., Cermak, S., Kaplan, J., Williams, M., Haranin, E., Bookheimer, S., Dapretto, M., & Aziz-Zadeh, L. (2021). Unique deficit in embodied simulation in autism: An fMRI study comparing autism and developmental coordination disorder. *Human Brain Mapping*, *42*(5), 1532–1546. https://doi.org/10.1002/hbm.25312

Kilroy, E., Ring, P., Hossain, A., Nalbach, A., Butera, C., Harrison, L., Jayashankar, A., Vigen, C., Aziz-Zadeh, L. & Cermak, S. A. (2022). Motor performance, praxis, and social skills in autism spectrum disorder and developmental coordination disorder. Autism Research

Kushki, A., Chau, T., & Anagnostou, E. (2011). Handwriting Difficulties in Children with Autism Spectrum Disorders: A Scoping Review. *Journal of Autism and Developmental Disorders*, *41*(12), 1706–1716. https://doi.org/10.1007/s10803-011-1206-0

Levisohn, L., Cronin-Golomb, A., & Schmahmann, J. D. (2000). Neuropsychological consequences of cerebellar tumour resection in children: Cerebellar cognitive affective syndrome in a paediatric population. *Brain*, *123*(5), 1041–1050. https://doi.org/10.1093/brain/123.5.1041

Licari, M. K., Billington, J., Reid, S. L., Wann, J. P., Elliott, C. M., Winsor, A. M., Robins, E., Thornton, A. L., Jones, R., & Bynevelt, M. (2015). Cortical functioning in children with developmental coordination disorder: A motor overflow study. *Experimental Brain Research*, *233*(6), 1703–1710. https://doi.org/10.1007/s00221-015-4243-7

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223. https://doi.org/10.1023/A:1005592401947

Lord C, Rutter M, & Le Couteur A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism & Developmental Disorders*, *24*(5), 659–685. https://doi.org/10.1007/bf02172145

Lu, S.-C., Rowe, P., Tachtatzis, C., Andonovic, I., Anzulewicz, A., Sobota, K., & Delafield-Butt, J. (2022). Swipe kinematic differences in young children with autism spectrum disorders are task- and age-dependent: A smart tablet game approach. *Brain Disorders*, *5*, 100032. https://doi.org/10.1016/j.dscb.2022.100032

Manto, M., Bower, J. M., Conforto, A. B., Delgado-García, J. M., da Guarda, S. N. F., Gerwig, M., Habas, C., Hagura, N., Ivry, R. B., Mariën, P., Molinari, M., Naito, E., Nowak, D. A., Ben Taib, N. O., Pelisson, D., Tesche, C. D., Tilikete, C., & Timmann, D. (2012). Consensus Paper: Roles of the Cerebellum in Motor Control—The Diversity of Ideas on Cerebellar Involvement in Movement. *Cerebellum (London, England)*, *11*(2), 457–487. https://doi.org/10.1007/s12311-011-0331-9

McKinney, W. S., Kelly, S. E., Unruh, K. E., Shafer, R. L., Sweeney, J. A., Styner, M., & Mosconi, M. W. (2022). Cerebellar Volumes and Sensorimotor Behavior in Autism Spectrum Disorder. *Frontiers in Integrative Neuroscience*, *16*. https://www.frontiersin.org/articles/10.3389/fnint.2022.821109

Mostofsky, S. H., Dubey, P., Jerath, V. K., Jansiewicz, E. M., Goldberg, M. C., & Denckla, M. B. (2006). Developmental dyspraxia is not limited to imitation in children with autism spectrum disorders. *Journal of the International Neuropsychological Society*, *12*(3), 314–326. https://doi.org/10.1017/S1355617706060437

Neely, K. A., Coombes, S. A., Planetta, P. J., & Vaillancourt, D. E. (2013). Segregated and overlapping neural circuits exist for the production of static and dynamic precision grip force. *Human Brain Mapping*, *34*(3), 698–712. https://doi.org/10.1002/hbm.21467

Noonan, S. K., Haist, F., & Müller, R.-A. (2009). Aberrant functional connectivity in autism: Evidence from low-frequency BOLD signal fluctuations. *Brain Research*, *1262*, 48–63. https://doi.org/10.1016/j.brainres.2008.12.076

Oliveira, M. A., Shim, J. K., Loss, J. F., Petersen, R. D. S., & Clark, J. E. (2006). Effect of kinetic redundancy on hand digit control in children with DCD. *Neuroscience Letters*, *410*(1), 42–46. https://doi.org/10.1016/j.neulet.2006.09.065

Pangelinan, M. M., Hatfield, B. D., & Clark, J. E. (2013). Differences in movement-related cortical activation patterns underlying motor performance in children with and without developmental coordination disorder. *Journal of Neurophysiology*, *109*(12), 3041–3050. https://doi.org/10.1152/jn.00532.2012

Paquet, A., Olliac, B., Golse, B., & Vaivre-Douret, L. (2019). Nature of motor impairments in autism spectrum disorder: A comparison with developmental coordination disorder. *Journal of Clinical and Experimental Neuropsychology*, *41*(1), 1–14.

Ringold, S. M., McGuire, R. W., Jayashankar, A., Kilroy, E., Butera, C. D., Harrison, L.,& Aziz-Zadeh, L. (2022). Sensory Modulation in Children with Developmental Coordination Disorder Compared to Autism Spectrum Disorder and Typically Developing Children. *Brain Sciences*, *12*(9), 1171.

Rinehart, N. J., Tonge, B. J., Iansek, R., McGinley, J., Brereton, A. V., Enticott, P. G., & Bradshaw, J. L. (2006). Gait function in newly diagnosed children with autism: Cerebellar and basal ganglia related motor disorder. *Developmental Medicine and Child Neurology*, *48*(10), 819–824. https://doi.org/10.1017/S0012162206001769

Riva, D., & Giorgi, C. (2000). The cerebellum contributes to higher functions during development: Evidence from a series of children surgically treated for posterior fossa tumours. *Brain*, *123*(5), 1051–1061. https://doi.org/10.1093/brain/123.5.1051

Roley, S. S., Mailloux, Z., Parham, L. D., Schaaf, R. C., Lane, C. J., & Cermak, S. (2014). Sensory Integration and Praxis Patterns in Children With Autism. *The American*

Journal of Occupational Therapy, *69*(1), 6901220010p1-6901220010p8. https://doi.org/10.5014/ajot.2015.012476

Rothi, G., Raymer, A., Ochipa, C., Maher, L., Greenwald, M., & Heilman, K. (2003). Florida Apraxia Battery-revised.

Sacrey, L.-A., Germani, T., Bryson, S., & Zwaigenbaum, L. (2014). Reaching and Grasping in Autism Spectrum Disorder: A Review of Recent Literature. *Frontiers in Neurology*, *5*. https://www.frontiersin.org/article/10.3389/fneur.2014.00006

Schmitz, C., Martineau, J., Barthélémy, C., & Assaiante, C. (2003). Motor control and children with autism: Deficit of anticipatory function? *Neuroscience Letters*, *348*(1), 17–20. https://doi.org/10.1016/S0304-3940(03)00644-X

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., Diamond, S. P., & Held, R. M. (2014). Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences*, *111*(42), 15220–15225. https://doi.org/10.1073/pnas.1416797111

Sivaswamy, L., Kumar, A., Rajan, D., Behen, M., Muzik, O., Chugani, D., & Chugani, H. (2010). A Diffusion Tensor Imaging Study of the Cerebellar Pathways in Children With Autism Spectrum Disorder. *Journal of Child Neurology*, *25*(10), 1223–1231. https://doi.org/10.1177/0883073809358765

Stoit, A. M. B., van Schie, H. T., Slaats-Willemse, D. I. E., & Buitelaar, J. K. (2013). Grasping Motor Impairments in Autism: Not Action Planning but Movement Execution is Deficient. *Journal of Autism and Developmental Disorders*, *43*(12), 2793–2806. https://doi.org/10.1007/s10803-013-1825-8

Stoodley, C. J. (2014). Distinct regions of the cerebellum show gray matter decreases in autism, ADHD, and developmental dyslexia. *Frontiers in Systems Neuroscience*, *8*. https://www.frontiersin.org/article/10.3389/fnsys.2014.00092

Teitelbaum, P., Teitelbaum, O., Nye, J., Fryman, J., & Maurer, R. G. (1998). Movement analysis in infancy may be useful for early diagnosis of autism. *Proceedings of the National Academy of Sciences*, *95*(23), 13982–13987.

TREVARTHEN, C., & Delafield-Butt, J. (2013). Autism as a developmental disorder in intentional movement and affective engagement. *Frontiers in Integrative Neuroscience*, *7*. https://www.frontiersin.org/article/10.3389/fnint.2013.00049

Tunçgenç, B., Pacheco, C., Rochowiak, R., Nicholas, R., Rengarajan, S., Zou, E., Messenger, B., Vidal, R., & Mostofsky, S. H. (2021). Computerized Assessment of Motor Imitation as a Scalable Method for Distinguishing Children With Autism. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(3), 321–328. https://doi.org/10.1016/j.bpsc.2020.09.001

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLOS ONE*, *14*(11), e0224365. https://doi.org/10.1371/journal.pone.0224365

Vaillancourt, D. E., Mayka, M. A., & Corcos, D. M. (2006). Intermittent Visuomotor Processing in the Human Cerebellum, Parietal Cortex, and Premotor Cortex. *Journal of Neurophysiology*, *95*(2), 922–931. https://doi.org/10.1152/jn.00718.2005

Van Overwalle, F., Ma, Q., & Heleven, E. (2020). The posterior crus II cerebellum is specialized for social mentalizing and emotional self-experiences: A meta-analysis. *Social Cognitive and Affective Neuroscience*, *15*(9), 905–928. https://doi.org/10.1093/scan/nsaa124

Verly, M., Verhoeven, J., Zink, I., Mantini, D., Peeters, R., Deprez, S., ... & Sunaert, S. (2014). Altered functional connectivity of the language network in ASD: role of classical language areas and cerebellum. *NeuroImage: Clinical*, *4*, 374-382.

Wechsler Abbreviated Scale of Intelligence—Second Edition—PsycNET. (n.d.). Retrieved April 11, 2022, from https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft15171-000

Williams, J. H., Whiten, A., & Singh, T. (2004). A systematic review of action imitation in autistic spectrum disorder. *Journal of Autism and Developmental Disorders*, *34*(3), 285–299.

Wilson, B. N., Kaplan, B. J., Crawford, S. G., & Roberts, G. (2007). The developmental coordination disorder questionnaire 2007 (DCDQ'07). Administrative manual for the DCDQ107 with psychometric properties, 267-272.

Wilson, P. H., Ruddock, S., Smits-Engelsman, B., Polatajko, H., & Blank, R. (2013). Understanding performance deficits in developmental coordination disorder: A meta-analysis of recent research. *Developmental Medicine & Child Neurology*, *55*(3), 217–228.

Wisdom, S. N., Dyck, M. J., Piek, J. P., Hay, D., & Hallmayer, J. (2007). Can autism, language and coordination disorders be differentiated based on ability profiles? *European Child & Adolescent Psychiatry*, *16*(3), 178–186. https://doi.org/10.1007/s00787-006-0586-8

Zwicker, J. G., Missiuna, C., Harris, S. R., & Boyd, L. A. (2011). Brain activation associated with motor skill practice in children with developmental coordination disorder: An fMRI study. *International Journal of Developmental Neuroscience*, *29*(2), 145–152. https://doi.org/10.1016/j.ijdevneu.2010.12.002