Neural Correlates of Interactive Word Learning Between Preschool-Aged Children and Caregivers

Sara Mosteller

A thesis submitted in partial fulfilment of the requirements of the University of East Anglia for the degree of Doctor of Philosophy.

Research undertaken in the School of Psychology, University of East Anglia.
September, 2020

Abstract

Combined processing between caregivers and children may facilitate word learning during the preschool years, reflecting successful transfer of representations between adults and children when speaking or encoding labels. This thesis investigated the neural correlates of early word learning during caregiver-child interactions. Caregivers taught their 32- and 54-month-old children the names of eight new objects within a simpler task in which children were taught three object names, and a more complex task in which children were taught five object names. Children's comprehension was later tested by an experimenter who presented them with pairs of objects and asked the child to select each by name. The caregiver's learning was assessed in a word production test. Functional near infrared spectroscopy (fNIRS) data was obtained from both the caregiver and child during these tasks. The first empirical chapter (Chapter 2) describes the methods development and reports a piloting study. Three sets of analyses on these data are then reported. In the second empirical chapter (Chapter 3), neural responses to the differing naming events were examined for both the caregiver and child. For children, naming of learned words was different to naming of words not learned within the right temporal cortex, indicating differences in social and language related processing. Further, caregivers' neural responses differed based on their child's learning within the left parietal cortex. The second set of analyses revealed that distinctiveness of object locations during the interactions predicted which words children learned. The third set of analyses, reported within the third empirical chapter (Chapter 4), investigated whether neural coherence was greater as caregivers and children directly interacted during word learning, compared with caregivers and children who were also teaching and learning words but who were not interacting with each other. After statistically correcting for multiple comparisons, no differences in coherence were found between pairs of interacting caregivers and children, and scrambled pairings of caregivers and children. Technical limitations of the research are discussed, along with implications for broader connections to the literature and suggestions for future directions.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my examiners for providing input on this thesis, specifically, Natasha Kirkham for helpful questions about the aims and limitations of the study and Teea Gliga for a meticulous consideration of the analyses. Thank you to my supervisory team for providing guidance on the research throughout. I would like to thank John Spencer for earlier input on the design and technical methods of the study, Larissa Samuelson for contributions to the study design and methods throughout and for thoughtful views on theory, and to Nadja Althaus for practical advice during the writing up on everything from statistics to writing style. I would like to thank my colleagues, especially mentioning Laia Fibla and Prerna Aneja, for expanding my research horizons, as well as Sam Forbes, Sobana Wijeakumar and Ajaz Bhat, for helpful advice on everything from statistics and neuroimaging methods to managing postgraduate student life. A big thank you to the student interns and research assistants whose attention to every detail and innovative ideas made the methods used in the study possible, as well as for creating a 3D model figure of the testing room: by first name and in chronological order, Ellie, Jeevun, Sushila, Ellie, Kai and Ciara. Another big thank you to the staff members who assisted with the data collection for this work, including Kate Mee and Joe Cassidy. I would like to express gratitude to Damian Milton for insightful advice at a time when I felt the most lost in my field. I would also like to thank my family for their continued support. Finally, a thank you to János, for unconditional friendship, a place to finish my degree, and for tirelessly doing little things to rebuild my health and my hopes. They say that love makes the impossible, possible, and now I understand why.

Author’s Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

Parts of this work have been presented at conferences:

Conference Presentations:

Mosteller, S., Samuelson, L.K., & Spencer, J.P. Neural mechanisms of communication and cooperation between children and parents. **Poster:** fNIRS 2018, Tokyo, Japan.

Mosteller, S., Samuelson, L.K., & Spencer, J.P. Neural mechanisms of communication and cooperation between children and parents. **Talk:** symposium *Developmental Aspects*, ESCON 2018: Transfer of Knowledge, Cologne, Germany.

Mosteller, S., Samuelson, L.K., Wijeakumar, S., Forbes, S.H., Ghose Coveney, S., Grewal, J., Johns, E. & Spencer, J.P. Neural mechanisms of spatial coordination during interactive word learning in early development. **Poster:** Expanding the Field 2018: Multi-disciplinary Developmental Dynamics, Norwich, United Kingdom.

Mosteller, S., Samuelson, L.K., Forbes, S.H., Wijeakumar, S. & Spencer, J.P. Shared Space and Neural Coherence in Early Word Learning. **Talk:** symposium *Kids in space: A dynamic coupling between space and language development* at the 48th Annual meeting of the Jean Piaget Society, Amsterdam, Netherlands.

Any ethical clearance for the research presented in this thesis has been approved. Approval has been sought and granted by the School of Psychology Ethics Committee at the University of East Anglia.

Name: Sara Mosteller

Signature:

Date:

SECTION 1
–
Interactive Word Learning Between Preschool Children and Caregivers

Chapter 1 – General Introduction: Interactive Word Learning Between Preschool Children and Caregivers

Hoff (2006) argued that, observably, social environments provide resources for children's vocabulary development above and beyond exposure to an analysable language model. These resources, as broadly described in the psychological literature, include caregivers' responsiveness to and contingency with the child's behaviour, joint attention, and explicit naming of objects. For example, a caregiver may name an object they notice their child is already looking at, establishing a common focal point between them and their child, and providing a naming input that is consistent with the child's visual focus. A survey of the relevant literature reveals that social connection facilitates children's word learning in two ways. First, caregivers and children share experiences of the outside world during play, which ultimately makes learning new words more salient and motivates learning language as a communicative tool for the child. Second, when interacting, caregivers and children actively shape language learning environments in ways that help the child make clear mappings between new words and their intended referents.

1.1 Caregiver-Child Interactions and Shared Experiences

At a fundamental level, sharing experiences may motivate early language-learning (Steels & Kaplan, 2000). Research that spans theoretical perspectives and methodologies suggests that, even from infancy, children are sensitive to the difference between sharing an experience of their environment with another person and experiencing their surrounding environment alone (Beier & Spelke, 2012; Carpenter et al., 1998; Henderson et al., 2002; Striano, Reid & Hoehl, 2006). Many studies of very early development have focused on joint attention, or the sharing of a common focal point between two or more individuals during an interaction. Typically, this common focal point might be an object that both individuals are exploring simultaneously. For example, an adult and toddler may be viewing the same object during stimulus presentations in which the object is named. It has been argued that, phenomenally, this is different from viewing the object alone because an individual's experience is more salient based on knowing that another person is focused on the same object with them (Zahavi, 2015). To support this hypothesis, children seem to be driven to learn new words in order to direct others' attention and to share attention with others (Hoff, 2006; Tamis-LeMonda, Kuchirko & Song, 2014). This is accomplished by eliciting the sharing of attention and other contingent social inputs that occur when they are focused on referents with their caregivers.

What is harder to answer than whether sharing attention occurs and why it motivates children to communicate, is how sharing attention directly changes anything about the cognitive processes underlying word learning. Mundy and Jarrold (2010) hypothesized that sharing attention with another person may enhance encoding of information based on the properties of neural networks alone. This is based on the principle that simultaneous activation of multiple networks increases the depth of encoding of external stimuli (for research supporting this theory, see Munakata & McClelland, 2003; Otten, Henson & Rugg, 2001). Sharing attention activates a dorsal, frontoparietal attention system involved in self- and other-referenced information processing, and this network may enhance simultaneous activation within visual and language processing networks. For example, children's encoding of an object, word or other stimuli may be enhanced via the broader neural response to sharing attention with their caregiver.

According to this hypothesis, activation within frontal and parietal attention areas may facilitate word learning and retention. Functionally, these areas are activated during episodes of shared attention as early as 14 months (Henderson et al., 2002). In support of the hypothesis that activation within this system might generate a more complex neural signature of learning, sharing attention during learning elicits a longer neural response over the left hemisphere during referent selection as early as 18 months. Hirotani et al. (2009) compared event-related potentials as toddlers 18-21-months old saw correctly labelled or incorrectly labelled objects for newly taught words. Toddlers were taught the object names in either a "social" condition while an experimenter made eye contact and spoke in a positive voice, or in a "non-social" condition while the experimenter did not make eye contact and spoke in a neutral tone. Both groups of toddlers showed evidence of having made associations between these taught words and the objects they represented via an early-going difference in response to congruent and incongruent pairings over the left inferior frontal gyrus (IFG). However, only toddlers who had been taught the object names in the social condition showed a later-going difference in waveforms when later presented with objects and names that did not match, over left temporal electrode sites. This was attributed by the authors to a failure to semantically integrate the incongruent word-object pairing after learning the original in a social context. By contrast, learning the exact same mappings without sharing attention did not result in this effect.

Convergently, sharing experiences may compensate for times when caregivers and children are not behaviourally in alignment with each other. While caregiver-child coordination can help children make correct word-object mappings in many cases, some research suggests that it's not a perfect system. In a review, Samuelson, Kucker and Spencer (2017) point out that sometimes caregivers name different objects to those their child is attending to, resulting in a lack of consistency between the visual and speech inputs the child receives concurrently (Pereira et al., 2014; Tomasello & Farrar, 1986; Yu & Smith, 2012). However, if these inputs are more salient when attention is shared compared with when it is not, children may then be biased to encode word-object mappings specifically when their caregiver is sharing attention to a referent with them. As a result, this bias may partially filter out incorrect mappings made when children are focused on a different object or aspect of their environment from what their caregiver is referring to when naming.

Another mechanism through which sharing experiences enhances language comprehension, and potentially word learning, is that it entrains perspectives between conversation partners. Individuals can begin a dialogue in very different places, but the sooner their perspectives fall in synchrony, the better they may be able to process each other's meanings. For example, processing meaning in another person's speech during dialogue can be bolstered by lower-level coupling at non-linguistic, sensorimotor levels. Garrod and Pickering (2009) argue that conversation is a form of joint action that relies on coupling at non-linguistic levels such as posture or speech rate, in order to align processing required at higher level, linguistic levels. The effects of representational alignment can be seen, for example, when conversation partners shift gaze to look at elements of a picture in synchrony rather than with the delay that can be seen as a listener follows a monolog (Richardson, Dale & Kirkham, 2007). This study indicates that when representations are in sync, language processing is more seamless as individuals make predictions about the objects or nouns their conversation partner refers to, orienting to referents along with, rather than after, the speaker.

1.2 Caregivers and Language-Learning Environments

The previous section revealed that more broadly distributed activation that includes, but is not limited to, language networks appears to facilitate word learning above and beyond a specialized language processing system alone. However, it can't specify in practical terms how contextual processing facilitates solving everyday word learning problems. In this section, this aspect is examined: namely, how speakers can direct children's mapping of new words to their correct referents through building on the statistical properties of the child's visual environment. In an environment with potentially hundreds of objects, caregiver-child interaction helps constrain a name to one specific object fairly quickly, leading to a rapid mapping between the two (Jesse & Johnson, 2016). Because children are sensitive to ostensive communicative cues from adults such as looking to and talking about objects, and these cues then positively affect their learning (Wu & Kirkham, 2010, see Sobel & Kirkham, 2012), social communication provides an additional basis of information from which children can directly draw inferences about the meanings of new words (see Tamis-LeMonda et al., 2014).

From the child's perspective, caregivers shape early language-learning environments through providing coherent visual and auditory inputs. In an environment that contains numerous objects, caregivers directly hold, point to or talk about objects in-context as they name them (Cartmill et al., 2013; Kalagher & Yu, 2006; Yu & Smith, 2013; Yu et al., 2009). For example, in environments that contain many possible referents for each spoken new word, adult speakers may name an object specifically when the child is focused on it, narrowing the number of possible real-world referents to a single focal point. To support this hypothesis, data from head-mounted cameras revealed that caregivers spontaneously named toys that their child was focused on more frequently than objects they were not (Suanda et al., 2016). Previous research suggests that children learn more when adults name objects that the child is looking at (Tomasello & Farrar, 1986), indicating that, by creating a direct correlation between what the child is seeing and the name they are hearing, social partners use attention to help children learn new words. Social interaction can help to regulate attention toward referents in order to facilitate retention of new mappings as well. From very early in development, caregivers talk about objects to help children sustain focus on them (Suarez-Rivera, Smith & Yu, 2019), potentially resulting in increased depth of object encoding and a more robust association between the word and object. These examples show that behavioural coordination between adults and children allows children to make rapid and accurate mappings between words and objects while exploring their environment, based on narrowing the space of possible referents to one. Previous studies examining free play between caregivers and children suggest that behavioural coordination emerges spontaneously during interactions (Kalagher & Yu, 2006; Suarez-Rivera et al., 2019; Yu & Smith, 2013) and can predict children's learning of new nouns (Samuelson et al., 2011).

As the previously discussed examples show, the gateway through which interaction impacts on word learning may be through influencing a variety of processes that facilitate making accurate word-object mappings (Samuelson et al., 2011; Suarez-Rivera et al., 2019; Sullivan et al., 2015). This concept can extend beyond explicitly naming objects and establishing joint attention, to facilitating binding between words and objects based on visuospatial organization as objects are named. Previous work suggests that auditory information related to events is indexed to the corresponding visual information via spatial location (Richardson & Kirkham, 2004) suggesting that children encode spatial information when playing with new objects. Further, spatial encoding may enable children to draw associations across multiple naming events through time. Spatial predictability in where objects are located during interactions primes children to make word-object mappings through boosting children's attention to objects based on their locations and through deepening encoding of word-object mappings within working memory (Benitez & Smith, 2012; Samuelson et al., 2011). This may be because children are faster to orient initially to objects that are named when they appear in a consistent spatial location as opposed to when their locations shift around.

Corresponding modelling work most notably shows that recurrent naming of an object in the same location results in a stronger word-object association than naming of an object in different locations, because object locations are encoded when objects are presented. Using the encoded spatial location can help resolve ambiguities in naming; for example, when children heard a new word, they were able to map the name to an absent object when the speaker tapped or snapped their fingers over the table at the location it had previously occupied. During naming of objects, a working memory trace develops that can then be used to associate the new word with a single referent based on indexing these together at the location

(Samuelson et al., 2011), a mechanism which can explain these experimental findings with toddlers.

Spatial predictability of object locations has been correlated with word learning during openly structured interactions in which caregivers taught their children new nouns, suggesting that caregivers use spatial structure to teach their child. When caregivers were given two objects and asked to teach their 17- to 22-month-old children the names, the percentage of time that caregivers consistently held each object in their right or left hand was correlated with better performance in a subsequent test of learning (Samuelson et al., 2011). Critically, the same factor that had predicted children's learning after being directly manipulated in a highly controlled study also predicted children's learning in an openly structured interaction with their caregiver.

The preschool years are characterized by marked changes in cognitive development (Kyttälä et al., 2014; Newbury et al., 2016; Vlach & DeBrock, 2017; 2019), language experience (Grassmann, Schulze & Tomasello, 2015; Kalashnikova, Mattock & Monaghan, 2015; Law et al., 2017), and social development (Saxe, 2013). On the one hand, many of the same processes observed in toddlers continue to facilitate learning new words during these years. For example, visual joint attention with an adult speaker remains pivotal to word learning at four to five years of age (Sullivan et al., 2015), though preschool-aged children show more initiative than toddlers in actively coordinating their attention with an adult speaker in order to learn. Likewise, between two and five years of age, individual differences in performance during working memory tasks are correlated with verbal processing, while visuospatial processing is related to vocabulary (Alloway, Gathercole & Pickering, 2006). Children also bring emerging new skills to word learning interactions, many of which can be characterized as explicitly inferring what infants and toddlers may rely more on caregivers to facilitate. For examples, children's participation in dialogue about new referents facilitates their word learning, while during naming, they can utilize explicit strategies like a mutual exclusivity bias to help resolve ambiguities in specific naming instances (Grassmann et al., 2015; Kalashnikova et al., 2015; Lewis et al., 2020).

In the preschool years, children are more sensitive to the significance of being taught new information and may attend selectively to being intentionally informed by an adult. The ability to determine what is pedagogically valuable emerges during these years, as prior work suggests that children ages 3-7 show variable responding to teaching-related cues within social interactions, varying based on the testing cohort and task (Bonawitz et al., 2011, Sobel & Corriveau, 2010; see also Sobel & Kirkham, 2012). Previously, this development has been demonstrated as children learn to make causal inferences that are required to understand the function of a new machine. For example, preschool-aged children may explore a new toy less on their own after being taught about it by an adult, as they perceive that they have fewer aspects to investigate on their own (Sobel & Sommerville, 2010). Interestingly, the study suggests that preschoolers are responsive to pedagogically relevant information in interactions with an adult, regardless of whether teaching them was intended. Children demonstrate further biases toward learning from interactions in which an adult is explicitly teaching (Sobel & Sommerville, 2009) and in which the information being taught is consistent with the data they have accumulated from their own observations (Sobel & Buchanan, 2009). Collectively, this work reveals that preschoolers directly respond to explicit teaching from an adult and also compare this teaching with beliefs from their own experience. This is consistent with research showing that 4-year-olds, but not 3-year-olds, discriminately learn word-object mappings when they believe an adult speaker is knowledgeable about the properties of objects that determine their label category (Sobel & Corriveau, 2010).

1.3 Neural Correlates of Interactive Word Learning

The following section explores how coordination between adults and children may enhance the child's underlying neural encoding of word-object mappings as new objects are named during interactions. Dynamic neural field models of word learning show that when hearing a name consistently co-occurs with seeing an object, a memory trace builds up for the pairing between the two, boosting activation for that specific pairing and simultaneously inhibiting the pairing of that name with other objects (Bhat, Spencer & Samuelson, 2018; Samuelson et al., 2011). By contrast, incorrect or inconsistent mappings can be made when an association drawn between one word-object mapping interferes with another mapping or when there is inconsistency in which objects and words are indexed together. When the child is learning consistent word-object pairings during interactions, a signature may appear in neural activation during naming that indicates the semantic association being made. This signature may be predicted from prior, computer-based tasks.

The neural correlates of semantic processing include distinct but interconnected networks. These include a left-lateralised language processing network, and activation within cortical networks that implement attention and working memory (Xu et al., 2016). Of these, the language processing network has been shown most consistently to be activated in language processing tasks. This network extends from Broca's area in the left inferior frontal gyrus (IFG), to Wernicke's area in the left posterior temporal lobe (Balsamo, Xu & Gaillard, 2006; Chou et al., 2006; Fiebach & Friederici, 2004; Grindrod et al., 2008; Hirshorn & Thompson-Schill, 2006; Hofstetter, Friedmann & Assaf, 2017; Liljeström et al., 2009; Romeo, Leonard et al., 2018; Wong et al., 2011; see also Fedorenko, 2014). Unsurprisingly, the language network is activated during processing of new words (Davis et al., 2009; López-Barroso et al., 2013; Nora et al., 2017; Takashima et al., 2019). For example, Takashima et al. (2019) trained 8- to 10 and 14- to 16-year-old children 30 new Japanese words, such as *tanuki*. Children were presented with word-picture pairings on the first training session, along with descriptive sentences such as "It likes to trick people." Children were then scanned one day later and again eight days later, shown sets of four pictures, and asked to select which was the match for the new word. These words were intermixed with pseudo-Japanese words that they had not been exposed to before, and in order to assess learning, they were also asked whether each stimulus was an existing word. As an additional control, children were presented with Dutch words (all children were Dutch speaking). A main effect of language was found; Japanese words elicited a stronger response in the left IFG area, whereas Dutch (familiar) words elicited a greater response within the posterior temporal lobe and inferior parietal lobe. This indicates language processing areas respond differently between hearing new or familiar words. Interestingly, no differences were observed in this paradigm between recently learned and pseudo words, limiting the scope of these effects to explain what activation is specific to word learning. It is important to note that, for children, language processing is less left-lateralized than it is in adults and can also be localized to the right inferior frontal gyrus (IFG, Berl et al., 2014; Nora et al., 2017). Even for adults, bilateral IFG activation may better help to hone language processing around information that is most relevant within noisy environments (Adank, Davis & Hagoort, 2012).

Attention, cognitive control and working memory are relevant processes involved in semantic cognition, that are all thought to support how word-object mappings are integrated, moment-to-moment. Among other things, they enhance the speed of processing and encoding of relevant stimuli necessary to map words to objects (Fedorenko, 2014). Working memory, for example, has been associated with vocabulary acquisition (for a review, see Gupta & MacWhinney, 1997) and with learning of news words (Archibald & Joanisse, 2013). Most frequently, attention-orienting and attention-regulating networks are activated during sentence processing and new word-learning tasks (Berl et al., 2014; Ferreira et al., 2015; Liljeström et al., 2009; López-Barroso et al., 2013) or these areas show structural plasticity during children's training in learning new words (Ekerdt et al., 2020).

Attention and working memory have been shown to be critical for children's language processing (Berl et al., 2015; Nora et al., 2017). When 4- to 12-year-old children were asked to evaluate whether statements about concrete nouns were true during an fMRI scan, they showed increased activation relative to baseline in left-lateralized language processing areas, such the left IFG. However, they also showed activation within the right IFG, the superior frontal gyrus, posterior cingulate cortex and the anterior cingulate cortex, areas generally associated with attention (Berl et al., 2014). Thus, attention appears to be directly involved in language processing as children evaluate descriptions of everyday nouns.

Attention also plays a role in learning new nouns for adults and especially for children. In one study, adults were trained on 40 sets of imaginary word-item pairings and then scanned while asked to classify these words and everyday words as belonging to a living or a non-living category. Their neural responses showed activation not just within left-lateralized language processing areas, but within attention areas such as the superior frontal gyrus and superior parietal cortex, that responded differentially to categorizing familiar versus recently learned words (Ferreira et al., 2015). While this network tends to be treated as a background support in most word learning studies, an unexpected finding from Ekerdt et al. (2020) suggests that the dorsal attention system may be primarily responsive to early word learning. Children were scanned in a structural imaging paradigm before and after completing eight weeks of learning nonsense word-animal pairs. While myelin density within the left posterior temporal cortex prior to training was associated with children's word-learning during the study, only the precentral gyrus showed changes in myelin density *after* training. Micro-structural plasticity within this area was interpreted as indicating that it was the dorsal attention system, not the language processing network, that changed the most as a result of the training. When putting together computational modelling and neuroimaging approaches, it appears that attention and working memory may functionally play a role in word learning. It remains an open question, whether distinct differences within these processes can be observed at a neural level based on whether word learning is taking place.

While the cognitive building blocks of attention and working memory enable children to encode word-object pairings, further work suggests that in interactive contexts, social computations that occur while observing a conversation partner impact on the neural responses to hearing objects named. Egorova, Shtyrov and Pulvermüller (2016) found that a neural network underlying the interpreting of actions was activated when processing the name of a new object after participants observed one person asking another what they wanted them to retrieve. By contrast, a region involved in semantic processing was activated when they previously observed the same person asking what these objects were called. What the first speaker was initially asking determined how adults processed the exact same object name from the second speaker. This indicates that different social computations that are tied to different interpretations of the speaker's intent may alter how an object name is processed. In another study with children, Rice, Moraczewski & Redcay (2016) presented 7- to 13-year-olds with video clips of an experimenter saying "The following items are on the breakfast menu: pancakes and fruit. I am trying to eat healthy, which one should I pick?". In one condition, children believed that the video feed was live and in another, they were told it was pre-recorded. During the "live" condition, children showed activation within the temporo-parietal junction, an area involved in perspective-taking, that was not present when they did not perceive that they were interacting. Therefore, children responded to the demands of perceiving their interaction was live with more effortful processing of the speaker's intentions. During the preschool years, the development of reasoning about others' mental states is critical (Wellman, Cross & Watson, 2001) and this ability is robustly present when mapping words to objects by four years of age (Sobel & Corriveau, 2010). This indicates that social computations about the caregiver's intention when teaching new words may change throughout these years, a trajectory prior literature has not yet mapped.

Finally, an open question remains: whether the caregiver's brain activity when naming new objects as they teach their child is different when their child is learning (i.e., when the teaching is successful) compared with then they are not. Previous work reveals that caregivers' responsiveness as teachers impacts directly on children's learning (see sections 1.1 and 1.2). Therefore, it would make sense that caregivers, as well as children, encode the environment as they name objects differently when they are successfully teaching their child. However, no studies to date have investigated the neural correlates of teaching new words. As will be more thoroughly discussed in the next section, previous findings indicate that relationships in brain activity between adults and children can be observed when they are interacting face-to-face (Piazza et al., 2020), and that alignment in brain activity between a speaker and listener predicts the detail of speech comprehension (Stephens, Silbert & Hasson, 2010). Like behaviour, the neural signatures of word learning can be explored from a social network perspective, as a combination of the caregiver's and child's brain activity while word-object mappings are made. For example, caregivers may respond to sharing attention with their child, may process what their child is looking at or saying more effortfully, or may perspective-take more, when successfully teaching new words.

Putting the pieces together, the previous literature is clear that language and other cognitive processes play a role in learning words, while leaving open more questions of how environmental context and social computations impact on word learning during a live interaction between caregiver and child. Fundamentally, previous work shows that the inferior frontal gyrus plays a pivotal role in processing nouns (Balsamo et al., 2006; Chou et al., 2006; Fiebach & Friederici, 2004; Grindrod et al., 2008; Hirshorn & Thompson-Schill, 2006; Liljeström et al., 2008; Romeo, Leonard et al., 2018; see also Fedorinko, 2014) and is also fundamental to word learning (Davis et al., 2009; López-Barroso et al., 2013; Nora et al., 2017; Takashima et al., 2019). Therefore, during interactions, this area may be activated as the child maps new words to objects. In addition, language processing has been shown to rely on cognitive processes such as attention and working memory (Archibald & Joanisse, 2013; Berl et al., 2014; Ekerdt et al., 2020; Ferreira et al., 2015). In an interaction, these processes are likely to be crucial for the child to hone attention to naming events and to remember word-object pairings as multiple objects are being dynamically explored. Finally, these processes do not operate in isolation when examining the neural responses to nouns within a (perceived) live interaction (Rice, Moraczewski et al., 2016), and in situations that involve referential ambiguity (Nieuwland, Petersson & Van Berkum, 2007). Processing of spoken nouns during a real-world interaction also requires social computations about the context in which these names are embedded, what the speaker is referring to and why. The neural correlates of social computations may be observed within the child's and caregiver's neural response to naming as caregivers teach their children new words, in a way that has not been observed within isolated, computer-based word learning tasks.

1.4 Hyperscanning and Neural Coherence During Interactions

Hasson et al. (2012) argue that individual brains can be studied not just as isolated processors, but as individual nodes embedded within a larger social network. The advantage of this approach is being able to observe how brain activity between two or more communicating individuals become mutually shaped by inputs from each other. For example, during a dialogue, the brain activity of one conversation partner has been shown to temporally entrain with that of the other as they exchange linguistic and non-verbal information, indicating that their underlying representations grow in alignment via the dynamics of the interaction. This may create neural variables that emerge at a more holistic level than individual social computations (Scholkmann et al., 2013), such as quantitatively observable relationships in brain activity that indicate when individuals are in alignment. A second advantage is that one can determine how underlying alignment in brain activity between two or more interacting partners sets the conditions for transferring information from one person to another. Nowhere is the hypothesis that alignment in brain activity facilitates information transfer better established than in the domain of verbal communication. Sharing experiences between caregiver and child may predict the child's overall word learning, and this can be approached through examining relationships in the caregiver's and child's brain activity continuously *throughout an interaction.*

Some prior work suggests that neural synchrony occurs during verbal communication. For example, a previous study revealed that temporal correlations in brain activity were associated with speech comprehension. Stephens et al. (2010) fMRI-scanned one individual as they told an unrehearsed story in English and another individual as they told an unrehearsed story in Russian. English-speaking participants were then scanned as they heard the stories, and their neural signals were correlated with those of the speaker who told a story in a language they comprehended and the speaker who told a story in a language they did not. In the English story condition, three types of correlations were found. Neural signals between the speaker and listeners were correlated, time point to time point, within lower-level auditory processing areas and within language processing areas, suggesting that the speaker and listener shared common semantic processing. Second, the speaker's and listeners' neural signals were correlated with a speaker-led lag over posterior, non-linguistic areas such as the temporo-parietal junction and praecuneus, and third, correlated with a listener-led lag within frontal areas such as the dorsomedial and dorsolateral prefrontal cortex. This suggests that an interplay exists in context-driven processing between the speaker and listener. Critically, the correlations within listener-led areas were in turn correlated with the detail in which the story was later remembered, indicating that they reflected the depth of speech processing. By contrast, correlations only occurred within lower-level auditory areas when non-Russian speaking participants listened to the story told in Russian, indicating that the correlations observed in the English condition reflected speech comprehension rather than auditory perception.

Multiple other studies have shown that local, phased-locked brain activity measured by fNIRS is correlated across frequency ranges that are relevant to making decisions in an experiment, as participants interact with each other (Cui, Bryant & Reiss, 2012; Liu et al., 2016; 2017; 2019). Hyperscanning refers to a method of scanning participants simultaneously as they are interacting and analysing data with the intention to observe relationships in brain activity between individuals. For example, a dyad may be engaged in a conversation, or in game play, as fNIRS data is collected from each individual (for a review of fNIRS hyperscanning, see Scholkmann et al., 2013). For example, Cui et al. (2012) scanned participants while they each pressed a button in response to a visual cue presented on a computer monitor. In a cooperative condition, participants timed their presses as closely as possible to each other. In a competitive condition, participants tried to be the first to respond. Finally, in an alone condition, each player passively observed the other playing the game. Brain activity was measured simultaneously from both participants over frontal cortical sites. Neural coherence was computed as the mean cross-correlation of transformed signals within corresponding channel[1] pairings, for each dyad, within each condition, above and beyond the coherence observed in these channels over a period of rest (for a technical review of the WTC analysis, see Grinsted et al., 2004). Greater neural coherence was found during cooperation as opposed to competition over a frequency range of 0.08-0.33 Hz, corresponding to periods ranging from 3-13 seconds, and was also found to be higher during any sort of interactive play as opposed to individual play within this frequency range. Significant values based on measurements from the scalp corresponded to the right superior frontal gyrus. By contrast, the actual response times were closer between participants in the competition condition. Thus, this area may be involved in social processing during live interaction.

Using the same wavelet transform coherence (WTC) analysis, neural coherence has been found within a number of different interactive tasks. Coherence has been observed between dyads during interactive Jenga play (Liu et al., 2016), when jointly completing a visual working memory task (Dommer et al., 2012), when making live eye contact (Hirsch et al., 2017), when engaged in an imitation task (Holper, Scholkmann & Wolf, 2012) and when interacting within complementary leader-follower roles (Jiang et al., 2015). Neural coherence within a mirroring network (intra-parietal sulcus, inferior frontal gyrus, and posterior temporal cortex) was observed during turn-based game play (Liu et al., 2017). Finally, neural coherence has been shown to be sensitive to individual differences based on relationship status (Pan et al., 2017), indicating that some advantages may also exist for caregivers and children who have years of interactive history.

Finally, neural coherence has been directly related to participants' comprehension of concrete descriptions of nouns and objects. Liu et al. (2019) systematically varied the communication mode (face-to-face, face-to-face without eye contact, or back-to-back) along with the syntactic consistency of presented stimuli. Dyads were alternately presented with images and partial sentences and, based on the image, asked to fill in the sentence. For example, Participant A was shown a picture of a cowboy tossing a sailor a book, and the sentence beginning "A cowboy hands a sailor...." and was asked to read the complete sentence to participant

---

[1] Within fNIRS neuroimaging, a channel refers to the measurement space between the optode source that emits infrared light and the optode detector that measures reflected light from the scalp.

B, “A cowboy hands a sailor a book”. Participant B was then shown an image and pressed a button to confirm whether it matched the image described; fifty percent of the time, the picture matched and 50% of the time it was inconsistent. In a consistent syntax condition, Participant B was then shown a picture and a sentence beginning like “A cowboy hands a sailor...” and the process was alternated, whereas in an inconsistent syntax condition, they were shown a sentence with a prepositional object only, i.e., “A cowboy hands a book”. If neural coherence was influenced by syntactic consistency and not merely face-to-face interaction, then coherence should differ based on whether the sentence structure presented to both participants matched.

Within this study, brain activity was measured bilaterally over the sylvian fissure (dividing the temporal lobe from frontal and parietal cortices), and the frequency range with maximum coherence for each channel was explored between periods of 14 and 100 seconds. Over the right posterior superior temporal cortex (pSTC), greater coherence was found between syntactically consistent sentences compared with other conditions. Coherence within this channel was positively correlated with accuracy for picture-description matching within this condition. Greater coherence was also observed during face-to-face interactions within the right temporo-parietal junction (TPJ), indicating that mentalizing might be more aligned between participants during live interactions. These effects were not found for scrambled participant pairs, and their relationship to communication remained significant when the global mean of coherence increase was included as a covariate into the statistical model for actual pairs. While the effects of this study do not consistently answer the original questions (is coherence greater when communication has consistent properties and does coherence predict communication quality), temporal alignment within the right posterior temporal lobe was instrumental to resolving inconsistencies in expected and heard syntax during dialogue. It is possible that a similar effect could occur with the introduction of new words into dialogue.

Neural coherence has previously been observed between adults and children during a variety of tasks, including game play (Miller et al., 2019; Reindl et al., 2018), and openly structured face to face interactions (Piazza et al., 2020). Piazza et al. (2020) tested infants 9- to 15-months of age, and an experimenter, in a together and apart condition. In the together condition, the infant and adult played together with items like a picture book and toys. In the apart condition, the child interacted with their caregiver while the experimenter told a story to another adult. Brain activity was recorded over the prefrontal and parietal cortices and the temporoparietal junction. Temporal relationships in brain activity were measured in each of these conditions by correlating the unprocessed signal within each channel measured from the experimenter with that of the child, time point to time point. Coherence in observed data was additionally validated by testing correlations from phase-scrambled data for every channel combination and using a bootstrapping procedure to estimate 20,000 of these combinations.

During the “together” condition in which the experimenter faced and interacted with the infant, differences in inter-subject correlation were found between prefrontal channels on both the experimenter’s and child’s head, between prefrontal areas on the experimenter’s head and parietal areas on the child’s head, and between prefrontal areas on the child’s head and one central and one parietal area on the experimenter’s head. Interestingly, the correlations between signals were predominately child-led by around three seconds, suggesting that coherence may have been driven predominately by the responsiveness of the adult to the child during the interactions. The authors argued that neural coherence that occurs between infants and adults during live interactions likely sets the conditions for language learning.

Studies with older children show that neural coherence is associated with cooperation toward a common goal between adults and children. Miller et al. (2019), tested 8- to 12-year-old children and their caregivers with the cooperative and competitive and independent button press tasks used by Cui et al. (2012). Greater coherence in the prefrontal cortex was found during the cooperative play, as opposed to the independent play. Similarly, Reindl et al. (2018) tested 8- to 13-year-old children in an adapted version of this paradigm. Children pressed a button to cause a dolphin to rise and catch a beach ball. During the cooperation condition, if their avatar and their caregiver’s avatar caught the ball simultaneously, they jointly won the trial. In the competition condition, their avatar had to catch the ball first. Children also played these games with an experimenter who they did not know. Brain activity was measured over frontal cortical sites. Coherence within each channel was measured as the number of timepoints that exceeded a threshold for chance, based on first establishing null levels of coherence in channel pairings from adults and children who were not interacting. Additionally, only channels with significantly greater frequencies of these above-threshold values within one condition, compared with the mean of a null distribution of 1200 scrambled channels, were analysed. Within the dorsolateral prefrontal cortex and frontopolar cortex, neural coherence was found between caregivers and children during cooperative game play. By contrast, no coherence above and beyond the baseline was found between children and strangers in any condition, or between caregivers and children when they were simply trying to press the button faster than the other person.

The above examples show that neural synchrony between interacting partners is associated with verbal processing, cooperation and face-to-face, child-adult interactions. However, prior empirical work has not established if and how neural synchrony is related to learning new words. Wass et al. (2020) list three potential mechanisms for how communicative behaviours could entrain oscillatory activity between interacting participants and sustain it through time. The first is through behavioural entrainment, or getting in synchrony with another person via shared patterns of movement, etc. The second is ostensive cues, or behaviours that explicitly indicate communicative intent. These behaviours may act as “edges”, resetting underlying phasic activity between the sender and receiver and establishing a common pattern between them. As an example, if a child and caregiver establish eye contact, this may function as an ostensive cue. The third mechanism is higher-level alignment that occurs via self- and other referencing, shared understanding and comprehension. Thus, within caregiver-child interactions, the behaviours that establish shared experiences may also establish neural coherence.

Additionally, the authors sketch out a possible mechanism for how neural synchrony could directly support early word learning, based on the hypothesis that specific learning events, like naming an object, can be influenced by the polarity of underlying neural activity. That is, some work suggests that the net shifts in oscillatory activity may be systematically related to the excitatory/inhibitory polarity of underlying neuronal activity (Busch, Dubois & VanRullen, 2009; although see Ruzzoli et al., 2019). The authors posit one could expect that if the child’s brain activity was properly timed to the naming of objects, these naming events would be more thoroughly encoded than if they are randomly timed (for related work on perceptual awareness, see Busch et al., 2009; Mathewson et al., 2009). According to this logic, coherence between the caregiver and child may intuitively improve the timing of naming such that the child makes more word-object mappings in response to the caregiver's naming of objects, specifically when the caregiver and child are in alignment.

Collectively this research and theoretical work reveal that neural synchrony can be observed between an adult speaker and listener and also between adults and children who are interacting. Additionally, temporal relationships in brain activity observed through these measures are sensitive to both cooperation and to speech comprehension. Less research has mapped neural synchrony to free play (i.e., an openly structured interaction) between adults and children (Piazza et al., 2020), and to the best of our knowledge, no research has directly explored neural *coherence* between adults and children during a naturalistic interactive task, or during any sort of word learning task. Therefore, while much work suggests that neural coherence could underlie word learning between caregivers and children based on its relevance to other tasks, and some work has suggested possible theoretical mechanisms for this relationship (Piazza et al., 2020; Wass et al., 2020), the literature has not directly established whether neural coherence between caregivers and children is associated with children's word learning during play.

1.5 Outstanding Questions

While much work has been done to map, in detail, how behavioural coordination and shared attention shape early word learning, the neural mechanisms are significantly less understood. While differential processing of new and familiar words has been investigated for young children within controlled laboratory tasks (Takashima et al., 2019), to the best of our knowledge, no functional imaging study has explored differences between learned words and words not learned within a live interaction between a caregiver and child. One outstanding question is: while their caregiver teaches them, how are children's neural responses to an adult's naming of objects they later show evidence of having learned (learned words), different from their responses to naming of objects they do not later show evidence of having learned (words not learned)? Word learning during a live interaction requires language, attention, working memory and social computations, and it is unknown how these different computations are activated for children when hearing new words.

In general, the neural correlates of teaching new object names have not been previously studied, and specifically, no one has examined caregivers' neural responses as they teach their child object names. One unanswered question is whether the caregiver's neural activation while naming objects that their child learns is different than it is to those their child doesn't learn, based on differences of processing that lead to more effective teaching. Caregivers may be more effectively engaged in language, cognitive or social computation during instances of naming that result in their child's learning. Differences in the effectiveness of their communication may be detected in their own neural data.

Another question is how encoding of new object names is influenced by behavioural coordination that occurs between the caregiver and child during interactions. Computational modelling work suggests that behavioural coordination could directly impact on neural responses to naming of new objects during interactions between caregivers and children (Samuelson et al., 2011). But how do caregivers connect with their child and what behaviours underlie the differences in responses to learned versus words not learned during interactions? Previous work has revealed that behavioural coordination has a clear function of helping children map each new word with a specific referent in space and time (Samuelson et al., 2011). Therefore, spatial organization may be one place to start in investigating how coordinating shared environments enables children to encode word-object mappings.

Finally, it is an open question whether neural coherence between the caregiver and the child while teaching and learning object names is correlated with the child's word learning (Piazza et al., 2020; see also Wass et al., 2020). In other words, can neural synchrony between the caregiver and child indicate whether they are in alignment enough for the child to learn? Based on previous neuroimaging studies that show that neural coherence predicts speech comprehension and speech-picture matching (Liu et al., 2019; Stephens et al., 2010), and from developmental research that shows neural coherence can be observed during interactions between adults and children (Miller et al., 2019; Piazza et al., 2020; Reindl et al., 2018), it is possible that neural coherence is positively correlated with children's word learning during face-to-face interactions with a caregiver.

1.6 Applications of Functional Near Infrared Spectroscopy (fNIRS)

Prior to discussing the aims and design of the current study, a brief introduction is provided to the neuroimaging technology used to measure brain activity in this thesis. Within neuroimaging paradigms, functional near infrared spectroscopy (fNIRS) measures the density of infrared light that has been passed through the scalp, skull and cortical tissues. NIR light (695-1000 nm) is passed through these tissues by an optode on the surface of the participant's scalp (a source) at a wavelength that is relatively transparent to these tissues, and that is predominately absorbed by the chromophore of hemoglobin proteins (see Ferrari & Quaresima, 2012). In the present apparatus, optical density that is measured by an additional optode positioned around three cm from the light source (a detector) can be converted using the modified Beer-Lambert law (Scholkmann et al., 2014) to indicate levels of hemoglobin concentrated within the microcircuitry of the cortical tissue. Because chromophores vary based on the level of oxygenation, often two or more wavelengths are emitted to measure the concentration within oxygenated vessels and deoxygenated vessels in these tissues (for reviews, see Ferrari & Quaresima, 2012; Scholkmann et al., 2014). From this point forward, oxygenated hemoglobin will be referred to conventionally as HbO, deoxygenated hemoglobin will be referred to as HbR, and the relationship between them will be referred to as neurovascular coupling. Importantly, hemoglobin concentration, measured by fNIRS, has been reliably correlated with the BOLD signal within fMRI research to measure neural activation during cognitive tasks (Fishburn et al., 2014; Wijeakumar et al., 2017; see Scholkmann et al., 2014). The event-related, averaged neural response to key events of interest, such as naming of new objects, is referred to as the hemodynamic response.

Functional near infrared spectroscopy (fNIRS) can be used to provide accurate information about where in the cortex activation is occurring (the spatial resolution), and about the time-course over which activation occurs (the temporal resolution). In order to determine where neural signals are originating, a geometry is created, consisting of a set array of sources and detectors with a distance from each other that is scaled to head size, so as to record from the same cortical areas in each participant (Wijeakumar et al., 2015). The spatial resolution of fNIRS, as in, how reliably the source of the light absorption in the brain can be determined, is thought to be superior to that of EEG (Lu et al., 2010), but is still greatly improved through image reconstruction methods. The path of diffusion through which light passes into cortical tissues can be modelled in order to provide a precise localization for the signal in the cortex, creating voxelated correlates to the task-related neural response (Doulgerakis et al., 2019; Wijeakumar et al., 2015; 2017). This enhances the spatial resolution capabilities of the measure and provides better standardization of the source of signals that is robust to minor variation in where optodes are placed on each participant's scalp. Further, the temporal resolution of fNIRS technology is considered good (Huppert et al., 2006; Tak & Ye, 2014), which enables accurate measures of the time course of the hemodynamic response, connectivity and coherence measures.

In addition to being able to localise signals precisely within the brain, fNIRS also has a critical advantage for research that is not available with other commonly used neuroimaging technologies for cognitive neuroscience research such as EEG and fMRI. fNIRS technology is robust to participant movement and can therefore be applied to work with young children (Buss et al., 2014; see Pinti et al., 2018; Soltanlou et al., 2018; Wilcox & Biondi, 2015) and with participants who are moving freely within naturalistic tasks (Lancia et al., 2018). In a review, Pinti et al. (2018) describe why fNIRS technology has advantages for studying social interaction. As social behaviours require movement for interacting partners to jointly explore their environment, follow gaze, and spontaneously imitate each other (as examples) the robustness of fNIRS to motion enables applications of the technology to mapping relationships between behaviour and brain activity that occur during interactions. Because participants can interact, fNIRS technology can capture how brain activity is driven by the moment-to-moment behaviours that occur during interactions.

Though barely touching on the depth of research and knowledge about fNIRS technology, the points made here reveal that this measure is optimally suited to studying interactive word learning between caregivers and children. Judgements on the quality of spatial resolution of the measure may vary (Tak & Ye, 2014) but when critical, this resolution can be improved via statistical modelling (Doulgerakis et al., 2019; Singh et al., 2005; Wijeakumar et al., 2015; 2017). The measure further provides good temporal resolution (Tak & Ye, 2014), enabling responses to be measured for spontaneous naming events that occur in real time during interactions, and that additionally can be used to determine neural synchrony between participants. Critically, fNIRS technology is robust to participant movement, enabling caregivers and children to explore new objects on their own terms, without restricting them, thus capturing behaviour that is closer to what occurs in the real world.

1.7 Aims and Design of the Current Project

The primary aim of the current project was to identify the neural mechanisms of early word learning, within naturalistic caregiver-child play. The project explored how children learn words during play with their caregiver, and what breaks down within behavioural coordination and brain activity when a caregiver is not successful in teaching their child new words. Data from the same study was analysed for all results presented in this thesis. Thirty-seven dyads of caregivers and their children (19 aged 32-months and 18 aged 54-months) sat across from each other at a table within a laboratory space. They were given two sets of new objects to play with. The first set consisted of three unusual toys (new objects), and the second set consisted of five unusual toys. Before beginning each task, an experimenter provided the caregiver with word-picture pairs to reference as they taught their children the object names. Caregivers and children then interacted with the toys. After the interactions with their caregiver, children were tested by an experimenter in a referent-selection task to determine which names they had learned. The child was presented with pairs of new objects on a tray and asked to select one by name. We recorded fNIRS from both the caregiver and child over frontal, temporal and parietal areas during their interactions and the following referent selection test. Within this study, we aimed to determine how the child's brain activity was different between naming of learned words and words not learned and also to test whether children's learning during the interactions would be influenced by behavioural and neural coordination with their caregiver.

In the second empirical chapter (Chapter 3), we investigated the neural correlates of interactive word learning between caregivers and preschool aged children within this task, conducting an event-based analysis of children's and caregiver's neural responses to naming of new objects. As planned, we analysed neural responses to naming across the interactions. Based on the previously established role of the left inferior frontal gyrus in semantic processing and word learning for adults and children, and the right inferior frontal gyrus in language processing for children (see section 1.3 for references) we hypothesized that, for children, the inferior frontal gyrus would respond differentially during naming of words the child learned versus during naming of words the child did not learn. However, neural responses to naming were simultaneously analysed over all of the areas we recorded from in order to explore how language, attention, working memory and social processing were associated with children's learning.

In order to further explore the role of behavioural coordination toward influencing children's encoding of object names, we investigated how the spatial organization within objects' locations supported children's ability to map new words to intended referents. Behaviourally, we tested whether the coordination of object locations between the caregiver and child influenced the child's learning. Based on previous computational approaches and experimental findings (Benitez & Smith, 2012; Samuelson et al., 2011), we predicted that the spatial consistency of object locations during play, measured continuously through time, could predict how successfully word-object mappings were made by the child. For example, if the dyads kept objects lined up as they named them and held them in either the right or left hand, children might use each object's location to index one name to that object, thereby making accurate word-object mappings. With more robust mappings made during play, children would then give more correct responses when asked to select each object by name during a comprehension test. In the second part of the study, we then established that one particular aspect of spatial coordination, the distinctiveness of locations that new objects occupied during the interaction, predicted which words children learned during the interactions.

In the final empirical chapter (Chapter 4), we investigated whether neural coherence between the caregiver and child during learning was correlated with the number of words the child learned. This was based on the hypothesis that observed alignment in brain activity is related to successful communication and behavioural coordination, and ultimately facilitates word learning. Over these same frontal, temporal and parietal regions, we calculated a mean coherence value based on the continuous time series of the fNIRS data throughout each caregiver-child interaction with the new toys. In other words, rather than looking specifically at hemodynamic responses to naming events within the individual child's and caregiver's brain, we calculated the robustness of temporal relationships in their brain activity as they interacted *through time*. Coherence was analysed within corresponding channel pairings between the caregiver and child. Coherence values were calculated using the WTC toolbox in MATLAB (Grinsted et al., 2004).

In order to validate the coherence values in our actual data, we compared the mean coherence, for each channel, with that of a null distribution of corresponding channels we created from mismatched caregiver-child pairs. Because ours was the first study to assess whether neural coherence between the caregiver and child through time would be correlated with the child's word learning, we took an exploratory approach to determine whether neural coherence was greater for caregivers and children who were learning words together, compared with caregivers and children who were completing the same task but not interacting with each other. Any channels that showed significantly higher coherence on average, compared with the mean of the baseline, would be correlated with the number of words the child learned. However, no such correlations were appropriate as neural coherence was not greater for caregivers and children who were directly interacting, compared with that of mismatched caregiver-child pairs.

SECTION 2

-

Experimental Chapters

Chapter 2: General Methods Development and Piloting

2.1 Overview of Methodological Aims

The present chapter provides an overview of the methods development and piloting of the thesis project. The primary aim of methods development and preliminary testing was to be able to extract clearly defined variables from behaviours that occur spontaneously during interactions. The primary challenge was to enable participants to interact in a controlled laboratory environment as they would at home while measuring and analysing specific variables of interest within the constraints of fNIRS and machine learning technology. In the present study, variables and events of interest were examined within an open task structure of word learning during caregiver-child play. Similar, though not identical, methods development can be seen in prior studies that examined child-adult interactions (Clerkin et al., 2016; Piazza et al., 2020; Smith, Jayaraman, Clerkin & Yu, 2018; Suarez-Rivera, Smith & Yu, 2019). As in this prior work, we did not manipulate the variables we examined, like naming or object presentation during teaching and learning, by indicating to caregivers how they should teach. Instead, we measured and analysed children's and caregivers' learning, neural responses to naming events, spatial consistency and neural coherence within the existing data.

In the present chapter, considerations relevant to the sample tested and the development of stimulus, apparatus, video coding and machine learning methods are each described. Then a small, preliminary study that built from and tested the capabilities of these considerations is reported in section 2.3. The primary purpose of the pilot study, which did not have enough power to find the differences of greatest interest, was merely to ensure that neural coherence could be observed between adults and preschool-aged children in an equivalent manner to how it could be observed between interacting adults. This was a relevant consideration given that the underlying patterns of neural activity may have some differing temporal structure at different stages of maturity, and that fewer studies had been conducted examining neural coherence between adults and children prior to collecting the data for this study. Should it be possible to observe similar levels of neural coherence between adult-child and adult-adult dyads who were engaged in play with the objects used in the study and should the below described methods plan be feasible for analysing data from caregiver-child dyads, then this would provide some indication that the methods were enough to proceed with the main study. Recent work has further confirmed the validity of neural coherence between interacting adults and children (Leong et al., 2017; Miller et al., 2019; Piazza et al., 2020; Reindl et al., 2018), further resolving that particular concern.

2.2. Preliminary Methods Development and Considerations

2.2.1 Age and Demographic Considerations

The study tested children between 29 and 56 months of age, including a fairly continuous recruitment between these ages in the pilot study and two discrete age groups of 32-month-old children (29 through 34 months) and 54-month-old children (51 through 56 months) in the main study. This age range is of theoretical interest but also presents methodological challenges. In terms of theory, the preschool years are pivotal to language learning and later literacy (Avons, et al., 1998; Cabell et al., 2015; Lonigan, 2007; Song et al., 2015; Stokes & Klee, 2009). Yet, less work has considered how this is related to the rapid and simultaneous development of multiple types of skills between two and five years of age, and our age groups fell on either end of this developmental timeframe. These occur in the domains of social cognition (Saxe, 2013; Sobel & Corriveau, 2010; Sobel & Sommerville, 2009), and working memory to facilitate verbal skill (Kyttälä et al., 2014; Newbury et al., 2016; Vlach & DeBrock, 2017; 2019). The preschool years are also important to study because interactions between adults and children impact on the quality of children’s word learning between two and five years of age, predicting their vocabulary (Liebeskind et al., 2014; Napoli & Purpura, 2018). Yet, as discussed in sections 1.3-1.5, prior work has not tapped the neural correlates involved in interactively teaching and learning new words during these ages.

Logistically speaking, the children needed to be old enough to be able to finish the task, engaging in the word learning and comprehension tests for a lengthy period of time and with an intensive prior setup. Only children around 30 months and older were likely to be able to participate in this study, given the length of the setup and task, and the number of words to be learned. For example, Hirotani et al. (2009), tested children aged 18- to 21-months in a word learning and joint attention task using EEG/ERP. However, this task had a far simpler setting up procedure, and also a shorter and simpler word learning task. Prior neuroimaging work in lengthier and more involved language processing tasks and using more spatially precise analysis methods had typically tested children no younger than four to five years of age (Ekerdt et al., 2020; Romeo, Leonard et al., 2018; Romeo, Segaran et al., 2018). The younger group of the present study essentially fell between these prior studies, at 29- through 34-months of age. Some aspects of the setup still needed to be considered in order to make them accessible for these younger children. Most notably, experimenters' roles were divided so that each child spent as little time in the testing room as possible. While one experimenter explained the task to the caregiver and went through the lengthy process of setting up the adult's fNIRS cap and checking the signals, another experimenter completed the vocabulary test with the child and let the child play with toys in a waiting area. Both experimenters then placed optodes in the child's fNIRS cap and adjusted the signals simultaneously to speed up the process. Children were offered snacks and drinks during the study. When moving, younger children are also more likely to pull out an optode, either intentionally or by accident. This was prevented through placing an additional cap over the NIRS cap. Finally, during periods in which the caregiver and child were not interacting, such as the comprehension test, further measures were taken to ensure caregiver and child were disconnected from each other even while in the same space. Caregivers faced away from their child, and could they not see each other's test. Portions of the test that involved the caregiver only, like their cap digitization and their production test, were completed after the child had already been taken to the waiting room by the second experimenter to play. Collectively, these adjustments enabled us to test 32-month-old children in a naturalistic paradigm with an exceptionally involved setup, thus observing a younger age group than was typically tested in the most similar neuroimaging studies. Future work could explore behavioural paradigms and tasks with scaled down word learning demands in younger children, exploring technologies that can test mutual prediction and synchrony, such as eye tracking and motion capture, as alternatives to a complicated fNIRS setup.

Throughout the piloting and final sample recruitment, we aimed to recruit a sample representative of the local population of a city and surrounding areas in the eastern part of the UK. Nurseries involved in the recruitment were selected from a database held at the university. Though this was not intended, the average household income reported by the present sample was greater than the average of the city and surrounding areas. In addition to this consideration, children had slightly higher-than-average vocabulary scores within the British Picture Vocabulary Scale (BPVS) -3. These overall vocabulary scores, as reported later in Chapter 3, were positively correlated with the number of words children learned in the present study.

Testing and space constraints made one aim of the piloting to create a mostly portable testing design. As a result, the presented work presents a starting path toward a *fully* portable design that would be more accessible to participants who find it difficult to commit to travel back and forth from a university, potentially expanding the reach of the present design. Portable setups could include emerging wireless fNIRS technology (Abtahi et al., 2016; Piper et al., 2014) and a way to place an overhead camera that is not permanently installed and that can be used to stably record object locations through time. However, all other aspects of the design, including the table and chairs, toys, cameras, and the board on which to affix the object photos and labels for the caregiver to read, were already part of a portable design that was fully packed away after each session, even in the present study conducted in a laboratory space.

2.2.2 Stimulus Development

The stimuli were chosen ahead of time and then tested in the piloting study through measuring the participants' learning and examining the overhead video feed to determine how easy it would be for a machine learning network to classify the objects. Objects chosen for the study were selected to have comparable size but maximal contrast in colour and shape, for easy classification by the machine learning algorithm and distinctiveness for the child. This was done prior to the piloting phase. Photos of the new and familiar objects that are later modelled in Figure 3.1, are shown here and described in Figure 2.1. Of these objects, the complexity and colour distinctiveness were matched between new and familiar sets. Additionally, the ratio of toys with animal-like characteristics was kept consistent across new and familiar toy sets (the yellow squishy worm, felt puppet and green caterpillar had animal-like characteristics in the new toy set, and the horse, rabbit and dog were included in the familiar object set). The new words chosen from the NOUN database (Horst & Hout, 2016) were divided into four new words with two syllables and four words with one syllable, though, as described elsewhere, their pairing was counterbalanced across the new objects.

Figure 2.1. Photographs of new toys (left in shown in black box) and familiar toys (right) used in the experiment. The set of three objects is pictured in the top row, and the set of five objects is pictured in the bottom row.

In order to counterbalance the pairing of new words with new objects, eight different sets of word-object pairings and corresponding comprehension test sequences were made that could be cycled through systematically. For each session sheet, the pairings of new words with new objects changed systematically. In the referent selection task, objects were each presented twice as a target and twice as a distractor. The means to determine the pairings and order were as follows. Randomly chosen combinations of the possible orders for the three and five object sets were used to construct a planned presentation sequence in each session. Each new and familiar object (and corresponding name) were tagged and then inserted into one of four session sheet templates formulated to ensure that each object was (a) presented as a target in the first half of the test and the last half of the test and also as a distractor in the first half of the test and in the last half of the test, and (b) that each object was presented once on the left and right of the other object as a target, and once on the left and right of the other object as a distractor. These measures ensured that the comprehension tests given to each child had equivalent difficulty, and that the relative side objects appeared on, which objects they were paired with, how spread out their presentation was, and how often they appeared as targets and as distractors and when, were each varied systematically across participants (see Appendix C). Essentially, this meant that even though children weren't presented with the exact same sequence of object pairings every time in the comprehension test (in order to counterbalance the presentation between different objects across participants), they each had an equally difficult comprehension test.

2.2.3 Event Coding

Here, the planning of the event coding paradigm is described, largely prior to conducting the study and with staff in the lab even prior to conducting a pilot study with actual participants. A language annotation software developed by the Max Planck Institute for Psycholinguistics, was used to code the naming events that occurred during conversations between caregivers and children. The software enables synchronizing audio and video recordings and selecting and marking events in the audio waveform along hierarchical tiers of organization (Wittenburg et al., 2006). Audio events are marked in minutes, seconds and milliseconds from the start of the recording. In the present study, the video recordings were converted into an audio (.wav) format. Then the video and audio file from the same interaction were inserted into a new file. Coders watched the video and marked the onsets of the flashes that indicated the beginning and end of each block. Then the audio was listened through while watching the video and each time a new word was named, the coder selected the event in the waveform, creating a record of the onset and offset of the spoken name. The object name spoken was typed into the annotation along with an indicator ("p" or "c") of whether the caregiver or child had said the new object name. The annotations were then processed with a script that computed the onset and offset of each naming event in milliseconds and output a spreadsheet with columns for the object name, who said the name, the onset and the offset of the event in milliseconds, from the onset flash. A new file was then made for the reliability coding (described in section 3.2.4).

The software has been successfully applied to annotating language events in prior studies (for examples, see Colletta, Pellenq & Guidetti, 2010; Gerstenberger, Partanen & Reißler, 2017). The primary limitation of the present approach is that the audio coding portion is immensely time consuming and requires multiple human coders. Recommendations for future work include exploring an automated machine learning analysis of the waveforms, involving training the machine algorithm to recognize the new words in a speech stream and then allowing the algorithm to predict when new naming events occurred in conversation. Otherwise, the scope of any project is limited to the time taken to find and mark the naming events for each dyad.

2.2.4 Machine Learning

In order to measure the object locations, we would need an open-source machine learning network specialized in object classification and that also predicted objects' locations. The methods for the machine learning were researched prior to conducting the piloting or the main study and then were experimented with throughout the data collection. This began by exploring the video feeds visually and reading about the constraints of object classification using open source CNNs to see if the testing environment worked within those constraints. The study further required a balance between speed and computational efficiency and accurate predictions. Though two CNNs were originally considered as options (Redmon et al., 2016; Zhang et al., 2016), Faster R-CNN was chosen for ease of varying training parameters and adequate information posted informally about how to train, fix, and apply the network for a similar purpose by users online. Convolutional neural networks (CNNs) typically require consistency between the training images and the new data in aspects such as camera angle, object appearance, and the environment background. Otherwise, the network cannot generalize from the objects it was exposed to in one fairly constrained set of examples, to data that may differ significantly in the distributions of these aspects in the real world. Some structural adaptations have been proposed to address these limitations for Faster R-CNN (Chen et al., 2018). In the present study, controlling the distribution of object presentations between the set of examples that was used to train Faster R-CNN to recognize the new objects, and the full dataset, could be mostly accomplished through controlling the environment and the camera. As described earlier in the present chapter, this was achieved by choosing simple, solid new objects with high colour and shape contrast, by stabilizing the camera angle and setting up the table in the same place each time, by making sure that new objects from different sets were covered up when not in use, and by removing brightly coloured objects in the surrounding environment, for some examples. However, ambiguity still existed in object detection even with a straightforward camera view, objects that were distinct and a neutral background environment. For example, participants may wear brightly coloured clothing or may hold the objects at angles that were not included in the training data. This led in some cases to missing data or inaccurate predictions. In this section, some further methods and limitations of the current approach are discussed from an experimental standpoint.

After controlling for distraction in the background environment as much as possible, the length and the number of training epochs and the minimum confidence were further explored in order to find the best fit to the new data. As described in section 3.2.9, we trained the network with a representative sample of our data: around 900 randomly selected video frames that varied across four randomly selected participants in the study for which we had annotated bounding boxes around each new object and the table. The number of training epochs, or the number of passes over the training data, were varied in an exploratory manner between 50 and 500 total epochs. We further varied the length of each individual epoch (as in, how many examples would be randomly selected before updating the model weights) between these values. The test results, the predictions made on approximately 160 images that were randomly withheld from the training set, were then visually inspected by two experimenters to look for signs of model overfitting, like drawing a box around only a small part of a whole object, and for missing predictions for clearly visible objects that may indicate the model had not been trained through enough epochs. The final number chosen was 100 epochs with an epoch length of 100. After determining this, the minimum prediction confidence required to draw a bounding box was varied between 50% and the default of 80%. A confidence level of >= 65% was selected by further visual inspection by two experimenters. This practical judgement was made to ensure that bounding boxes were drawn around the objects as often as possible but without a major increase in inaccurate predictions (i.e., boxes drawn around something that was not the object). Two different versions of the network were trained for each object set (three objects: the blue jack, keychain and yellow squishy worm; five objects: green caterpillar, felt puppet, brightly coloured blocks, paint roller and red mesh structure). These parameter settings were explored separately for the CNN corresponding to each set size but were kept consistent between them during the later data processing. After the data processing around 100 frames per object were randomly selected and marked for accuracy. In the data, objects were accurately predicted in 86% of the frames where they were at least 25% visible. Further research could increase the accuracy of the model, but even more importantly, its ability to generalize to new environments. This requires updating the technology used within a rapidly developing field of computer vision, and also to utilize networks trained on exceptionally large pre-existing image datasets to achieve greater expertise in object recognition across varied contexts.

Figure 2.2. An example of the machine learning predictions of the new object locations for the three object set. Faster R-CNN output bounding box coordinates around each toy and the table in each video frame. The object locations analysed were the x and y midpoint of the highest confidence bounding box, drawn around each object.

2.3 Piloting Study

The study design was piloted with seven caregiver-child dyads with children ranging continuously in age from 30- to 56-months, and seven adult-adult friend dyads recruited from the undergraduate student population at a local university. The objective of the study was to determine whether neural coherence could be measured within adult-child dyads during a word learning task. Because prior work that had specifically examined speech comprehension and neural synchrony had done so in adult-adult dyads (Liu et al., 2019; Stephens et al., 2010), the pilot data was used to compare neural coherence between adult-adult dyads and adult-child dyads. This is because the children's thought processes and the neural signatures that reflect them are less mature, and some processes may occur at differing timescales (Wass et al., 2020). The aim of the analysis was to explore the capabilities of the behavioural task and the neural coherence measure even prior to the implementation of more rigorous baseline controls. This was done via comparing coherence within task relevant frequencies and non-task relevant (heartbeat) frequencies and comparing between the following factors. The pilot study examined differences of task (between new object training versus a familiarization with everyday objects), age pairing (between adult-child and adult-adult dyads), and number of objects (3 new objects and 5 new objects). Finally, the piloting study was used to iron out some aspects of the methods. These included determining how to best assess how many words caregivers and children learned, and how to make the space suitable for a later machine learning analysis, each discussed in this report.

2.3.1 Participants

Fourteen adults (7 dyads) were tested in the final paradigm with three and five objects. Adults were recruited as friends and were assigned on arrival to be the "teacher" (in the role of the caregiver) or the "learner" (in the role of the child). Seven caregiver-child dyads were recruited. Children who participated ranged in age from 30 to 54 months.

2.3.2 Stimuli and Apparatus

In the pilot study, adult teachers were in the role of the caregiver and adult learners were in the role of the child. Caregivers and children (or adult teachers and learners) were seated across from each other, at a medium sized white table, in a well-lit testing room. A felt board was positioned behind the person in the learning role, in view of the teacher or caregiver. Images and labels for new objects were placed on the board for the caregiver to reference while teaching object names. The stimuli, described in section 2.2.2, can be seen in Figure 2.1. A full description of the apparatus for the fNIRS recording can be seen in section 3.2.3, and a model of the apparatus can be seen in Figure 3.2. Optical density measurements via fNIRS were recording for caregivers and children, and for adult teachers and learners, during the session.

2.3.3 Procedure

This is an abbreviated version of the full procedure used in the pilot and main study, that is described in full in section 3.2.3. In the pilot study, adult teachers and caregivers could see and teach the object names, while adult learners and children learned them. First, the teacher was brought into the experiment room and was set up for testing. They were told they would see pictures and names of the objects on the board and that they needed to teach the child/adult learner the names of the toys. While the task was explained, the teacher was also fitted with a NIRS cap and the quality of the NIRS signals were checked. Afterward, we brought the child or learner into the testing room and fitted them child with a NIRS cap and checked the quality of the NIRS signals. Once we had completed the setup, we began the NIRS and video recordings and marked the onset of the interaction with a synchronized flash and a simultaneous stimulus marker in the NIRS recording. Like in the main study, in the first two learning blocks, the participants were given 2.25 min per block (45 sec per object * 3 objects, total 4.5 min) to discuss the names of the objects. In between these blocks, the participants were given three familiar objects and left to interact for 1.5 min.

After the second training block was completed, the teacher was moved to face away from the child/learner and an experimenter interacted with the learner for an additional 1.5 minutes using the familiar toys. The experimenters then began the comprehension test. During each trial, different experimenters would ask the teacher and learner to retrieve the target object. The learner retrieved the object while the teacher pointed to the correct photo on the display card. This was repeated across twelve comprehension trials: six trials where they were asked to select a familiar object and six trials where they were asked to select a new object. The entire word learning task and comprehension test were then repeated with the set of five new objects. Participants were given 3.75 minutes in each interaction (45 sec per object * 5 objects, 7.5 total minutes). The second comprehension test consisted of twenty trials: ten trials in which the participants were presented with a pair of familiar objects and asked to select one of the objects alternating with 10 trials in which they were presented with a pair of new objects and asked to select one of the objects. At the end, an experimenter asked the teacher and learner, separately, the name of each new object, saying “Can you tell me what this is called?”, as an additional word production test.

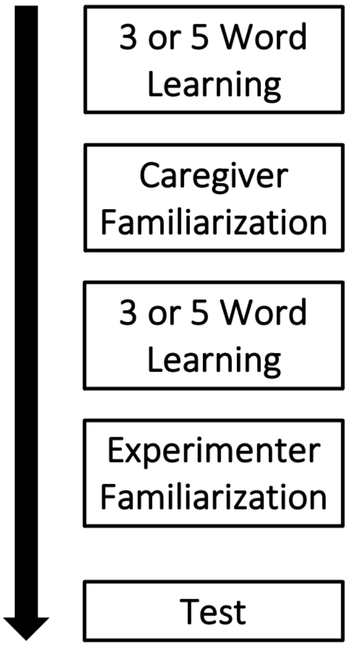

Figure 2.3. A schematic showing the procedure. This succession was repeated twice within each session, for a three object set in which the word learning blocks lasted 2.25 minutes, and the five object set in which the word learning blocks lasted 3.75 minutes. The familiarization periods lasted 1.5 minutes each. The comprehension test then consisted of trials that were repeated twice for each new and familiar object in the set.

2.3.4 Results

2.3.4.1 Behavioural Results: Adult learners scored, on average, 98.2% accuracy during the new object trials of the referent selection test, ranging from seven to eight learned words out of eight possible. Adults' immediate recall of the object names was nearly perfect. In the production test, adult learners scored a 78.7% on average, ranging from four to eight words remembered out of eight possible. Children scored an average of 64.2% in the comprehension test, ranging from three to eight learned words, but an average of only 21.4% in the production test, ranging from zero to three words remembered later. For this reason, performance in the production test was taken as an indication of adults' learning, but performance in the referent selection (comprehension) test was taken as an indication of children's learning.

2.3.4.2 Neural Coherence Analysis and Results: A mixed effects linear model was run with the Task (new object training versus the familiarization) as a within subjects factor, and Group (child-caregiver dyads or adult friends dyads) as a between subjects factor. Random effects in the model included a dyad identifier and the fNIRS channel. The data was first filtered so that coherence in the task-relevant range was greater than the approximate mean of 0.3/1 for each channel (observation) retained. The remaining mean neural coherence across each task between periods of 3.2-12.8 seconds was the dependent variable (Cui et al., 2012). Overall, an effect of task was found, with coherence during teaching and learning new words significantly lower than during play with the familiar objects, $t_{127} = -3.27$, $p = 0.001$, Figure 2.4. The degrees of freedom represent the number of viable channels times the number of dyads. This indicated that the process of learning something new broke neural synchrony down significantly. By contrast, there was no difference in coherence based on whether or not the participants were caregivers and children, or adult friends, $t_{127} = -1.53$, $p = 0.128$, and no interaction between the learner's age and task type, $t_{127} = 0.72$, $p = 0.474$.

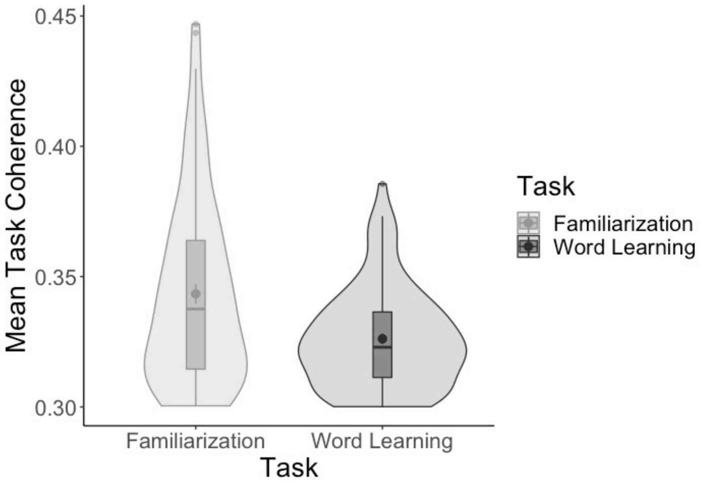

Figure 2.4. Boxplots that show neural coherence within the task-relevant frequency band (corresponding to a period length of 3.2-12.8 seconds), for the familiarization in which caregivers and children played with the regular toys, and the new object training (word learning) in which the caregiver taught their child object names. Channels with low coherence in the task-relevant frequency band (< .3) were omitted from this analysis.

As a control, we examined coherence during the range of 0.6-1.0 seconds, or the range of resting heartbeat. At these period lengths and using the filtered dataset, no effect of task (familiar objects vs new objects) was found, $t_{127} = -0.98$, $p = 0.329$. By contrast, child-adult dyads had lower heartbeat-driven coherence compared with adults, $t_{127} = -2.31$, $p = 0.023$, while no interaction was found between new vs. familiar objects and age, $t_{127} = 0.51$, $p = 0.614$.

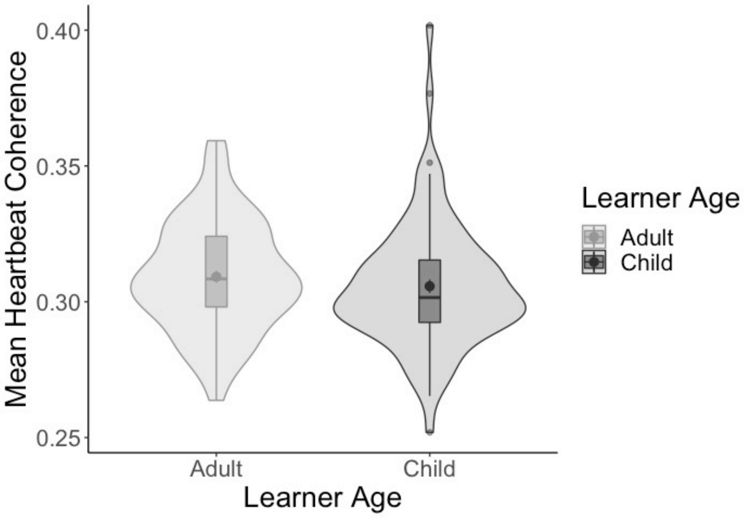

Figure 2.5. Boxplots that show neural coherence within the heartbeat-relevant frequency band (corresponding to a period length of 0.6-1.0 seconds), for adult-adult friend and caregiver-child dyads. Channels with low coherence in the task-relevant frequency band (< .3) were omitted from this analysis.

As an exploratory measure, we tested whether coherence was higher in the easier task with three objects or in the more difficult task with five objects. When we compared coherence between the types of new object trainings, we only found a marginal effect of task, with coherence being slightly lower in the longer and more difficult task with five objects $t_{65} = -2.00$, $p = 0.050$. No effect of learner age was found, $t_{65} = -0.45$, $p = 0.658$ nor interaction between task and age, $t_{65} = -0.25$, $p = 0.802$.

2.3.4.3 Brain-Behaviour Relationships: To further explore the data, a distribution was created of the channels based on the mean coherence across the task-relevant frequency band and during the new object trainings. Those channels that were, on average, above one standard deviation from the mean coherence (0.307 + 0.006), were selected for further comparisons. The coherence values within these channels, channel 1, located over the left inferior frontal gyrus (IFG), and channel 18, located over the right posterior temporal cortex (PTC), were tested for correlations with the number of words children and the number of words adult learned. However, neither channel was correlated with the words participants learned, for channel 1, $r_{12} = -0.298$, $p = 0.301$, for channel 18, $r_{12} = 0.293$, $p = 0.308$.

2.3.5 Conclusions and Adjustments Made

Conducting and analysing the pilot data revealed some key aspects of the study design. As the pilot study was conducted, we practiced minimising the time needed to set up and to have the child in the testing room. This included completing the caregiver setup first. Most adult-adult dyads were run prior to having caregivers and children into the lab to pilot the full session, so this was practiced on adults until the experimenters had minimized the setup time and time in between tasks. The final procedure was predominately planned prior to the piloting study but was practiced throughout the adult testing so as to be ready to test children with the complicated apparatus.

The hyperscanning analysis revealed that coherence was comparable within task-relevant frequency bands examined in prior work between child-adult dyads and adult-adult dyads, meaning that we could test child-adult dyads and observe equivalent task relevant effects as had previously been explored only in adults (Liu et al., 2019; Stephens et al., 2010). The pilot data indicated that neural coherence during a task-relevant frequency band (3.2-12.8) that had been used in previous studies, even those with a very different task (Cui et al., 2012), showed differences based on task while the ‘heartbeat’ range (0.6-1.0) did not. Not finding differences by task in the cardiac range suggested that differences in coherence observed based on task were less likely to be owing to physiological noise in the data, and instead were more likely to be driven by engaging in the task. However, a number of questions remained, including whether coherence would be intact during the new object training within a larger sample, and whether greater power would be afforded to find statistical evidence of brain-behaviour relationships.

During the pilot study, we determined that adult learning could be assessed using a later production test, which showed distributions of retention that were similar to children in the comprehension test. By contrast, the selection test was too easy for the adults, as they were accurate nearly 100% of the time. The production test was then applied to all caregivers who were teaching to assess which words they learned in the main study and to rule out that they learned the same words during interactions as their children.

The earlier methods development, described in section 2.2, was largely tested in the pilot study. Early inspection of the video data during the pilot testing led to changes to the environment in order for the machine learning analysis to be feasible.

We removed any objects from the testing space that had colours similar to our new objects and covered new objects up when participants were not interacting with them. Though the overhead camera was permanently positioned, piloting was used to adjust the audio and video recordings so that we could see and hear the participants while they were moving. The side camera was adjusted to enable video coding of the referent selection and audio coding of the naming, while the overhead camera, which did not have an audio file, was adjusted only for a complete view of the table. Having multiple cameras enabled us to record audio with enough clarity to mark the naming events in the recordings later, having a viable video + audio recording of the comprehension test so that coders could hear which object was asked for and see which objects the child selected. Meanwhile, the overhead camera position was essential to establishing stable measurements of the object locations that could be referenced to the table.

In the following chapters, the methods and results of the main study are reported. The methods development, apparatus and procedure described here were applied to the main study. As described in this section, what much of the preparatory work and the piloting revealed was the need to structure the testing apparatus and surrounding environment in order to be able to extract variables of interest using fNIRS and machine learning technology. Most notably, this included minimizing the time that children needed to be in the testing room, simplifying the environment for a later machine learning analysis of object locations, deciding how the naming events would be coded, and confirming that task-relevant neural coherence could be measured with adult-child dyads.

Chapter 3: The Impact of Neural and Spatial Coordination on Learning During Encoding of Object Names

3.1 Introduction

Interactions between adults and children facilitates early vocabulary development, above and beyond what is learned from ambient language input (Romeo, Segaran et al., 2018; see also Hoff, 2006; Tamis-LeMonda, Kuchirko & Song, 2014). For example, children's ability to make correct word-object mappings during interactions is facilitated by the coordination of their attention and working memory with an adult speaker (Cartmill et al., 2013; Kalagher & Yu, 2006; Suanda, Smith & Yu, 2016; Sullivan, Mundy & Mastergeorge, 2015; Tomasello & Farrar, 1986; Yu & Smith, 2013; Yu et al., 2009). At a neural level, previous work suggests that children's event-related responses in a referent selection task are more sensitive when the initial encoding of the new words and objects occurred as they shared attention with an adult (Hirotani et al., 2009). The implication is that, early in development, social interaction supports word learning via coordination between adults and children, ultimately enhancing the child's underlying neural encoding of pairings made between objects and spoken new words (Mundy & Jarrold, 2010; Samuelson et al., 2011).

Traditionally, the neural correlates of language development have been studied within computer-based tasks that do not tap the influence of interaction on word learning in the real world. This previous work points to the importance of localized speech and language processing, but also attention and working memory. Foremost, children's language-learning can be functionally mapped to a language-processing network between Broca's area in the left inferior frontal gyrus and Wernicke's area in the left posterior temporal cortex (Balsamo, Xu & Gaillard, 2006; Berl et al., 2014; Romeo, Leonard et al., 2018; Romeo, Segaran et al., 2018; Takashima et al., 2019) and also to attention and working memory networks (Berl et al., 2014; Ekerdt et al., 2020; Nora et al., 2017). Additionally, it is known that the neural correlates activated during processing concrete nouns are more broadly distributed when children believe themselves to be engaged in interaction with another person (Hirotani et al., 2009; Rice, Moraczewski & Redcay, 2016). It is clear that contextual social computations are made during the processing of concrete nouns when children perceive that they are sharing representations, compared with when they are knowingly presented with pre-recorded stimuli while alone. However, the neural signatures of word-learning *through direct, live social interaction* and how they are influenced by behaviour are virtually unmapped by previous work.

The present study aimed to determine whether neural signatures of children's word learning could be detected within the child's and caregiver's neural responses to naming of new objects during their interactions. Functional near infrared spectroscopy (fNIRS) technology enabled a naturalistic testing paradigm in which participants could interact with each other and with objects on the table (see Pinti et al., 2018). The preschool years were chosen to assess interactive word learning, given their pivotal role in early language development and later literacy (Avons, et al., 1998; Cabell et al., 2015; Lonigan, 2007; Song et al., 2015; Stokes & Klee, 2009), and that this age group has been studied less because they are often considered too young for functional neuroimaging research with high spatial resolution in language-learning tasks.

During this task, caregivers taught their 32- and 54-month-old children the names of eight new objects during free play, while fNIRS data was recorded from both the caregiver and child and the interactions were videotaped. We aimed to analyse event-related hemodynamic responses following instances of naming. The first analysis determined whether the time-locked, neural responses to naming of objects the child learned were significantly different within the child's and caregiver's brain activity, compared with those words the child did not learn (for previous work that discussed a two-person social neuroscience framework, see Hasson et al., 2012; Wass et al., 2020).

A second set of analyses explored whether children's learning could be predicted by behaviour that occurred during the interactions. Previous work had shown that the consistency in new objects' spatial positions during play with caregivers facilitated children's memory for their names (Samuelson et al., 2011), and, more generally, that visuospatial working memory is correlated with verbal working memory (Alloway, Gathercole & Pickering, 2006), and particularly with children's receptive and expressive vocabulary scores (Stokes et al., 2017) during the preschool years. Taken together, prior work suggests that development of spatial and language processing domains influence each other, potentially pointing to the role of spatial coordination as a key aspect of interactive word learning between caregivers and preschool aged children.

3.2 Methods

3.2.1 Participants

We recruited a sample of 68 preschool aged children from a city and surrounding areas in the UK. These included two age cohorts, 38 32-month-olds (*M* = 32.2 months, *SD* = 56 days) and 30 54-month-olds (*M* = 53.6 months, *SD* = 57 days). Of those, we excluded 24 children who did not want to wear the NIRS cap, and seven children because of technical errors. The final sample consisted of 19 32-month olds (*M* = 32.5 months, *SD* = 52 days) and 18 54-month olds (*M* = 53.7 months, *SD* = 58 days) and their caregiver.

3.2.2 Stimuli and Apparatus

Caregivers and children were seated across from each other, at a medium sized table, in a well-lit testing room. We videotaped the session from an overhead camera that was mounted to the ceiling, and with two additional cameras placed at opposite sides of the table. A felt board was positioned behind the child, in view of the caregiver. We attached images and names of the objects to the board for the caregiver to reference while teaching their child the object names. During the word learning task, our new stimuli consisted of eight unusual toys, including a three-object set of a yellow squishy toy, a large decorative key ring, and a large blue jack, and a five-object set of a wooden caterpillar that had been painted solid green, a miniature paint roller, a red mesh structure, brightly coloured blocks, and a small felt puppet, (see Fig 3.1). During breaks, children played with two sets of familiar toys that consisted of a small plastic horse, a cup and a toy train for the three-object set, and a toy bed, rabbit, dog, small shoe, and small plastic sunglasses for the five-object set.

Both the caregiver and child were fitted with a stretchy NIRS cap that delivered near-infrared light at wavelengths of 690 nm and 830 nm via eight fibre optic cables (sources) and measured optical density via 16 fibre optic cables (detectors). We used a 48-channel TechEn ® CW7 system, with a geometry that was divided between both participants. Optodes were fitted securely into the cap with grommets. Given that they were moving freely, participants wore an additional cap over the NIRS cap in order to hold the optodes in place. Optode locations were recorded in 3D space using a digitization device that registered the locations of head landmarks and the optodes. Though we did not analyse these data for the current study, participants were each fitted with a Pupil Labs head-mounted eye tracker that consisted of a camera that pointed outward from the participant's forehead to provide an egocentric view at 100 diagonal degrees, which captured most of the participant's visual field, as well as an eye camera that recorded x and y coordinates of the participant's gaze direction.

Figure 3.1. New toys (left) and familiar toys (right) used in the experiment.

3.2.3 Procedure

The caregiver was first taken into the testing room by an experimenter while a second experimenter completed the British Picture Vocabulary Scale®, version 3 (BPVS-3) with the child (Dunn & Dunn, 2009). The BPVS-3 measured the child's vocabulary, for example, by presenting the child with a set of four pictures and asking them to point to one of the pictures, e.g., "Can you point to 'house'?" Meanwhile, the caregiver was told they would see pictures and names of the objects on the board and that they needed to teach the child the names of the toys while they played with them as naturally as possible. While the task was explained, the caregiver was also fitted with a NIRS cap and head-mounted eye tracker. After the caregiver had been set up, including checking the quality of the NIRS signals, we brought the child into the testing room and fitted the child with a NIRS cap and eye tracker. We then inspected and made final adjustments to the NIRS signals for both participants and calibrated the eye trackers. Once we had completed the setup, we began the NIRS and video recordings and initiated a flash and a simultaneous stimulus marker in the NIRS recording. These events marked the beginning of each task for all cameras and the neuroimaging data.

At the start of the word learning block between caregiver and child, an experimenter pronounced the names of the new objects while the caregiver viewed photos of the objects and their labels on the board. Participants were then left to interact as naturally as possible. In the first two word learning blocks, the caregiver was given 2.25 min per block (45 sec per object * 3 objects) to interact with the child and teach the child the names of the objects. Across both blocks, the participants were given 4.5 min to interact with the new objects. In between the two word learning blocks the caregiver was given three familiar objects (the toy horse, cup and train) and left to play with the child for 1.5 min. This familiarized the child with these toys prior to testing. After the second training block was completed, the caregiver was moved to face away from the child and an experimenter played with the child for an additional 1.5 minutes using the familiar toys. This helped ensure the child was comfortable interacting with the experimenter prior to the testing phase.

The experimenters then began the comprehension test. The child and caregiver (still facing away from each other) were presented simultaneously with pairs of the familiar objects alternating with pairs of new objects, by the experimenters (the actual toys were presented to the child while a display card with their photos was shown to the caregiver). During each trial, the experimenter would ask the child to retrieve the target object. The child retrieved the object while the caregiver, unable to see the child and out of their child's view, pointed to the correct photo on the display card. In total, participants completed twelve comprehension trials: six trials where they were asked to select a familiar object and six trials where they were asked to select a new object.

The entire word learning task and comprehension test were then repeated with the set of five new objects. Because we had increased the number of objects, participants were given 3.75 minutes in each interaction (45 sec per object * 5 objects) for a total of 7.5 min of interacting with these objects across both blocks. The second comprehension test consisted of twenty trials: ten trials in which the participants were presented with a pair of familiar objects and asked to select one of the objects alternating with 10 trials in which they were presented with a pair of new objects and asked to select one of the objects.

Finally, in order to test children's and caregiver's ability to produce the object names, an experimenter asked the child the name of each new object, saying "Can you tell me what this is called?" The child was given a few seconds to answer. The caregiver faced away from the child during this task and wore headphones. Subsequently, caregivers completed an identical production test, to assess their learning. We obtained a digitization of all optode locations on the child's and caregiver's head using key landmarks from the 10-20 reference system, including digitizing the nasion, right and left ear, Cz and inion as landmarks.

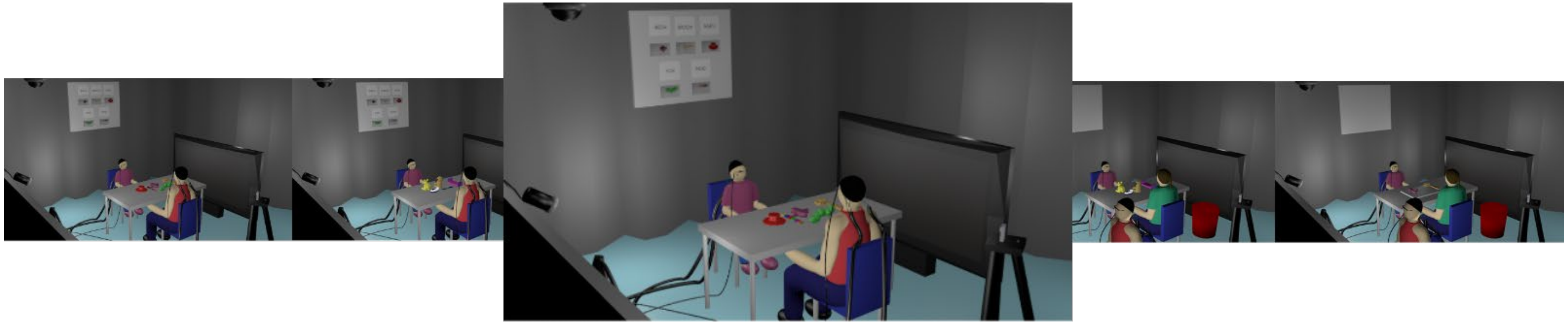

Figure 3.2. The apparatus and procedure. From left to right, (1) caregiver and child learning the names of five new objects. The caregiver referenced the object pictures and corresponding names on the felt board positioned behind the child while teaching their child (2) caregiver and child playing with familiar objects, (3, enlarged), second interaction with the new objects (4) a familiarization with the child and experimenter while the caregiver faced away and (5) the comprehension test in which an experimenter separately presented the child and caregiver with pairs of objects and asked them to select each new and each familiar object twice. The child is shown selecting an object from a tray, while the caregiver, facing away, points to photos of the objects on a display card.

3.2.4 Observer Coding of Behaviour

Using a waveform of the participants' speech, we marked the onset and offset of each time the caregiver or child named an object during each caregiver-child interaction with the new objects. A language annotation software (Wittenburg et al., 2006) was used for the marking of naming events, using the procedure described in section 2.2.3. Coders also watched videos of the comprehension test to select the object the child picked during each trial. If the child picked up both objects during the trial, the coder judged which toy the child ultimately selected based on the last object they remained holding or handed to the experimenter upon request. If the child selected the correct object on both trials for a given name, that object name was marked as a learned word. If the child picked the correct object on only one trial or on neither trial, that name was recorded as a word not learned. The coder then listened to a video of the child's production test and if the child had later produced any of the words marked as not learned from the comprehension test, that word was updated to be a learned word (this occurred only a few times in all of the data). The child's production test was scored along a five-point Likert scale that the coder used to rate the accuracy of the child's pronunciation. The scale ranged from 1 = "not at all" to 5 = "perfect pronunciation". Any of the child's pronunciation attempts that were scored as a three (close), and up, were counted as correct/learned. Ultimately, the child's learning outcome was calculated as the percentage of learned words out of the total of eight possible words over the course of the session. The caregiver's learning was based on a scoring of their production test, identical to the child's scoring.

Data from approximately 22% of participants in the study were double-coded. Reliability between video coders was calculated for each participant and then averaged across participants. For the naming events, the percent agreement between coders was assessed based on the number of events for which both coders had independently marked the onset of a naming event within 250 ms of each other, over the total number of coded events. Inter-coder reliability was 97.73%. For the comprehension test, reliability was assessed based on the percentage of trials in which coders agreed on which object the child selected, with an overall inter-coder reliability of 99.59%. Finally, reliability on scoring the production test was assessed based on whether coders agreed on the participant's accuracy in naming each object within one level of the Likert scale, with an overall reliability of 99.22%.

3.2.5 Scoring the British Picture Vocabulary Scale (BPVS-3)

Each child's original score was based on the ratio of words they identified out of the number of words they were presented with. Because the BPVS-3 was not standardized for our youngest age group (who were 2.7 years old on average, while the test is standardized starting at three years), we standardized our sample based on the original scores in each age group we tested (standardized score = child's original score – mean original score/sample standard deviation). For these older children, we found a strong correlation between the scores we standardized based on our sample and the population-based scores, $r_{15} = 0.99$, $p < 0.001$.

3.2.6 Determining the fNIRS Geometry

We designed the geometry in which fNIRS optodes would be organized in order to cover the maximum number of regions of interest discovered in the fMRI literature (Wijeakumar et al., 2015). We conducted comprehensive searches for neuroimaging studies of word learning and social interaction in Google Scholar, PsychINFO, and PubMED databases, using keywords including "rapid word object mapping", "word learning", and "child word learning", as well as "social interaction" and "social language development", and to capture spatial processing, from a previous meta-analysis of the neural correlates of visual working memory (Wijeakumar et al., 2015, Table 2). We projected the ROIs onto an MNI atlas for an adult and a 30-month old child using AtlasViewerGUI in HomER2, overlaying these ROIs with the cortical coverage from potential cap geometries based on a digitization taken of a woman and of a 30-month-old child. The final geometry we selected covered the superior frontal cortex, inferior frontal gyrus, posterior temporal cortex and parietal cortex (see Appendices A and B).

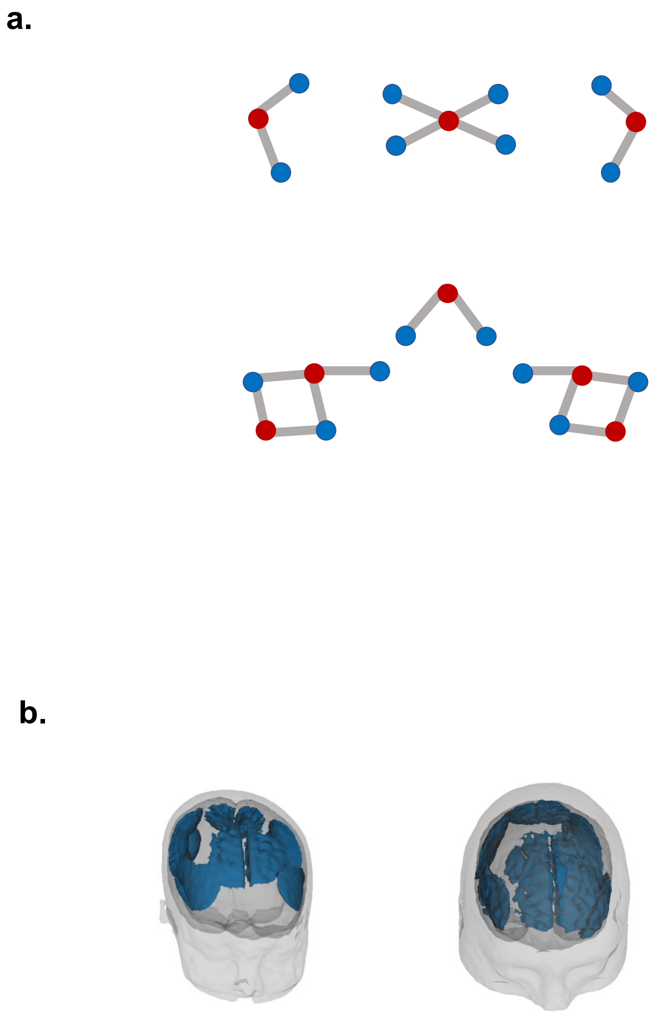

Figure 3.3. The cortical coverage of the fNIRS, including a.) The NIRS geometry array with the front of the head positioned toward the top. Six clusters are arranged bilaterally covering superior/prefrontal and inferior frontal, posterior temporal and parietal areas. Sources are shown in red and detectors are shown in blue. b.) Cortical coverage of the NIRS recording on a representative child (left) and adult (right).

3.2.7 fNIRS Pre-Processing

The onsets of naming events audio coded during the interactions were inserted into the NIRS data. After separating the data collected from the caregiver and the child, we conducted signal processing on the unprocessed data using HomER2 software (Huppert et al., 2009), We eliminated channels with signal measurements that did not fall between 80 dB and 130 dB, dB=20*LOG10(*y*), where y is measured as light intensity, converting the light intensity measures into optical density (OD) units. This enabled us to remove low quality channels prior to processing the data. The pre-processing pipeline identified motion artefacts based on changes in optical density greater than 0.4 au within one second or with a SD greater than 50. A targeted principal component analysis (tPCA) corrected for components identified as motion (Yücel et al., 2014). Large, abrupt changes were identified in the data and removed because these were due to motion, and naming events during which large motion artefacts had occurred were not analysed. Data slower than 0.016 Hz and faster than 0.5 Hz were then bandpass filtered, leaving only a range of data that is associated with task-relevant brain activity, and the remaining motion artefacts were removed by channel. After processing the data, concentration values for oxygenated hemoglobin (HbO), deoxygenated hemoglobin (HbR) and total hemoglobin (HbT) were computed for each channel using the modified Beer-Lambert law. The processed data consisted of these concentration values for HbO and HbR at each time step.

3.2.8 fNIRS Image Reconstruction and GLM

A forward simulation was used to determine the sensitivity of light measure within each channel of the fNIRS recording. Head 10-20 landmarks for each participant, including the nasion, inion, Cz, and right and left ear, as well as the location of each optode on the participant's cap, were first run through a transformation algorithm to correct minor measurement errors in the digitization readings. AtlasViewerGUI in HomER2 was then used to relax the digitized points onto the Collins head atlas for the adults (Aasted et al., 2015), and a 3-year-old child's structural MRI head atlas for the children (Fillmore et al., 2015; Richards et al., 2016; Richards & Xie, 2015; see Wijeakumar et al., 2015). To determine the sensitivity of each channel to measuring absorption of infrared light in skin, skull and cortical tissues, Monte Carlo simulations were run with 100,000,000 photons using the projected digitization-to-atlas geometry for each participant (for more information about the algorithm used, see Fang and Boas, 2009). The final output of the Monte Carlo simulations was a sensitivity profile of the light measurements over the whole head, for each wavelength recorded.

The participant's head volume and sensitivity profile were then converted into NIFTI images. We used NeuroDOT, a MATLAB toolbox designed for post-processing and spatial registration of fNIRS data, to localize signals to the participant's head volume. NeuroDOT accomplishes spatial registration (e.g., localizing signals that were measured from the scalp to their source in the individual participant's brain) while retaining the temporal structure of the data. Thereafter the inherent phase structure that is maintained in this data enables a more accurate localization of signals across varying cortical depth, compared with methods that average out this structure prior to localizing the signals (Doulgerakis, Eggebrecht, & Dehghani, 2019). To control for physiological noise in the data, we conducted a global signal regression (Forbes et al., 2021). We then conducted a general linear model (GLM) on aggregated hemodynamic responses within the image reconstructed data to determine beta (β) values of the averaged hemodynamic responses to naming of objects. Beta values consisted of an estimated averaged hemodynamic response for each participant, relative to baseline, for naming events in each type of task and condition. Finally, these beta images were transformed into a common, MNI atlas space for caregivers and children. The beta images were converted into a binary mask for each subject. These masks were summed across individuals to create a group intersection mask for children and for caregivers (separately) that consisted of the voxels common to 70% of the participants' individual masks. The resulting voxelated mask was applied prior to calculating relative differences in hemoglobin concentration for each participant based on chromophore, learning, task and the child's age. In other words, the final fNIRS data that was analysed consisted of a single, separate beta value within each voxel that represented the level of neural activation to the naming events that occurred for HbO and HbR, for words that were learned and that were not learned, and that occurred during the interaction versus the comprehension test. Voxels were only included if measurements were taken from that area of the brain in 70% of the participants. Voxels that were only recorded from for under 30% of participants were not considered in the analysis.

3.2.9 Machine Learning Predictions of Object Locations

We used a machine learning network to classify the toys and determine their locations within each recorded video frame from the overhead camera. The methods development can be viewed in section 2.2.4. TensorFlow, an open source deep learning network developed by Google, provided the foundational architecture for the training (Abadi et al., 2016). Specifically, we trained Faster R-CNN, a convolutional neural network architecture that relies on deep learning and is specialized for image classification, using convolutional feature maps to recognize the new objects and the table in our videos. After being trained, the network fed suggested object locations into a fast-processing regional proposal network that predicted the type and location of an object in the video frame (Zhang et al., 2016). In order to enable Faster R-CNN to make predictions about the class and location of our objects, we trained the network with a representative sample of our data: around 900 randomly selected video frames were taken from the overhead camera during the word learning interactions and annotated with bounding boxes drawn around each object.

After training the network through 100 epochs, we assessed the accuracy of the trained network by inspecting the reliability and accuracy of bounding boxes drawn at >= 65% confidence in a representative sample of a few hundred frames. At least 100 of these frames were independently judged by two individuals. Once we had trained the network for the three and five object sets, we then extracted video frames from the recordings of the task and allowed the trained model to make predictions about the locations of objects. For each video frame, Faster R-CNN produced a class label for each toy (e.g., “green caterpillar”) and rectangular coordinates in image space.

The location predictions began on the first frame in which the caregiver or child touched each object. From that point on, using normalized coordinates of the bounding boxes, we re-computed the locations of the objects, referenced to the location of the table, within each video frame. The bounding box coordinates for each toy that had been computed relative to the centre of the table were used to analyse the data.

3.3 Analysis 1: Neural Correlates of Interactive Word Learning Between Caregivers and Children

In the first analysis, we investigated the neural correlates of children’s word learning during the caregiver-child interactions. Primarily, we aimed to determine whether the child’s neural responding to naming of new objects differed between naming of learned and of words not learned during interactions with their caregiver. We compared event-related neural responses to the caregiver’s naming of objects within the learning interactions. We also examined later recognition of the objects as the experimenter named objects during the comprehension test. Hemodynamic responses to naming events were analyzed over frontal, temporal and parietal cortical regions.

Because the inferior frontal gyrus (IFG) area has been shown to play a pivotal role in semantic and noun processing (Balsamo, Xu & Gaillard, 2006; Chou et al., 2006; Fiebach & Friederici, 2004; Grindrod et al., 2008; Hirshorn & Thompson-Schill, 2006; Liljeström et al., 2009; Romeo, Leonard et al., 2018; see also Fedorenko, 2014) and word learning (Davis et al., 2009; López-Barroso et al., 2013; Nora et al., 2017; Takashima et al., 2019), we made a prediction that the child's inferior frontal gyrus (IFG) would respond differentially to naming of words they learned or did not learn. A single analysis of event-related neural responses to naming over the full cortical area covered tested this specific hypothesis though only in the context of exploring whether activation within domain-general, including attention, working memory (Archibald & Joanisse, 2012; Berl et al., 2014; Ekerdt et al., 2020; Ferreira et al., 2015; López-Barroso et al., 2013) and social cognition regions (Hirotani et al., 2009; Rice, Moraczewski & Redcay, 2016; Rice & Redcay, 2016), also contributed to word learning in our task. Finally, we conducted an identical analysis examining the caregiver's hemodynamic responses to naming and investigated whether these responses differed between words the child learned or did not learn. Based on the critical role of caregiver's behaviour in teaching new words to their child, this analysis would determine whether the caregiver's brain activity could be predicted based on whether their child was learning. In other words, the analysis tested whether the caregiver's neural signature while they taught their child was unique when they were effectively communicating the new object names.

3.3.1 Statistical Analysis and Results

3.3.1.1 Behavioural Results

On average, 32-month-old children learned 4.05/8.00 words during the study (*SD* = 2.04), while 54-month-old children learned 5.39/8.00 words (*SD* = 2.03), t_{35} = 2.09, *p* = 0.044. Children's learning over the entire experiment ranged from 1-8 total object names. From the three-object set, children on average learned 54% of words while from the five-object set, children learned an average of 62% of words. The BPVS-3 scores for 32-month-olds were (*M* = 37.26, *SD* = 12.99), and for 54-month-olds were (*M* = 70.22, *SD* = 10.02). The standardized score for the older age group was 110. Vocabulary was correlated with the number of words learned in the study,

$r_{35} = 0.475$, $p = 0.003$. Additionally, caregivers and children remembered different object names $\chi^2_1 = 1.51$, $p = 0.220$ and did not learn words at a similar rate, $r_{35} = 0.063$, $p = 0.713$. Therefore, differential responding in brain activity between words that the child learned or did not learn, was related solely to the child's learning and not the caregiver's own learning.

3.3.1.2 fNIRS Analysis

We carried out an analysis plan to compare neural responses to naming of learned versus words not learned during the interactions and the children's comprehension test. A mixed effects linear regression model was conducted comparing the image-reconstructed beta maps based on age (32 vs 54 months) as a between-subjects factor, and learning (naming of learned vs words not learned) and chromophore (HbO vs HbR concentrations) as within-subjects factors. The dependent variable was the magnitude of the fNIRS hemodynamic response in the child's brain within the voxelated space, relative to baseline. We compared the beta values to determine whether there were statistically significant neural responses to naming within clusters of adjacent voxels. In order for a cluster to be considered a significant effect, differential responding between conditions must be present in a minimum number of adjacent voxels, or a large enough area (see below). We were not able to conduct any comparisons with naming of familiar objects because the familiar objects were not named frequently enough by the caregiver *(Median* = 18 events) to analyse along with the caregiver's naming of new objects (*Median* = 104 events).

AFNI software was used to conduct the regression on the group-level voxelated beta maps, using the 3dLME function. To control for family-wise error given the number of voxels in which we were analysing beta values, we used the 3dClustSim function that determined how many adjacent voxels the cluster would need to span. In particular, we employed a mixed-ACF approach that has been shown to be an effective control for event-related designs (Cox et al., 2017). Only those results with a canonical hemodynamic response (a neurovascular coupling pattern that shows a statistical interaction between HbO and HbR) are reported here, as an inverse relationship between chromophores would be indicative of task-relevant brain activity (Buss et al., 2014).

3.3.1.3 Statistical Comparisons and Results

Table 3.1. Child Results – Word Learning

Effect	**Cluster**	**ROI**	**Hemi**	**Volume (mm^3)**	**Centre of Mass**		
					x	**y**	**z**
Chromophore (Hb)	Supramarginal Gyrus	TPJ	R	228	54	-49	45
Learned x Chromophore (Hb)	Fusiform	STS	R	325	57	-56	12

A mixed effects linear regression model was conducted to identify neural clusters that were significantly activated to naming events in each condition. The neural clusters found through this analysis can be viewed, by condition, in Table 3.1. In the analysis, the child's age (32 months vs 54 months) was included as a between-subjects factor, whether the child learned the word (learned word vs word not learned) and chromophore (HbO vs HbR) were included as within-subjects factors. Mean estimations of changes in activation (beta values, µM), were the dependent variable. This model revealed a main effect of chromophore, (HbO vs HbR) in the right supramarginal gyrus. This response was consistent across all of the naming events, whether the child learned the object name or not (see Figure 3.4, a). An interaction between learning and chromophore was found in the right fusiform gyrus (see Figure 3.4, b.). Opposite going relationships between HbO and HbR were qualitatively observed in the child's hemodynamic response to naming of learned words, versus during naming events that did not result in learning.

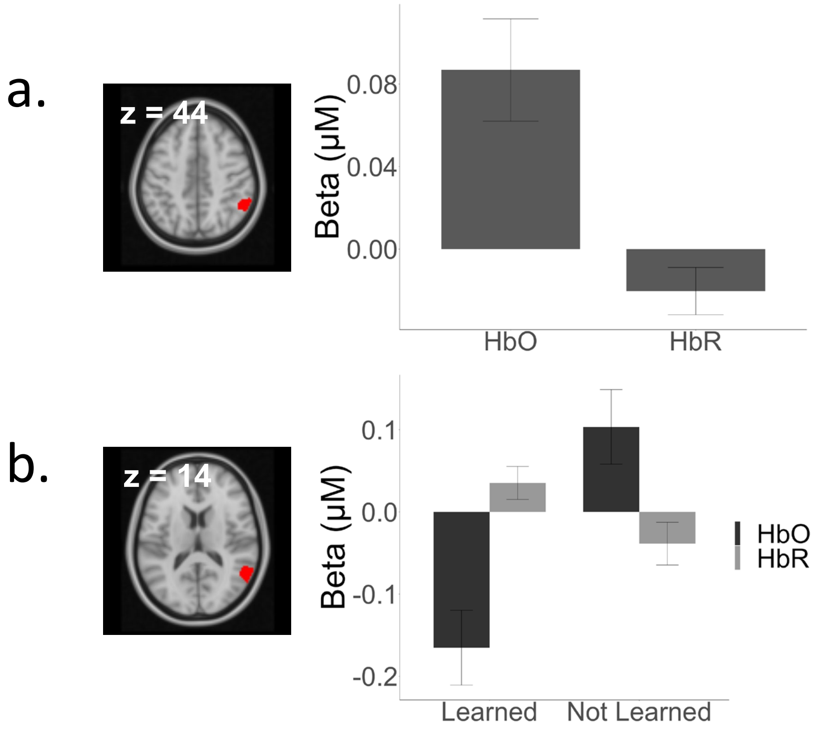

Figure 3.4. The fNIRS effects for children during word learning, including a.) Across all naming, children's brain activity showed a main effect of chromophore in the right parietal cortex. b.) Differential responding in the right fusiform gyrus between naming of words the child learned and did not learn. Bar graphs show the averages and error bars the standard error.

Table 3.2. Child Results – Comprehension Test

				Volume	Centre of Mass		
Effect	**Cluster**	**ROI**	**Hemi**	**(mm^3)**	**x**	**y**	**z**
Chromophore (Hb)	Visual Association Area	NA	L	217	-48	-73	15
Age x Chromophore (Hb)	Angular Gyrus	TPJ	L	232	-44	-59	50

By contrast, no differences were observed during the comprehension test between hearing new words the child recognized, new words the child did not recognize, and familiar words the child already knew. Instead, a main effect of chromophore within the left visual association area suggests that children responded to hearing objects named by processing the possible referents. An interaction between age and chromophore in the left angular gyrus showed that, qualitatively, the younger children had positive HbO activation and negative going HbR suppression in a language processing area in response to hearing a new object named, whereas older children showed the opposite pattern of effects.

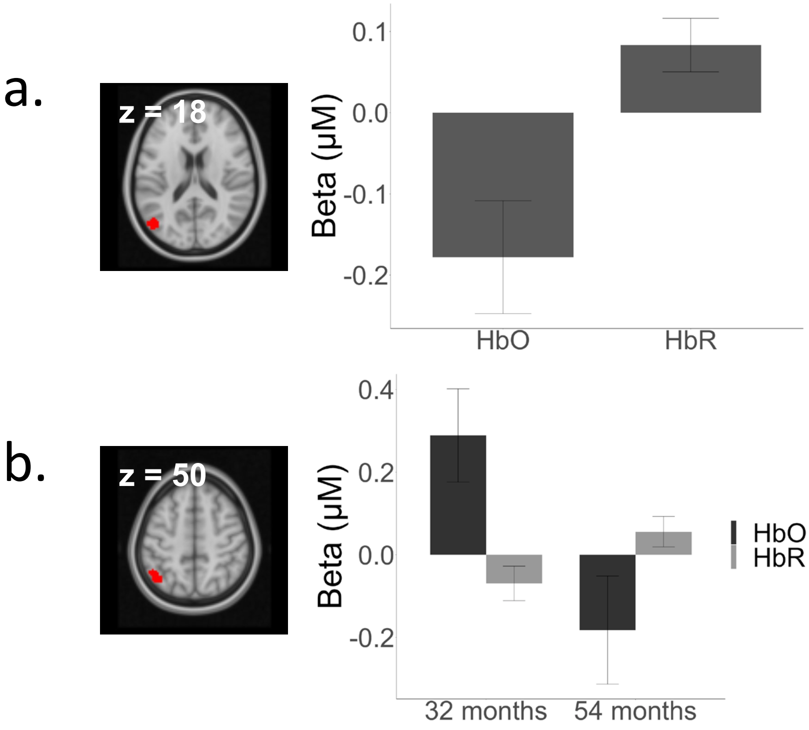

Figure 3.5. The fNIRS effects for children during the comprehension test, including a.) Across all naming, children's brain activity showed a main effect of chromophore in the left visual association area. b.) Differential responding in the left angular

gyrus between younger (32 months) and older (54 months) children’s responses. Bar graphs show the averages. Error bars show the standard error.

Table 3.3. Caregiver Results – Word Teaching

Effect	Cluster	ROI	Hemi	Volume (mm^3)	Centre of Mass x	y	z
Age x Chromophore (Hb)	Dorsolateral PFC	NA	L	85	-3	41	52
Learned x Chromophore (Hb)	Angular gyrus	IPL	L	183	-54	-68	16

These same factors were then examined in the caregiver’s brain activity. In order to examine whether an effect existed based on the child’s learning, within the caregiver’s neural responses to their own naming, we conducted a mixed effects linear model with the child’s age (32 months vs 54 months) as a between-subjects factor, and the child’s learning (learned word vs word not learned) and chromophore (HbO vs HbR) as within-subjects factors. The dependent variable was the mean change in activation (µM) within the caregivers’ hemodynamic response to naming. This model revealed an interaction between the child’s age and chromophore within the left prefrontal cortex that, qualitatively, appeared more positively activated as caregivers taught younger (32-month-old) children compared with older (54-month-old) children. The model also revealed an interaction between the child’s learning and chromophore, within the left angular gyrus. This suggests that caregivers’ social and semantic processing during naming was different when successfully teaching their child.

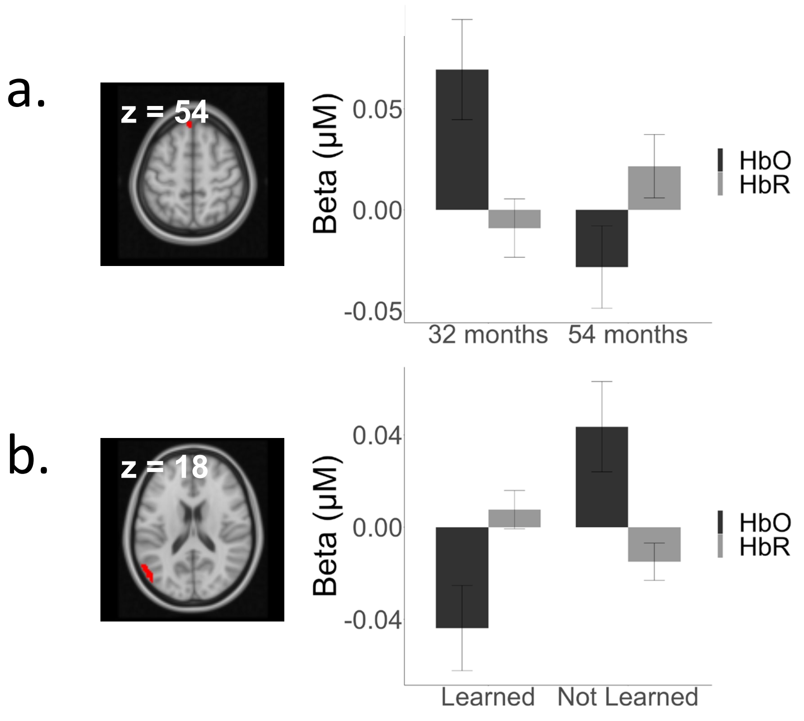

Figure 3.6. The fNIRS effects for caregivers during the word learning interactions, including a.) Caregivers showed heightened activation in the left prefrontal cortex when teaching younger children and b.) Differential responding in the left angular gyrus between naming of words the child learned and did not learn. Bar graphs show the average, and error bars show the standard error.

3.3.2 Discussion

We aimed to explore the neural correlates that impact on early word learning within caregiver-child interactions. To summarize, the behavioural data showed that children varied greatly in the number of words they learned out of eight possible words, that older children learned, on average, roughly an extra 1.5 words compared with younger children, and that children and caregivers did not learn the same words during the interactions. For children, the fNIRS data revealed differences within a right posterior temporal cortical region to naming events and based on learning. During the comprehension test, children showed activation changes in the left visual association area and angular gyrus/temporoparietal junction. Importantly, as caregivers taught their children, changes in activation were observed in the left prefrontal cortex and in the left angular gyrus, the latter based on the child's learning.

In order to determine what associations have been drawn in previous research between each region that showed activation in our task and associated cognitive functions, we conducted a location-based search within the Neurosynth database (Yarkoni et al., 2011) drawing from metanalyses of functional neuroimaging studies that have found activation within the closest range of our centre of mass that returned under 75 studies (between three and four mm of our centre of mass). Only studies that did not focus on specific populations, such as older adults, were considered in these reports.

First, the results for children are considered. The right supramarginal gyrus showed positive HbO activation across all naming events in our study. An ROI-based survey of the literature revealed that the most frequent association was attention (Bankó et al., 2011; Huster et al., 2011; Kühn, Haggard & Brass, 2009; Sato et al., 2016; Schulte et al., 2009; Simanova et al., 2014; Spreng et al., 2014; Vossel, Thiel & Fink, 2006; White et al., 2014), followed by social stimuli, including faces and eye gaze (Decety et al., 2004; Etzel et al., 2016; Frühholz et al., 2011; Iidaka et al., 2006; Spreng et al., 2014; Williams et al., 2005). Activation within this area in response to naming may indicate heightened attention toward instances of naming as well as looking to see where the speaker was attending, for example.

The interaction between learning and chromophore, observed within children's right fusiform gyrus showed differential responding with heightened HbO activation when children *were not* learning from instances of naming. A search within the Neurosynth database that returned previous findings within three mm of the closest available centre of mass to that of our finding indicates that this area has previously been associated in general with social processing (Arsalidou et al., 2010; Blackwood et al., 2003; Contreras et al., 2013; Engell & McCarthy, 2013; Hooker et al., 2006; Kitada et al., 2009; Koike et al., 2016; Sakaki, Niki & Mather, 2011; Schreuders et al., 2018; Tashjian et al., 2018; Thompson et al., 2007; Wutte et al., 2012) and to a lesser extent, with language processing (Habeck et al., 2012; Kepinska et al., 2018; Lin et al., 2011; Majerus et al., 2006; Newman et al., 2013; Quinn, Taylor & Davis, 2017). Interestingly, the interaction suggests that what is happening when the child is not learning from instances of naming may be more interesting in some sense than what is occurring when they are. The activation patterns suggest that during instances of naming, children's brain activity may be more heightened afterward when they are not learning.

During the comprehension test, the main effect found for naming events in the left visual association area, was associated primarily in previous literature with planning, decision making and control (Brown et al., 2012; Corbetta et al., 1998; Grosbras et al., 2001; Guo et al., 2013; Jamadar et al., 2015; Johnson-Frey, Newman-Norlund & Grafton, 2005; Limanowski, Lutti & Blankenburg, 2014; Pergola et al., 2013; Prado, Van Der Henst & Noveck, 2010; Sabb et al., 2007; Salminen et al., 2016; Van der Laan et al., 2012). Differences in age group found within the left angular gyrus were associated with processing and memory recall, largely related to verbal stimuli (Blondin & Lepage, 2005; Braunlich, Gomez-Lavin & Segar, 2015; Dobbins et al., 2002; Donaldson, Petersen & Buckner, 2001; Goghari & MacDonald, 2009; Guo et al., 2011; Han et al., 2010; Kim et al., 2017; Vogel, Petersen & Schlaggar, 2012; Wendelken, Chung & Bunge, 2012; Yokoyama et al., 2002; Zhang et al., 2014). This reveals differences in processing demands in the comprehension test with age but does not provide insight to differences in recognition for learned words and words not learned.

Although no signatures of learning were observed in the inferior frontal gyrus when neural correlates were examined over the whole head, children showed heightened activation within a temporal area that was shown to be functionally associated in general with language. Interestingly, this region was also shown to be more active after a word not learned was named, suggesting a need to further examine the time course of activation during speech both prior to, and following, instances of naming. Also, it suggests that what is occurring when a word is not learned, could in some cases be the more robust marker. The combined association with unspecified social processing suggests that when not learning, children may have spent more time looking at the face or hands and trying to determine the intended referent.

In support of prior work showing that multiple cognitive processes are involved in processing new words, children's neural responses across *all* naming events suggested that heightened attention following a naming event was a primary response involved in mapping new names to objects (Ekerdt et al., 2020). Activation within the right posterior temporal cortex may indicate the child's engagement and sustained attention to their caregiver's labelling of objects during the learning interactions, as well as the possibility that language processing may be distributed more bilaterally for these children, a hypothesis that is not possible to test in the present literature. For example, this effect indicates that children experienced heightened levels of attention to new words embedded in the caregiver's speech stream and suppressed distracting inputs. This activation may contribute to word learning directly, given that children process language more bilaterally than adults, or indirectly, by bringing to the forefront aspects of the environment that have semantic relevance.

As caregivers taught their children, an interaction was found between age and chromophore in the left prefrontal cortex that was more active following instances of naming when teaching younger children. The previous literature suggests that this area is involved in interpreting emotional and social relevance (Caplan et al., 2006; Coaster et al., 2011; Fiddick, Spampinato & Grafman, 2005; Hartwright, Apperly & Hansen, 2014; Kruse et al., 2016; Lewis et al., 2005; Moser et al., 2009; Sakaki, Niki & Mather, 2011; Tamir & Mitchell, 2011; Walter et al., 2009). This could indicate that caregivers worked harder to see what their younger children were thinking or doing when naming the objects. The functional association with social relevance was also true of the area that interacted with children's learning, in the left angular gyrus (Campagne et al., 2016; Canessa et al., 2005; Sakaki, Niki & Mather, 2012; Waugh, Hamilton & Gotlib, 2010). Further, this area was also frequently activated in language tasks (Khader et al., 2010; Muehlhaus et al., 2014; Sabb et al., 2007).

3.4 Analysis 2: Spatial Coordination Between Caregivers and Children During Word Learning

Previous work has shown that spatial cognition supports the communicative inferences that are required for word learning, above and beyond a speaker's explicit labelling of new objects (Benitez & Smith, 2012; Samuelson et al., 2011). Children encode object locations when hearing spoken names, providing an index for binding words to specific referents within working memory (Axelsson et al., 2016; Samuelson et al., 2011; see also Samuelson, Kucker & Spencer, 2017). This then facilitates forming associations between specific new words and their referents. In one prior study that compared their relative effect, spatial location was more heavily relied on than other features that could be used to bind words with objects, such as color (Samuelson et al., 2011). This may reflect the unique constraints on word-object mappings that spatial organization offers, including that no two objects can occupy the exact same locations.

Within naturalistic social interactions, spatial coordination has been shown to predict word learning during play between caregivers and children. When caregivers and children were given two new objects and asked to teach their 17- to 22-month-old children the names, spatial structure emerged spontaneously for some dyads as, overall, caregivers spent the majority of time (0.75%) holding each object in either their right or left hand. However, left-right consistency of manual holding also varied between caregiver-child dyads (0.32-1.0%). Critically, this variation was positively correlated with the number of children's correct responses in a later comprehension test in which they were asked to select each object by name (Samuelson et al., 2011). Thus, early in development, caregivers successfully use encoding of object locations to teach their children new words.

At a behavioural level, we tested a hypothesis that the consistency of object locations during the word learning interactions would predict how many words children learned (Samuelson et al., 2011), using a more difficult task than that used with toddlers, with three and five new objects.

3.4.1 Statistical Analyses and Results

We examined the spatial distributions of object locations in two ways. We derived these measures using the adehabitathr package in R (Calenge, 2006). The first measure indicated how consistently each object remained on the table left or right during the session. This measure was computed based on the probability weighting in each grid square, that indicated how many observations of the object were located there. The left-right consistency was a proportion of the weights summed on the main side / the weights summed on both sides. The second measure required calculating the volume of intersection that each object shared with the volumes occupied by other objects on the table (Millspaugh et al., 2004). A kernel density estimator was used to calculate the volume occupied by each object during each interaction, consolidating the unprocessed data into the space that the object most frequently occupied (Walter, Onorato & Fischer, 2015). The volume of overlap, which we scored based on this measure, was the average proportion of space each object shared with the others on the table. From now we will refer to the averaged volume of intersection as the volume of overlap.

3.4.1.1 Confirmatory Analysis

To determine whether there was a relationship between spatial consistency and overall amounts of word learning, we first conducted a planned correlation between the percentage of left-right consistency of the object locations during the three and five object training blocks and the number of words the child learned during those sessions. The left-right consistency was not significantly correlated with the total number of words learned $r_{35} = -0.185$, $p = 0.274$, and when further broken down by task, it was not associated with learning either in the three-object task, $r_{35} = 0.233$, $p = 0.164$, or in the five-object task, $r_{35} = 0.086$, $p = 0.612$. *3.4.1.1.1 Adaptation: Naming Events*: Intervals of object locations from the onset of naming events through five seconds later were examined separately, revealing no overall relationship between the consistency with which objects remained on the dominant left/right side, $r_{35} = 0.027$, $p = 0.875$. No relationships were found when this was broken down by task, for three-objects, $r_{35} = 0.235$, $p = 0.161$, and for five objects, $r_{35} = -0.054$, $p = 0.751$.

3.4.1.2 Exploratory Analyses

Because object position consistency was not a predictor of word learning, we conducted correlations between the volume of overlap and the number of words learned. The volume of overlap, measured as the kerneloverlappr function in adehabitathr that called the volume of intersection function, did not predict learning overall, $r_{35} = -0.227$, $p = 0.177$, and was a weak predictor of learning in the three-object task, $r_{35} = -0.314$, $p = 0.058$, though not at all in the five-object task, $r_{35} = 0.019$, $p = 0.912$. When only the interval around the naming events was considered, the volume of overlap was not correlated with the total number of words learned, $r_{35} = -0.226$, $p = 0.179$, was robustly correlated with the number of words learned from the three-object set, $r_{35} = -0.429$, $p = 0.008$, and was not correlated with the number of words learned from the five-object set, $r_{35} = -0.026$, $p = 0.878$. This finding with three objects withstood a Benjamini-Hochberg statistical correction for six comparisons.

Because neither the left-right consistency of object locations nor the volume of overlap between them predicted the overall learning, and because visual inspection of the data revealed that spatial structure varied considerably between objects in the same interaction, we next examined whether these spatial measures and other relevant variables could predict *which individual word*s were learned by each child. We conducted a set of mixed effects logistic regression models using the glmer function in R programming software. In each case, the dependent variable was modelled with a binomial probability where possible outcomes were “learned” or “not learned”. (For a practical review of applying mixed effects logistic regression models, see Sommet & Morselli, 2017; for a theoretical review of applications to linguistics research, see Quené & Van den Bergh, 2008). Statistical significance was met if $p < 0.05$. Spatial variables were taken from five seconds following the naming events.

In the first logistic regression model, fixed effects included the number of objects as a factorial predictor, and the volume of overlap as a continuous predictor. The child’s age group was added as a covarying factorial predictor into the model. A dyad (participant) identifier was included as a random effect. A main effect of age was found, $z = 2.08$, $p = 0.038$ and number of objects, $z = 2.08$, $p = 0.038$ and a main effect of was found of volume of overlap, $z = -2.83$, $p = 0.005$ as well as a marginal interaction between the number of objects and volume of overlap, $z = 1.78$, $p = 0.075$, see Figure 3.5.

To rule out the possibility that the object position consistency would predict which words children learned, we conducted an additional, identical logistic regression model with left-right consistency. The only significant predictor in this model was age, $z = 2.18$, $p = 0.029$. No other predictors were significant, including the number of objects, $z = 0.64$, $p = 0.522$, the left-right consistency, $z = 1.30$, $p = 0.193$ or the interaction between number of objects and left-right consistency, $z = 0.05$, $p = 0.957$.

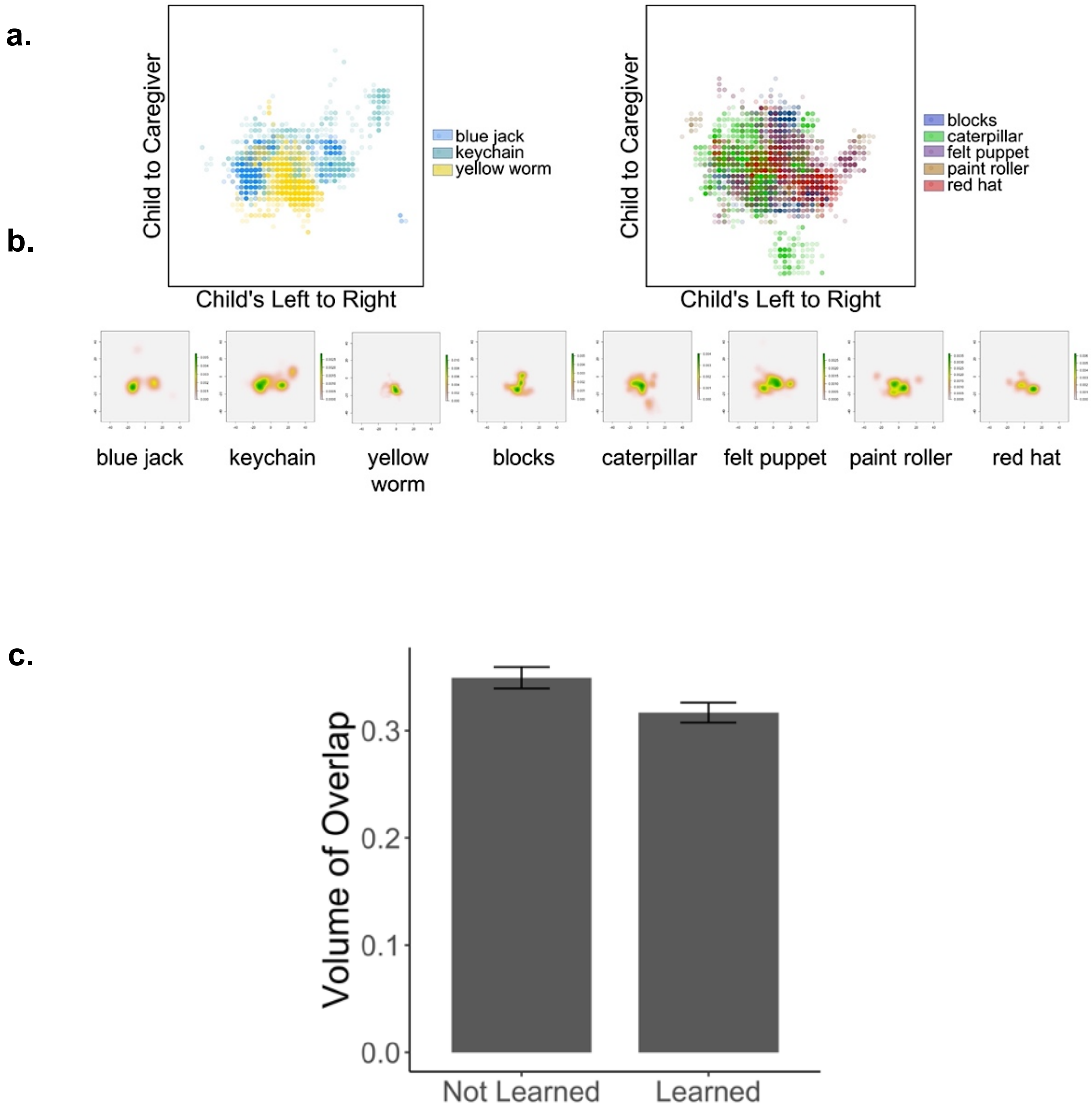

Figure 3.7. The impact of volume of overlap on the child's learning, including a.) Examples of the spatial distribution and overlap of individual toys within an interaction including three objects (left) and an interaction including five objects (right). Units are normalized pixel coordinates referenced to the table midpoint, with the x axis showing the table from the child's left to right and the y axis showing the table from the child's side (lower) to the caregiver's side (upper). b.) Plots of the volumes of locations occupied by each toy (calculated using a kernel density estimator). (c.) A bar graph showing the difference volume of object overlap between words that were learned and words that were not learned.

Additional models revealed that neither the caregiver's number of naming events for each object, the child's number of naming events for each object, or the child's vocabulary (measured by the BPVS-3 survey) predicted children's word learning. Within this model, age did not predict learning, $z = 0.14$, $p = 0.889$, nor did children's naming, $z = 1.02$, $p = 0.307$, nor did caregivers' naming, $z = 0.90$, $p = 0.367$. No interactions were found between age and child's naming $z = 0.83$, $p = 0.407$, between age and caregiver's naming, $z = 0.53$, $p = 0.598$, between the child's and caregiver's naming, $z = -1.10$, $p = 0.272$, or between age, the child's naming and the caregiver's naming, $z = -0.66$, $p = 0.511$.

A fourth model explored the child's age (scaled and centred), and the child's standardized BPVS-3 score as fixed effects, and the number of objects as a random effect. The dependent variable for this model was the BPVS score collapsed across word learning "successes" and "failures", for each child. Within this model, older children were more likely to learn words, $t_{35} = 2.62$, $p = 0.013$, while children's BPVS-3 score was a marginally significant predictor, $t_{35} = 1.96$, $p = 0.058$. There was no interaction between the child's (standardized) BPVS score and age, $t_{35} = 0.70$, $p = 0.487$.

3.4.2 Discussion

We aimed to determine whether caregivers who use spatial consistency, for example, lining objects up on the table, would be able to more effectively teach their children new names than those who don't. Results showed that object locations varied within dyads, thus, overall, caregivers did not seem to make an intentional effort to keep toys in a consistent region of space. However, the names of those objects that did occupy unique regions of space during a naming interaction were more likely to be learned, compared with those that did not. As evidenced by the lack of relationships between the left-right consistency of object locations and learning, where objects moved around during the interaction did not matter, in and of itself. What mattered was how consistently the object occupied unique regions of space that did not overlap with those frequently occupied by other objects, which we have referred to as the volume of overlap between objects.

Computational modelling approaches suggest that, from the child's perspective, spatial overlap among objects may interfere with mapping each name to a specific object because spatial location fails to provide a unique index for word-object mappings (Samuelson et al., 2011). This is consistent with studies that have directly manipulated spatial consistency or spatial predictability by switching locations between objects rather than merely varying the object locations (Benitez & Smith, 2012; Samuelson et al., 2011). In such cases, the word learning was poorer. In retrospect, the distinctions between right and left that were analysed in Samuelson et al. (2011) could have been more appropriate for a task involving two objects, and that in a more complex task involving three and five objects, this distinction may not capture the encoding of location that ultimately predicts children's learning. The present work reveals that children's encoding of spatial location is heavily influenced by the uniqueness of the locations each object occupies throughout the interaction, and that this has an influence on how effectively children learn words. Our findings suggest that distinctiveness of object locations functions as a key aspect of social coordination between the caregiver and child that supports word learning. From the child's perspective, distinctiveness of object locations establishes a common ground with their caregiver in which to index names with intended referents.

3.5 General Discussion

The present chapter examined the neural and behavioural correlates of early word learning during a play interaction in which caregivers taught their 2- to 5-year-old children the names of new objects. The findings revealed that neural responses that were time-locked to naming differed between words the child learned and words they did not learn. Within the child's brain activity, responses to naming in general revealed underlying responses to naming events within the right supramarginal gyrus, associated with attention, while the neural correlate that differentiated naming of learned from words not learned was most closely associated, based on similar findings in the prior literature, with social and language processing. During the comprehension test, all children responded to hearing the name of the referent they would need to retrieve with activation that elicited action planning and decision. However, an interaction between age and chromophore suggested that younger children put more effort into recall, trial by trial.

As caregivers taught their children, neural responses to their own naming of the new objects revealed interactions between (the child's) age and chromophore, and between the child's learning and chromophore. These findings revealed that the neural signatures of early word learning were evidenced within both the caregiver's and the child's brain activity, supporting the general view that neural signatures of learning in real-world contexts can be explored at the level of social networks (Hasson et al., 2012; Wass et al., 2020). As these results show, when caregivers named objects, differing neural signatures were observed depending on the child's age, and whether the child was learning as the objects were being named. This suggests that caregivers' own processing was different when their children were learning from their object naming. However, in the present study, it is not possible to quantify whether early word learning that emphasizes the combination of neural signatures between an adult and child would uniquely signify learning is taking place. The impact of coordination between caregivers and children was also observed at a behavioural level. The second set of analyses, examining the impact of object locations during the interactions and during naming on learning, revealed that spatial coordination between the caregiver and child predicted which words the child would learn. Specifically, the volume of overlap each object shared with other objects on the table interfered with mapping words to objects, and this effect was most robust for the specific object names children were more likely to learn. There was no evidence to suggest that coordination in the volume of overlap was intentional, as in, that caregivers and children deliberately planned to keep objects in distinct regions while teaching and learning. Yet, this variable still affected children's learning.

Key findings in the study differed from the original predictions, while remaining consistent in an unexpected way with the literature that had motivated these predictions. The first prediction we tested was that differential responding in the right or left inferior frontal gyrus (IFG) would be observed in children's hemodynamic responses to learned versus words not learned. This prediction was based on previous work with children and adults, suggesting that the IFG shows different patterns of activation between new and familiar words and is generally activated in language processing tasks. Instead, differential responding between learned words and words not learned was observed in the right inferior temporal cortex, associated with both social processing and language. Therefore, while the original prediction was not supported (no differences were observed in the IFG between naming of learned words and words not learned in the whole-brain analysis), the finding is consistent with the hypothesis that differences in language-related processing would be observed based on the child's learning. Informatively, this correlate of learning in the right temporal cortex had a double association with social processing, suggesting a possible close relationship between language-specific and general social processing within the child's neural signature of learning.

Another prediction, that the child would learn more words when objects were consistently placed during learning, directly matched a previous study in which children's overall number of correct responses in a comprehension test were correlated with the consistency in which caregivers held each object in their right or left hand (Samuelson et al., 2011). However, in the present study with older children and involving three and five objects the left-right consistency of object locations during the learning interactions bore no statistical relationship to the child's learning. Thus, this prediction was not supported. However, the finding within the exploratory analysis, that the volume of overlap between objects predicted learning, was fully consistent with previous studies that had motivated our original prediction. In each of these studies and an accompanying computational model, the object locations had either involved distinct locations or had involved direct overlap in location while the child was learning the names (Benitez & Smith, 2012; i.e., either each object remained in the right or left hand or switched hands frequently, Samuelson et al., 2011). Our finding suggests that, as hypothesized, spatial predictability did impact on children's word learning, but contrary to what we had expected, the children in our study encoded spatial location relative to the locations of other objects, relying on the distinctiveness, rather than the consistency of the objects' locations. Finally, contrary to our initial prediction, no spatial variable directly predicted individual differences in children's learning. Instead, the volume of overlap predicted which words the child would learn.

With regard to the fNIRS measurements taken during the study, there are some technical limitation to consider. As coverage of the cortex is greatly limited in hyperscanning paradigms that use a single system to record from two individuals, employing short distance sources that can be used to regress physiological signals out of the data is relatively uncommon in hyperscanning studies. Rather than using a signal regression to remove globally presenting physiological noise from the data, future work could employ measurements at multiple lengths from the light source, creating a more precise means to regress this noise from the task-relevant signal components (Berger, Goodwin & Gaudet, 2012).

The study also presents directions for future work. The present study did not measure what participants were doing with the objects, such as where they were looking or when they were interacting with the objects. Future work could explore relationships between the caregiver's and child's holding of objects during the interactions and their neural responses to naming of learned words and words not learned. For example, it is possible that a broader range of spatial metrics impact on learning when looking and holding are taken into account, such as object position consistency.

Our neural findings suggest that within interactive, real-world paradigms, word learning is reflected in children's neural responses to naming of learned words and words not learned, as well as those of their caregiver. Interestingly, what happened when children were not learning at times became the most salient marker. For example, evidence of increased neural activation during instances of naming in the right temporal cortex for children, and in the left parietal cortex for caregivers each occurred during naming of words not learned. Further, the volume of overlap had a negative impact on children's likelihood of learning any given word. Collectively, this may indicate that caregivers and children had more seamless experiences or put more cognitive effort into predicting each other's behaviour before instances of naming (rather than during) when learning was successful. This would explain why brain activity after naming was greater after naming of words not learned, because of the additional effort required to make sense of conversational inputs in those cases. One hope is that this work will provide future research with grounds for confirmatory testing of these key findings. Additionally, it is hoped that the task described here will provide the foundation for future development in naturalistic testing paradigms to investigate the neural correlates of early word learning.

Chapter 4 – Neural Coherence During Interactive Word Learning

4.1 Introduction

The early development of brain structure and function is shaped by social interaction (Belsky & De Haan, 2011; Fox, Levitt & Nelson, 2010), much of which takes place between caregivers and children. Crucially, caregiver-child relationships shape early language development through day-to-day interaction in which caregivers label, hold and talk about objects (Cartmill et al., 2013; Kalagher & Yu, 2006; Yu & Smith, 2013; Yu et al., 2009). However, the underlying neural states that facilitate word learning during interactions are unknown (Wass et al., 2020). In addition to momentary changes in brain activity during discrete naming of objects, relationships in brain activity between a caregiver and child that are sustained through longer periods of time may facilitate children's mapping of words to objects.

Temporal alignment in neural signals, like alignment in behaviour, can indicate that individuals are representationally in synchrony. Moreover, neural alignment may be a more sensitive measure of whether interacting partners comprehend each other and are experiencing a similar underlying representation. As an analogy, at a behavioural level, interacting partners align their gaze, movements and even choice of words to enhance the information shared through speech (Garrod & Pickering, 2009). This ultimately results in increased sharing of visual attention during dialogue, meaning that individuals are viewing the same situation in a more similar way during interactive conversation (Richardson, Dale & Kirkham, 2007). Behaviorally, spontaneous alignment that occurs in verbal and non-verbal communication suggests that entrainment between individuals leads to better communication.

However, at a neural level, entrainment in brain activity between a speaker and listener through time has been shown to play a direct role in speech and language comprehension. For example, when English-speaking participants listened to a telling of an unrehearsed story in English, brain activity between the original speaker and the listeners was correlated, time point to time point, within language and other functional neural networks. In other words, neural (BOLD) signals measured from the speaker's and listener's brain activity moved together continuously through time during the telling of the story. Within frontal areas, the correlation values, or the robustness of alignment in the neural signals, predicted the detail in which the story in English was later remembered. These correlations were not observed when they heard a story told in Russian (Stephens, Silbert & Hasson, 2010). Therefore, the temporal alignment of neural signals suggests that individuals are sharing similar representations that can be driven by speech, and that the robustness of entrainment predicts the depth of comprehension.

Temporal relationships in brain activity have been observed between participants as they *directly* engaged in dialogue, using hyperscanning, a technique of collecting neuroimaging data on two or more individuals as they interact with the intention to compare relationships in their brain activity (for a detailed discussion, see section 1.4). Liu et al. (2019) found that entrainment within the right posterior temporal cortex was more robust when participants who alternated in speaking and listening roles were given sentences with a common syntactic structure, meaning that similarity in the sentence structure that each individual spoke facilitated the other's comprehension. Neural entrainment was measured within decomposed time-frequency space, after the fNIRS signals had been separated out by frequency components, in order to measure the neural coherence (Cui, Bryant & Reiss, 2012). Coherence within the posterior temporal cortex specifically predicted how well listening participants were able to determine whether pictures they were shown matched the speaking participant's description. These studies indicate that cross-correlations in neural signals correspond to successful speech processing, and that wavelet coherence in particular can quantify this effect.

Early in development, neural entrainment between interacting adults and infants has been shown to increase during face-to-face interaction, as opposed to looking away (Leong et al., 2017), or conversing apart (Piazza, et al., 2020). Overall entrainment can be statistically explained, in part, by entrainment that occurs specifically during communicative behaviours, such as establishing eye contact, or smiling (Piazza et al., 2020). For children ages 7-13, neural coherence has been shown to increase during cooperative, joint action between caregivers and children, when compared with competitive or individual play conditions, or interactions with an adult stranger. For example, neural coherence within frontal areas was higher when children and caregivers were told to press a button at the same time in a game, as opposed to competing to press the button first (Miller et al., 2019; Reindl et al., 2018). Taken together, this research suggests that correlations in brain activity are associated with behavioural coordination.

One remaining question, however, is whether neural coherence between caregivers and children could predict children's word learning. Another is *why* neural coherence would impact on learning. No relationship between coherence and interactive word learning has been empirically established. However, coherence during periods of eye contact and during infants' smiling have been shown to contribute uniquely to neural coherence between and adult and children, suggesting that communicative behaviours bring adults and infants more in alignment with each other (Piazza et al., 2020). It has been speculated that this may be because ostensive social cues increase neural entrainment by functioning as "edges" that reset temporal patterns of brain activity between two individuals, aligning them and helping to establish common representational ground (Wass et al., 2020). Further, the processes elicited in predicting or processing another person's behaviour, moment-to-moment, may increase coherence between interacting partners on a moment-to-moment basis (Hamilton, 2021). For example, successful interaction might require keeping track of how a conversation partner is viewing or holding objects, which may then impact encoding which spoken words correspond to which referents. Communicative behaviours that occur during face-to-face interactions may directly boost alignment in brain activity, setting the conditions for learning new words. On the other hand, neural coherence at the beginning of interactions has been shown to predict individual differences in later levels of cooperation between caregivers and children as they jointly played a simple game (Reindl et al., 2018), indicating that the impact of coherence as a shared state can set the conditions for cooperation between caregivers and children, even when disconnected in time. Based on these influences between neural coherence and communicative behaviour, one would expect that dyads who show greater neural coherence will also teach and learn more effectively.

No study to date has directly investigated whether neural coherence between caregivers and children during an openly structured interaction can predict word learning. Therefore, we aimed to answer whether neural coherence between caregivers and children would predict how many words the child learned. Functional near infrared spectroscopy (fNIRS) was measured from both the caregiver and child as the caregiver taught their child the names during play. The main research question was: will neural coherence between caregiver and child during play with new objects be correlated with the number of object names the child learns?

However, in order to investigate whether neural coherence can predict how many words were learned, we first needed to establish which areas of the cortex were showing neural coherence that was relevant to interactively teaching and learning. Neural coherence can be driven by background brain activity that is unrelated to a task, by similarity in patterns of physiology such as breathing or heart rate, or even by sharing an environment or doing the same task, regardless of whether the individuals are interacting (Burgess, 2013). Prior work has used a resting baseline to control for alignment in neural signals that occurs as a result of basic physiology, as opposed to legitimate experimental results. For example, participants may be asked to sit still with their eyes closed for a period of a few minutes before completing a task in order to establish a resting baseline (Cui et al., 2012; Liu et al., 2019). However, coherence that is higher when interacting in the same task than during rest can include multiple underlying causes. As stated above, these can include similarity in underlying brain activity that results from merely doing the same task as another person, and also from interacting with another person. Thus, even the aforementioned work that identified and only retained channels with coherence higher than the resting baseline, also compared coherence between different conditions for those channels that were higher in coherence than the resting condition. For example, after retaining channels that were higher in coherence compared with resting baselines, Cui et al. (2012) then tested for differences in these channels between independent, competitive and cooperative game play during a computer-based task. Only those channels that were higher in coherence in the cooperative or competitive condition were reported. This indicates that interacting with the other person, not merely doing the same task that they were, drove the coherence. Alternatively, Liu et al. (2019) compared coherence to a resting state in order to determine which frequency ranges to examine in the hyperscanning analysis. Those frequencies that differed from the resting baseline were included in the analysis.

It is more difficult to establish the range of experimental conditions necessary to distinguish between the relevant impact of the task structure and the impact of interaction on neural coherence within openly structured interactions. This is because, for some tasks like teaching and learning in-person, there must be two interacting partners rather than two separately engaged individuals whose brain activity can be recorded in parallel. Therefore, it would be difficult to establish complete control conditions with independent learning. Either the learning phase would not involve instances of naming as there would be no conversation partner, or the child would be observing two different people teach and learn but not participate in learning themselves, for examples. Prior work that studied naturalistic interactions where partners did not have distinct teaching and learning roles have used pseudo-resting conditions to compare with different types of interactions in the experimental task. For example, Liu et al. (2016) had participants discuss everyday topics during the rest period, and then compared coherence between cooperative, obstructive and independent Jenga play and the rest conditions. With adult-child interaction, additional considerations can make it difficult to match the stimuli between interactive tasks and resting baselines. For example, in the non-interactive condition by Piazza et al. (2020), the experimenter who read to the child during the interaction condition turned away from them and read to another adult while the child could see, and while the child interacted with a caregiver instead. This more variable type of baseline shows why it is difficult to establish task-related structure to baselines in naturalistic interactions, and the prior work shows why it is important to examine resting baselines along with interactive and non-interactive task conditions. Finally, proper resting baselines have not previously been used with children in hyperscanning studies because it is not typically possible to ask infants and young children to sit still with their eyes closed.

One common and parsimonious way that interaction-derived coherence is validated against these background factors is by comparing data from participants who were interacting, compared with participants who completed the same task but who were not interacting with each other (Hirsch et al., 2017; Liu et al., 2019; Reindl et al., 2018). Within these studies, channels that did not show a significant difference between actual and scrambled participant pairs were often excluded from further analyses. One benefit of this approach for word learning interactions is that it can separate coherence that occurs as a result of interacting with that person, from coherence that occurs just from generally participating in the same task of teaching and learning new words. Ideally, like in adult studies, these types of baselines may be used in combination with resting baselines (see Liu et al., 2019), dividing baseline physiology from coherence driven by task structure, and coherence driven by task structure from coherence driven by the synchronizing of representations between two interacting partners. Because previous research suggests that neural coherence is robustly related to moment-to-moment behaviours during interactions (Piazza et al., 2020) and because we were interested in establishing whether neural coherence during interactions could predict how many words children learned, we applied a baseline of scrambled caregiver-child pairs to validate our coherence measure, prior to exploring any relationships with behaviour. If neural coherence between interacting caregivers and children was significantly greater than neural coherence between scrambled dyads within any channels, those channels would be tested for correlation with the child's word learning.

4.2 Methods

4.2.1 Participants

The data analysed in this chapter is the same as that collected in Chapter 3, consisting of a final sample of 19 32-month olds (*M* = 32.5 months, *SD* = 52 days) and 18 54-month olds (*M* = 53.7 months, *SD* = 58 days) and their caregiver. Participants were recruited from a city and surrounding areas in the UK.

4.2.2 Stimuli and Apparatus

Caregivers and children were seated across from each other at a table. Caregivers could see the names of photos of objects on a felt board that was positioned behind the child. The new stimuli consisted of eight unusual toys, including a three-object set of a yellow squishy toy, a large decorative key ring, and a large blue jack, and a five-object set of a wooden caterpillar that had been painted solid green, a miniature paint roller, a red mesh structure, brightly coloured blocks, and a small felt puppet, (see Fig 4.1). fNIRS data was recorded, as described in Chapter 3.

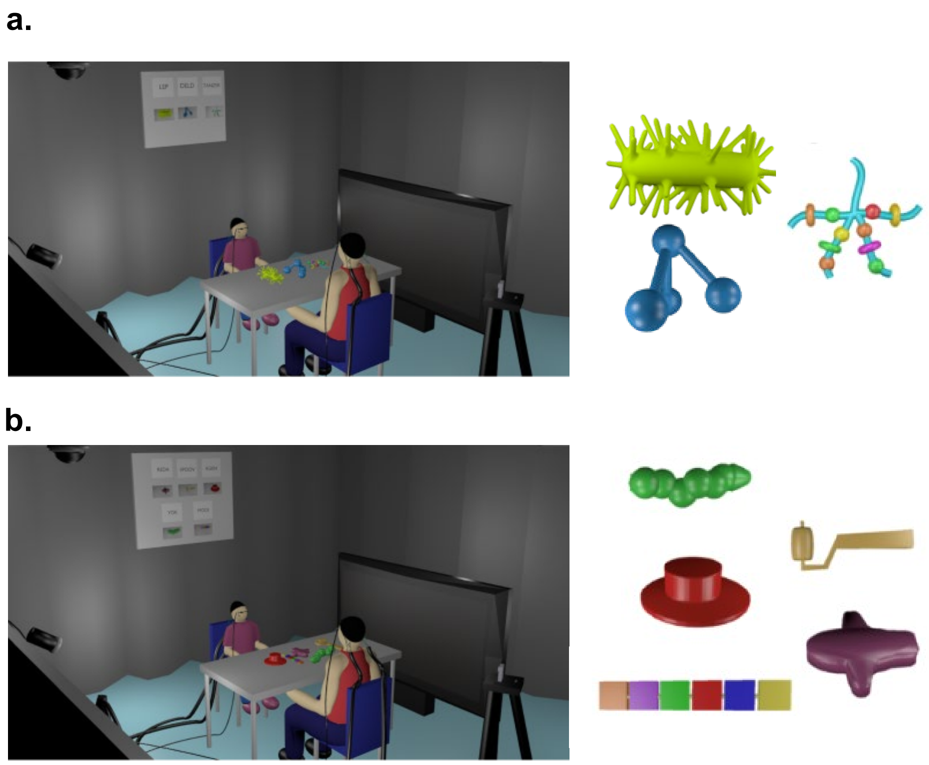

Figure 4.1. The stimuli and apparatus, including a. left) A model of the experiment setup and the first task in which the caregiver taught their child the names of three new objects. (right) The objects are pictured. b. (left) A model of the second task in which caregivers taught their children the names of five new objects. (right) The set of five new objects is shown. The caregiver could see photos and names of objects on a felt board positioned behind the child. fNIRS data was recorded from both the caregiver and child during these tasks.

4.2.3 Procedure

For a detailed description of the procedure, see Chapter 3, and for relevant methods development, see chapter 2. To summarise, the caregiver and then the child were fitted with an fNIRS cap and signals were inspected for quality. At the beginning of the session, the child completed the British Picture Vocabulary Scale (BPVS-3) survey. Caregivers then taught their children names of eight new objects. For the first three new objects, the caregiver was given 2.25 min (45 sec per object * 3 objects) to interact with the child and teach the child the names of the objects. In total, the participants were given 4.5 min to interact with the first three new objects. The experimenters then began the comprehension test. The child and caregiver (while facing away from each other) were presented simultaneously with pairs of the familiar objects alternating with pairs of the new objects, by the experimenters.

During each trial, the experimenter asked the child to retrieve the target object. The child retrieved the object while the caregiver, unable to see the child and out of their child's view, pointed to the correct photo on a display card. In total, participants completed twelve comprehension trials: six trials where they were asked to select a familiar object and six trials where they were asked to select a new object. The entire word learning task and comprehension test were then repeated with another set of five new objects. Because we had increased the number of objects, participants were given 3.75 minutes (45 sec per object * 5 objects) in each word learning session for a total of 7.5 min. The second comprehension test consisted of twenty trials: ten trials in which the participants were presented with a pair of familiar objects and asked to select one of the objects alternating with ten trials in which they were presented with a pair of new objects and asked to select one of the objects. Finally, we tested whether the caregiver and child could produce the names of the new objects from memory. The caregiver faced away from the child. An experimenter then asked the child the name of each new object, saying "Can you tell me what this is called?" The child was given a few seconds to answer. Caregivers completed an identical production test, to assess their learning.

4.2.4 Determining the fNIRS Geometry

The geometry in which optodes were arranged was designed to cover the main regions of interest from previous studies that bore the highest relevance to the present study (for a full methods description, see Wijeakumar et al., 2015), including studies from social interaction, language and word learning, and visual working memory. The final geometry covered the superior frontal cortex, inferior frontal gyrus, posterior temporal cortex and parietal cortex (for a full list of studies used to design the geometry, see Appendices A and B). To view the cap geometry used, see Figure 4.2.

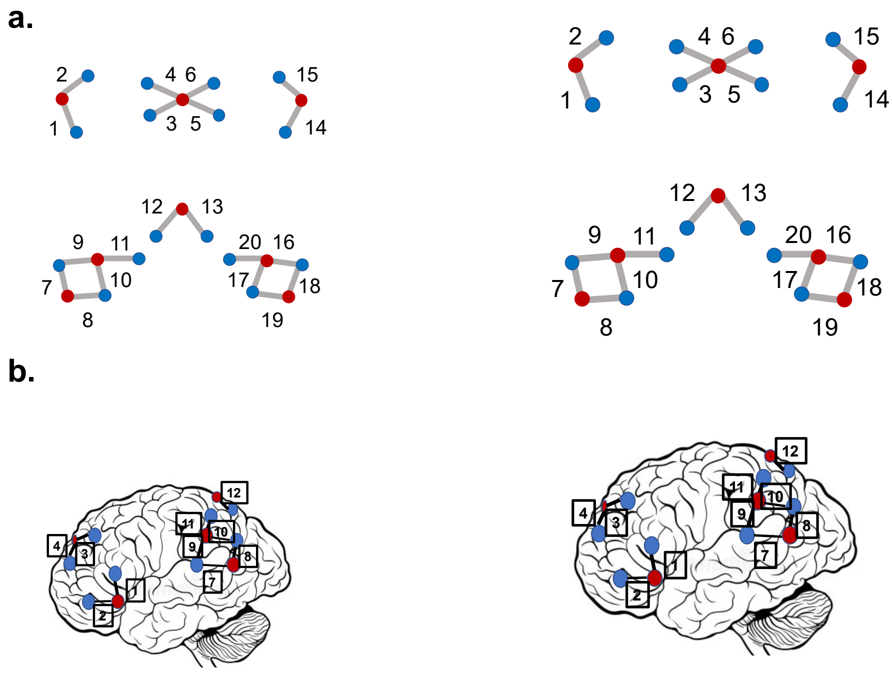

Figure 4.2. The fNIRS coverage, including a.) the NIRS geometry array for child (left) and caregiver (right) with the front of the head positioned toward the top. Six clusters are arranged bilaterally covering superior frontal, prefrontal and inferior frontal, posterior temporal and parietal areas. Sources are shown in red and detectors are shown in blue. b.) Cortical coverage of the cap geometry laid over the left hemisphere of a child brain (left) and adult brain (right). It is important to note that the gap geometry shown here is the same as in Chapter 3.

4.2.5 fNIRS Pre-Processing

The data collected from the caregiver's head and the child's head were pre-processed separately, using a method described in more detail in Chapter 3. Data was pre-processed with nearly the same steps for both analyses. Signal processing on the unprocessed data was conducted using custom parameters within the HomER2 software (Huppert et al., 2009). Signal measurements were omitted entirely if they did not fall between 70 dB and 140 dB, dB=20*LOG10(*y*), where y is measured as light intensity, converting the light intensity measures into optical density (OD) units. The pre-processing pipeline identified motion artefacts based on changes in optical density greater than 0.4 au within one second, or with a SD greater than 50. A targeted principal component analysis (tPCA) corrected for components identified as motion (Yücel et al., 2014). Data slower than 0.016 Hz and faster than 0.5 Hz were then bandpass filtered and the remaining motion artefacts removed by channel. After processing the data, concentration values for oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR) were computed for each channel using the modified Beer-Lambert law.

4.2.6 Wavelet Coherence Analysis

In order to measure the alignment in brain activity between the caregiver and child across each word learning interaction, we conducted a wavelet transform coherence (WTC) analysis on corresponding channels recorded from each person. The coherence measure requires the NIRS signal to be decomposed into time-frequency components. For example, the function identifies and separates the pattern of oscillatory activity in the unprocessed fNIRS signal at different periods. Then a wavelet function is used to bandpass filter the original time series within those periods. The Wavelet Transform Coherence (WTC) function then calculates coherence as the cross-correlation *between* the two signals that had been transformed into time-frequency space (for more in-depth explanation of this analysis, see Grinsted, Moore & Jevrejeva, 2004). We used the WTC toolbox in MATLAB to calculate the coherence values. This toolbox has previously been used to measure neural coherence between interacting partners using fNIRS (Cui et al., 2012; Hirsch et al., 2017; Miller et al., 2019), indicating that it was appropriate for our task.

Based on systematic visual inspection of time-frequency plots of our fNIRS coherence data, as well as previous literature (Hirsch et al., 2017; Liu et al., 2019), we determined that coherence in our data was most concentrated over frequencies of 6 to 30 seconds, or 0.16-0.03 Hz. Thus, the mean coherence level, across all frequencies within this range, was computed for each dyad and each interaction, including the earlier and later block with three objects and an earlier and later block with five objects (Cui et al., 2012). Where possible, we padded the time window around which coherence was calculated with 40 seconds of data at the beginning and end of the block in order to ensure continuity of the wavelets at longer periods.

4.2.7 Validation of Coherence Via Permutation

One concern with coherence analyses is the likelihood of finding coincidental coherence that is due to participants' engaging in a similar task (Burgess, 2013). In order to ensure finding coherence levels that were driven by *interactive* learning and not merely through engaging in the same task, we generated an artificial dataset based on pairing time series within corresponding channels, between all combinations of caregivers with other children in the study who were not interacting with them (for studies which have used a similar baseline, see Hirsch et al., 2017; Miller et al., 2019; Reindl et al., 2018). Before conducting this analysis, we trimmed all recordings to be the same length between participants in order to fit an autoregressive (AR1) function to the time series. Two participants were omitted from this baseline, though not the later analysis, due to having recordings that were shorter than the others. Ultimately, we calculated the mean within, and then across, all of the scrambled pairs for both tasks, resulting in a single value against which to compare our data for a statistically significant difference.

4.2.8 Statistical Analysis

We first aimed to determine whether any of our channels contained coherence above the baseline, using a protocol that had been directly used in previous studies that were similar to ours (Cui et al., 2012; Hirsch et al., 2017; Miller et al., 2019; Reindl et al., 2018). The justification for using this procedure is that a null distribution based on scrambled caregiver-child pairs would establish what aspects of coherence were due to underlying physiology and task structure, rather than the interactive teaching and learning process. Therefore, we conducted 20 one-sample t-tests to compare the mean coherence value in our data for that channel across the entire session, with the mean of the null distribution. HbO values were selected this analysis, based on visual inspection of plots comparing overall distributions and observed and scrambled data for HbO and HbR that indicated that HbO values showed the greatest difference between actual data and the baseline. A Benjamini-Hochberg correction for multiple comparisons was applied to the t-tests (alpha = 0.1). For the output of this comparison, see Figure 4.3.

4.3 Results

The mean of the observed data of caregivers and children who were interacting ($M = 0.325$ $SD = 0.006$) was similar to the mean of the null distribution of scrambled caregiver-child channels, comprised of all possible combinations of caregivers and children who were engaged in the same task but who were not interacting ($M = 0.320$, $SD = 0.002$). Following a Benjamini-Hochberg correction for multiple comparisons, we did not find a difference in coherence between any channel in the interacting pairs and the mean of the scrambled caregiver-child pairs ($p > 0.05$). The threshold for a significant p-value set by the Benjamini-Hochberg correction was 0.005, for the p value of highest significance. Here are the actual results, by channel, none of which met the significance level required by the correction. For channel 1, $t_{35} = 0.12$, $p = 0.902$. For channel 2, $t_{36} = 0.26$, $p = 0.799$, For channel 3, $t_{31} = 1.10$, $p = 0.279$. For channel 4, $t_{31} = 0.885$, $p = 0.383$. For channel 5, $t_{30} = 0.69$, $p = 0.496$. For channel 6, $t_{30} = 0.69$, $p = 0.496$. For channel 7, $t_{32} = 1.61$, $p = 0.117$. For channel 8, $t_{30} = 0.62$, $p = 0.542$. For channel 9, $t_{30} = 0.85$, $p = 0.399$. For channel 10, $t_{33} = 0.12$, $p = 0.908$. For channel 11, $t_{35} = 0.41$, $p = 0.685$. For channel 12, $t_{27} = 1.14$, $p = 0.261$. For channel 13, $t_{28} = 0.883$, $p = 0.930$. For channel 14, $t_{33} = 0.47$, $p = 0.644$. For channel 15, $t_{36} = 0.93$, $p = 0.359$. For channel 16, $t_{32} = 0.14$, $p = 0.890$. For channel 17, $t_{35} = 0.98$, $p = 0.332$. For channel 18, $t_{31} = 0.25$, $p = 0.801$. For channel 19, $t_{30} = 2.39$, $p = 0.023$. For channel 20, $t_{33} = 0.210$, $p = 0.835$. As the results show, none of the channels had significantly higher coherence compared with the baseline, indicating that there was not enough power to detect coherence increases based on what was happening in interactions between caregivers and children. Therefore, we did not continue to explore brain-behaviour relationships within the interacting pairs in any channels.

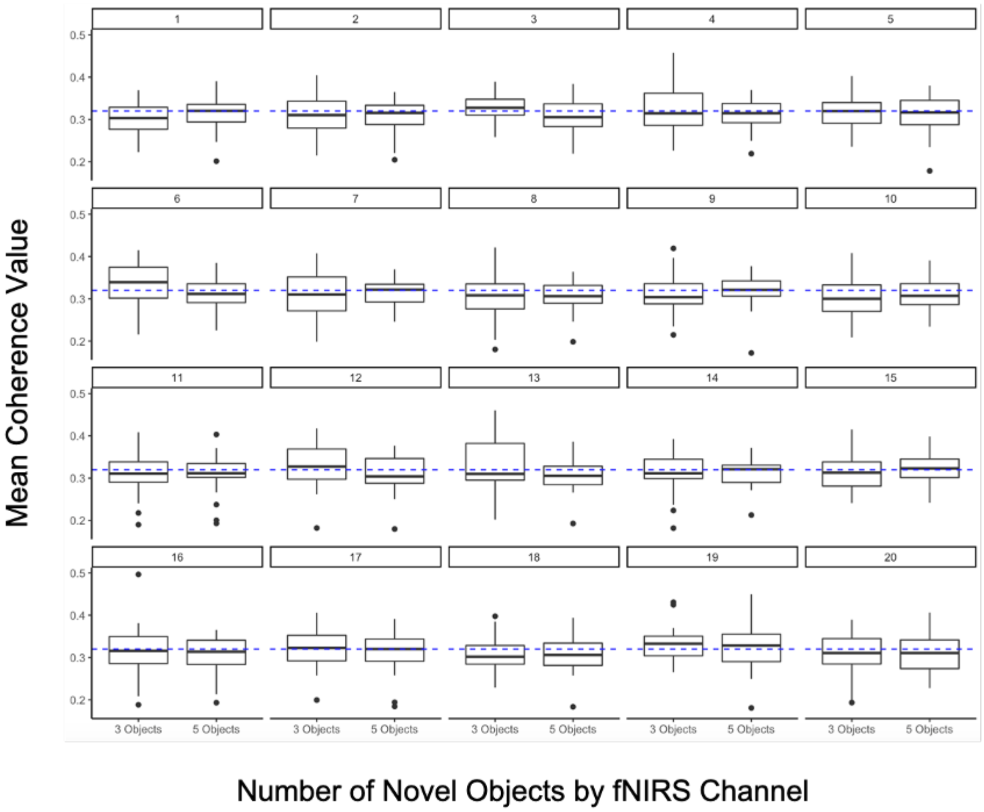

Figure 4.3. Boxplots of the coherence data for each fNIRS channel, broken down by channel (numbered at the top of each plot) and by task (the simpler task with three new objects or the more difficult task with five new objects). The overall mean of the null distribution of channels from scrambled caregiver-child pairs is shown as a blue dashed line.

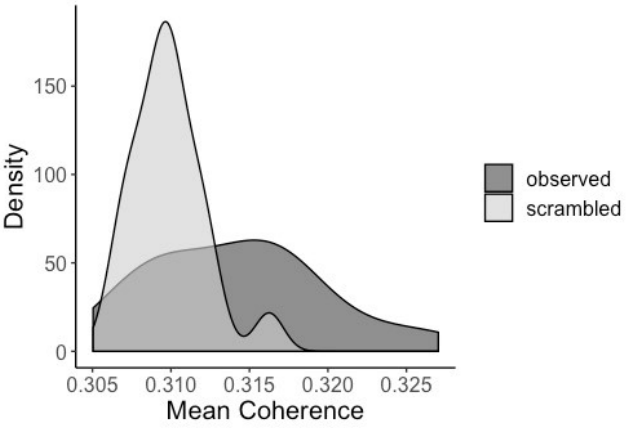

Figure 4.4. A density plot of the coherence levels, by channel, of scrambled caregiver-child pairs, overlaid with the distribution of the actual data. Each channel was averaged across participants, thus there are 20 values (the average value of each of 20 channels) within each distribution. The figure reveals that the distributions of these two data sets were different, with the null distribution having far less variability than the observed data, even though no statistical differences were found between channels.

Given that no channels were significantly higher in coherence than the null baseline, no specific channels could be correlated with the number of words learned. In other words, no specific area of the cortex was uniquely different for interacting caregivers and children, which would be needed to reflect task-relevant brain activity. Therefore, no theoretical motivation existed based directly on previous work, to explore relationships between word learning and coherence within this study.

4.4 Discussion

An emerging question within the hyperscanning literature is how neural coherence, measured throughout interactions, contributes to early word learning (Piazza et al., 2020; Wass et al., 2020). We aimed to determine whether neural coherence between the caregiver and child as caregivers named new objects would predict the child's learning. However, prior to being able to investigate the impact of neural coherence on learning, we failed to observe greater levels of neural coherence between interacting children and caregivers compared with that of caregivers and children who were engaged in the same task but who were not interacting. Thus, our findings do not lend support to the hypothesis that behavioural coordination directly drives alignment in underlying brain activity between communicating partners (Hasson et al., 2012; Piazza et al., 2020; Wass et al., 2020). Otherwise, what was happening during the interactions would increase the coherence in brain activity between caregivers and children who were interacting compared with those who were not. Nor does it support a hypothesis that neural coherence can reliably predict individual children's word-learning during interactions within local, task-relevant brain areas as has been found within other studies that used the same protocol.

One explanation is that neural coherence between interacting partners does increase during interactions, but that this is a subtle phenomenon compared with group-level differences in individual brain activity between conditions, the traditional type of functional imaging analysis. After conducting a Benjamini-Hochberg correction for multiple comparisons, Liu et al. (2019) found greater coherence between interacting partners within one channel in an array of 20 channels. In a within-subjects study, their sample size was 84 adult-adult dyads. This suggests that greater statistical power may be needed to find task-related neural coherence between interacting partners than is required to find task-related differences in individual brain activity. A future direction could be to explore neural coherence between caregivers and children with a larger sample size, likely requiring data collection paradigms that are more portable.

Without having access to a larger sample size that could provide the power to detect coherence that is specific to interacting pairs, there may still be ways to measure, validate and draw brain-behaviour relationships from neural coherence in a sample of 37 dyads. This would require addressing some limitations and/or exploring how coherence can predict learning in a within-subjects paradigm. To address a first limitation, we could quantitatively determine where to look for coherence rather than determining this through visual inspection. Liu et al. (2019) examined coherence between frequencies of 0.01-0.7 Hz during the task, and analysed data within a smaller frequency range that differed maximally from coherence during a resting period. To implement an analysis like this, a basic resting period such as having a caregiver and child sit facing forward and alone interacting with a toy, could provide a measure of resting coherence from which to find the best window in which to look for task-relevant coherence. Likewise, Hirsch et al. (2017) explored and reported significant coherence values over a subset of frequencies extracted during the analysis rather than *a priori*. The frequency range in which we visually observed the highest levels of coherence (6-30 seconds) overlaps with these studies (Liu et al. found coherence between 20 and 50 seconds with a shorter task; Hirsch et al. found coherence between 12 and 24 seconds). However, because we determined which frequencies we wanted to extract based on visual inspection, we could potentially have missed the range of maximal coherence within our task. Finally, some research suggests that overall behavioural outcomes during caregiver-child interactions are better predicted by coherence during earlier task blocks, compared with later ones (Reindl et al., 2018). Thus, a different consideration is that calculating neural coherence within the first half of the training interaction with three objects and with five objects, rather than averaging across both the first and second half, may prove more promising.

Finally, more care should be taken to reduce noise that contributes to neural coherence from background brain activity and from physiological influences on changes to HbO and HbR concentration. While periods of rest may be a problematic baseline for hyperscanning studies (Reindl et al., 2018), subtracting coherence during periods of rest from coherence during the task *prior* to validating with scrambled pairs may eliminate background noise from these comparisons (Liu et al., 2019). In developmental studies in particular, more precise methods are needed for establishing what aspects of coherence can be attributed to underlying similarity of brain activity at rest, and as a result of being engaged with a task. Alternatively, to reduce noise associated with physiology prior to comparing interacting dyads with scrambled pairs, it might also be necessary to subtract physiology-related components from the dyad-level data, prior to comparing interacting dyads with non-interacting controls (see Kirilina et al., 2013 for a review of a method to subtract frequency bands associated with physiology). While, to the best of our knowledge, this has not been previously done in a hyperscanning study, it could potentially reduce the amount of noise in the coherence values and increase the likelihood of finding a difference between interacting and scrambled caregiver-child dyads.

It is expected that neural coherence within a small proportion of corresponding channels in an array is significantly greater than coherence within scrambled participant pairs. Yet, within a handful of relevant studies we reviewed, this average was 6.5% of all channels/ROIs tested (Hirsch et al., 2017; Jiang et al., 2015; Liu et al., 2019; Reindl et al., 2018). Additionally, only two of these studies report using a correction for multiple comparisons in testing individual channels or ROIs against a scrambled baseline. Therefore, one possible explanation for many previous effects is that they were spurious: the result of conducting numerous statistical tests, and that with appropriate corrections, null results like ours would be more common. Even though previous studies that did not use these corrections may have had significant results, neural coherence across tasks and through time may still not be specific to interacting dyads.

Different sorts of brain-behaviour relationships could be explored to test a new hypothesis that relationships between coherence and word learning do not need to involve overall differences between actual and scrambled caregiver-child pairs, but still could be related to learning. For example, Jiang et al. (2012) found that neural coherence could be used to classify whether communicative behaviours such as turn taking in dialogue, were occurring during face-to-face or back-to-back interactions, even though they were not able to validate their coherence measure against scrambled pairs within each condition. Thus, it may be possible to statistically determine differences in intervals in which naming of learned words predominately occurred, and intervals in which naming of words not learned predominately occurred, based on the coherence values across these time intervals.

In other words, coherence may be different when the same children are learning versus when they are not.

Our findings did not suggest that a robust overall difference exists in coherence between caregivers and children who were interacting directly, versus those who were interacting with a different partner while doing the same task. Future work should be undertaken to boost confidence in these null results or to ultimately disprove them, including controlling noise from physiology or other sources in the unprocessed data prior to comparing coherence in actual caregiver-child dyads with those of scrambled dyads. However, new avenues for research are also opened, most importantly, an exploration of whether coherence can be used as an indicator that successful communication is taking place regardless of whether *overall* differences in coherence between actual and scrambled dyads are observed, or whether coherence can predict intervals of time in which a same child is learning, compared with those in which they are not learning.

SECTION 3

–

General Discussion and Conclusion

Chapter 5 – General Discussion and Conclusion

5.1 Chapter Overview

The present thesis investigated the neural correlates of interactive word learning between caregivers and children. In two interactive training blocks, caregivers taught their 2- to 5-year-old children the names of three and then five new objects while playing naturally. Following these interactions, children were presented with alternating pairs of familiar and new objects and were asked to select one by an experimenter. Though facing away from their child, caregivers saw the same object pairs and were separately asked to point to a picture of the correct referent. Hemodynamic response data to instances of naming were measured from both the caregiver and child during each of these tasks, and neural coherence across each naming interaction between the caregiver and child was calculated between corresponding channels on the caregiver's and child's head. Given that previous neuroimaging research had predominately tested children in computer-based word learning tasks that lacked a live, social-interactive context (Ekerdt et al., 2020; López-Barroso et al., 2013; Nora et al., 2017; Romeo, Leonard et al., 2018; Rice, Moraczewski et al., 2016; Takashima et al., 2019), our primary aim was to determine how brain activity predicted learning during an interaction. First, we investigated the neural correlates of interactive word learning between caregivers and children in response to naming, comparing hemodynamic response averages from naming of new words that were learned and that were not learned by the child. In order to connect differences in hemodynamic responses to behaviour that occurred during the interaction, we then explored whether spatial coordination between the caregiver and child across the session and during the naming events themselves could predict children's learning. Finally, we compared neural coherence between caregivers and children who were interacting during these tasks with that of mismatched dyads. In this chapter, the findings from these studies will be reviewed, along with limitations and methods considerations, and broader theoretical implications.

5.2. Summary of Main Study Results

5.2.1 Chapter 3

Within this chapter, children's and caregivers' hemodynamic responses to instances of naming were first compared for words the child learned, versus words the child did not learn. Neural responses to naming events differed, between words the child learned and did not learn, within both the caregiver's and the child's brain activity. We then analysed the behavioural data and found that the distinctiveness of object locations (which were measured by the volume of overlap) predicted which words the child learned. The present study did not find evidence that dyads intentionally kept objects in distinct positions in order to teach and learn, or even that activation within their neural responses was greater during learning. Taken together, these findings suggest that coordination between caregivers and children enables children to learn new words, and that it may be easier to learn when children and caregivers are processing in compatible ways.

Results revealed that children's neural responses differed between learned words and words not learned within the right inferior temporal cortex. Meta analytic work that used the Neurosynth database revealed that this area is functionally associated with social and language processing. This may indicate that interrelated language and social-pragmatic processing underlie differences between responding to naming of learned words and words not learned during interactions. Activation within the right supramarginal gyrus, found in response to instances of naming in general, was functionally associated with attention across a variety of different tasks. This may suggest that hearing new objects named elicits attention to what is being referred to. For example, this may occur when children process speech within complex environments (Adank et al., 2012). It is possible that the interaction observed suggests that, during learning with an array of multiple new objects, children inhibit their attention to distracting stimuli.

An analysis that examined the caregiver's neural responses to naming revealed an interaction between learning and chromophore within the left angular gyrus, functionally associated with determining social relevance and with language and category processing. Once again, the oxygenated hemoglobin (HbO) concentration showed positive activation in this analysis when naming words that were not learned. Further, differences were observed in the left prefrontal cortex based on the child's age, indicating that caregivers may have worked harder to process social information when teaching their younger children. Some work suggests that children construct their own learning experiences with the environmental resources available (for related work, see Trawick-Smith, Russell & Swaminathan, 2011; see also Gopnik & Wellman, 2012). In the present study, there is evidence that children's and caregivers' processing varied together, supporting the applicability of a two-person neuroscience framework to studying early word learning (Hasson et al., 2012; Wass et al., 2020).

We also tested whether one path to the child's learning could be measured via the spatial consistency of object locations and the volume of overlap between objects during naming interactions with the caregiver. We found that children's learning was positively predicted by the distinctiveness of object locations (which we assessed by the volume of overlap). This finding was consistent with previous literature that had found younger children's word learning was poorer when object locations were switched during naming (Benitez & Smith, 2012; Samuelson et al., 2011). This finding suggests that children struggle to learn more than one new name in a location that had been occupied by multiple objects. Taken together, results suggest that children's mapping of words to objects is influenced by a range of factors including environmental, behavioural and cognitive coordination with their caregiver.

5.2.2 Chapter 4

We additionally explored an outstanding question in the literature, whether neural coherence between caregivers and children during interactions in which the caregiver named objects could predict the child's word learning (Hasson et al., 2012; Piazza et al., 2020; Wass et al., 2020). Within this study, we used a wavelet transform coherence (WTC) analysis (Cui et al., 2012) to compare neural coherence within matched channels between interacting caregivers and children with a distribution of matched channels between mis-matched caregivers and children. After applying a correction for multiple comparisons, we did not find any channels that were significantly more coherent within interacting dyads, compared with dyads who were engaged in the same task but who were not interacting. This may be due to lack of power in our sample (for a significant finding in a related task with a larger sample, see Liu et al., 2019), or to the manner in which we chose to calculate coherence. Thus, one future direction is to explore differences in neural coherence within-subjects, between intervals of time in which more learned words were named, and intervals of time in which more words not learned were named. Further methodological considerations regarding the WTC analysis can be found in section 5.3.4.

5.3 Limitations

The study breaks new ground in terms of applying a naturalistic, neuroimaging design to study learning during caregiver-child interactions. However, multiple technical limitations and considerations with regard to the design and the interpretation of the data could be addressed in future work.

5.3.1 Technical Limitations Within the Behavioural Paradigm

For the scoring of the British Picture Vocabulary Scale (BPVS-3), original scores of children's vocabulary were bound to reflect age-driven differences, and moreover, we were not able to use the population-based standardization typically used to score this survey, because the assessment had not been standardized for our youngest age group. Standardization of the BPVS-3 begins at three years of age. As a solution, we standardized all of the scores for both age groups within the full sample of children recruited regardless of whether they had finished the study. When we calculated the survey-based standardized score of the older age group, we realized that the children in our sample had slightly higher vocabulary scores than expected though within one standard deviation. Thus, while they are valid to our sample, the *standardized* scores we used reflect the distribution of these children's vocabularies and may only marginally compare with the survey-based standardization.

Children's referent selections were sometimes messy during the comprehension test, including picking up more than one object and requiring the experimenter to prompt a second time or reaching for a favourite object rather than the object asked for by the experimenter. We created a margin of error in which children must not retrieve the incorrect object during familiar object trials more than six times during the session, because we could assume they knew the names of these objects and therefore their selections provided a basis for evaluating how on-task they were when selecting new referents. However, it is likely that some responses during the new object trials did not reflect what the child knew. Additionally, it is possible that in some cases, a child who knew the requested object name was on-task during one of the two times they were asked to select it, and that they were off-task on the other, leading us to assume they had not learned the word. This would mean that sometimes a word the child knew would be counted as a word not learned in the analysis, and sometimes a word the child didn't know would be counted as a learned word.

5.3.2 Technical Limitations Within the Machine Learning Paradigm

Within the machine learning analysis, Faster R-CNN relied to some extent on the stability of the surrounding environment in order to recognize objects whose locations varied frame-to-frame. This has been discussed in greater detail in section 2.2.4. A pseudo-naturalistic design offered greater stability than a fully real-world design due to the surrounding equipment and environment of the lab being consistent, and due to consistency in the overhead camera placement. However, variations such as participant clothing and objects brought into the environment by participants occasionally led to erroneous predictions by the network. For example, in a case where a participant was wearing a bright yellow shirt, their forearm may be mistaken in some instances for the yellow worm toy. In some cases, this created noise in the unprocessed data or resulted in missing data even when the objects were visible.

While, in theory, the adehabitathr package offered a conceptually valid method for measuring object locations in our study, it had not been previously used for a similar type of data (Calenge, 2006; Walter et al., 2015). Conceptually, a method designed to measure the volume of overlap in the home ranges of animals should also measure the volume of overlap between the areas most associated with each object. When fit to the object location data, the function was flexible enough to find multiple high-density locations when objects frequently occupied more than one location during the interaction. Presently, this package was a user-friendly and suitable tool for the research question, though previously untried in this context.

Most importantly, the CNN was only trained to detect object locations and classify the objects in the videos. Therefore, the study did not factor in when objects were being held or viewed by participants. We were able to start measuring each object after it had been held the first time by the caregiver or child because this aspect could be manageably determined by observer video coding. However, we could not directly make a connection between the child's attention to objects or their holding/their caregiver's holding of objects and the impact of overlap during these interactions. Previous work focused on embodied spatial coordination, such as measuring which hand predominantly held each object during caregiver-child interactions (Samuelson et al., 2011), but our study examined spatial coordination more indirectly than it had previously been tested. We assumed that object locations changed based on holding and that objects were predominately attended to during naming, but only actually measured the object locations following the first time each object was held (based on video coding).

5.3.3 Technical Limitations Within the fNIRS Analyses

The primary methods limitation in the event-based NIRS analysis was the variability in the number of times children named each object, (*M* = 4.45 x/object, *SD* = 3.41) Further, the number of possible children's naming events did not reliably afford comparable statistical power to that of adults' naming, (*M* = 12.61 x/object, *SD* = 4.76). One challenge of the present study is that even though it is within a neuroimaging paradigm that typically has conventions for spacing between events of interest, there is no way to control for when participants name objects, in the same way that the intervals of time between stimulus presentations in computer-based tasks can be controlled for as a part of the experimental design. The neuroimaging data had a fair amount of motion artefact and a low signal-to-noise ratio that was corrected in the event-based analysis via removing a substantial number of individual channels. Additionally, sparse channel arrangements and limited cortical coverage for each participant resulted in limited methods for controlling for physiology. In the present event-based analysis, we approached this with a global signal regression and via only reporting effects that showed a statistical interaction between chromophores.

Because the neural coherence analysis was conducted at an earlier time point and in the channel space, there were fewer channels removed due to signal noise and there was not a correction for physiological noise, as this was done during the image reconstruction of the event-based analysis. Finally, the analysis of neural coherence controlled for physiology post-hoc, by comparing actual and scrambled pairs (Reindl et al., 2018). In future, omitting periods in which physiology is known to occur prior to extracting coherence data may reduce the inherent noise from physiology and lead to a statistically significant difference between coherence within channels from actual and scrambled pairs. Kirilina et al. (2013) used a wavelet analysis on multiple channels recorded from the scalp to determine which frequencies contained physiological signals and discovered that physiological contamination in the data occurred within distinct periods of around 3, 10 and 34 seconds. Since the WTC analysis used in the present study (Grinsted et al., 2004) analyses neural coherence within a specified time-frequency space, these frequencies could simply be removed from the initial coherence calculations.

5.3.4 Hyperscanning Interpretations and Limitations

In a review, Hamilton (2021) argued that confidence in hyperscanning findings can be bolstered by a neurocognitive theory and better relationships drawn between neural coherence and behaviour. Such a theory could do more to explain why coherence is greater during active communication and cooperation than when individuals engage with the same task. This is critical because finding that brain activity is similar between individuals who complete the same task does not indicate anything that is not already assumed in studies that examine group-level responses to stimuli across individuals. Instead, hyperscanning should reveal more precise mechanisms above and beyond this principle. The hypothesis of *mutual prediction* (Kingsbury et al., 2019) posits that alignment in brain activity between interacting partners occurs because the individuals simultaneously control their own behaviour and predict their partner's behaviour. Studying mutual prediction in interactions requires consideration of the behaviours that occur during the interactions, and how these behaviours are related to the corresponding alignment in brain activity. This may occur at multiple levels, such making sensorimotor predictions about what a conversation partner will do, and semantic predictions about what they intend to say next (Stephens et al., 2010; Wass et al., 2020). Further, it may involve computations about establishing or maintaining joint attention (Piazza et al., 2020).

Mutual prediction, broadly applied, opens up a new avenue for interpreting neural synchrony during teaching and learning new words. However, few prior studies thoroughly examined the role of prediction even in their own findings. For example, in the study that revealed correlations between speakers and listeners (Stephens, Silbert & Hasson, 2010), the aim was to demonstrate that speech comprehension was correlated between a speaker and listeners when the speech was understood. Anticipatory activity, or signals that aligned with a listener-led lag, proved to be a robust indicator of the depth of speech comprehension. Predicting what someone would say was a robust marker of both speech comprehension and neural synchrony between speaker and listener. Liu et al. (2019) tested this concept experimentally by manipulating the syntactic consistency in the stimulus presentation as speakers and listeners alternated roles. Neural coherence was greater when sentence structures were consistent across switches. In the present study, coherence and its relationship to learning may have had either a speaker- or listener-led lag (indicating the importance of processing and prediction), and it could be that neural coherence is higher for interacting dyads with a lag like this, and that coherence at a lag may be correlated with learning. However, this was not a planned investigation of the present study. Instead, only synchrony that occurred time point to time point was analysed. Further exploratory work could systematically explore lags between caregivers and children (Piazza et al., 2020) in a word learning task. For example, a speaker-led lag may exist during times in which adults frequently name objects, such that the caregiver anticipates what the child will see and hear before naming the object, while the child processes the mapping of word to object after the fact. Conversely, a listener-led lag may occur if the child predicts that an object is likely to be named or needs to be named, while the adult may look to see whether the child made the mapping after naming the object.

In contrast to the way that mutual prediction has sometimes been articulated, when it comes to teaching and learning words, the primary aim is not always to understand how participants are reasoning about each other’s actions. As Wass et al. (2020) explain, in studies that examine language comprehension, often the aim is to explain how representations in a common task (like learning object names) align via achieving a shared understanding of semantic meaning. Interactions can establish common ground in higher level cognitive processes involved in comprehension, narrative and shared meaning (Fishburn et al., 2018; Simony et al., 2016; see Stolk et al., 2014). This appears to blur the line between coherence that occurs due to task structure (such as teaching and learning new words) and coherence that occurs through reasoning about the other person. However, in early language learning, much work suggests that mapping words to objects is influenced by actions that happen during interactions (Cartmill et al., 2013; Kalagher & Yu, 2006; Samuelson et al., 2011; Sullivan et al., 2015; Yu & Smith, 2013; Yu et al., 2009), as well as the content of dialogue as it relates to observed stimuli (Sobel & Corriveau, 2010). Actual and long-term understanding of word meanings are more often integrated across multiple contexts, some of which do not involve face-to-face interaction even though they involve ambient conversational input or artefact. The present scope of hyperscanning studies is to examine how varying factors, including dialogue, environmental organization and gesture that occur during interactions impact on word learning. These types of factors likely play the greatest role in the rapid word-object mappings made and tested within a single session.

Much prior hyperscanning work that involved interactions between participants has focused on cooperative tasks in which individuals have equivalent knowledge and must align their behaviours (Cui et al., 2012; Miller et al., 2019; Reindl et al., 2018). For word learning studies, another aspect is the need to consider differing roles and knowledge that is possessed by the teacher but not the learner. For example, if children do not know an object name, they technically can't predict what will happen in the speech stream. However, prior work suggests that children do understand when naming is likely to occur and may even use environmental information such as where an object was located before when it was named, to predict the naming event (Benitez & Smith, 2012). An important process to observe in cases where one person is teaching the other information is how behavioural coordination that occurs between the teacher, learner and new objects translates into both predictive processes (e.g., I know that this space is where that object was named before) as well as processes that resolve ambiguity following naming (Samuelson et al., 2011). The present work did find evidence that event-related responses to naming in particular are different during successful communication for both caregivers and their children. This shows that one general advantage of the hyperscanning approach is that it opens up the possibility of examining relationships in brain activity across time during an interaction.

One limitation that exists in the present hyperscanning work is the lack of considering the impact of interdependent processes occurring within interactions, especially in determining when successful learning was taking place and when it was not. A block analysis of coherence within corresponding channels, once baseline corrected, can reveal whether there are overall increases in neural coherence between interacting partners, but cannot directly link these to moment-to-moment behaviour or tell us whether similarity in brain activity may occur from non-corresponding pairings (within different areas in the brain). This would distinguish whether coherence was higher between caregivers and children at some points in time and lower at others, and whether these differences were related to children's learning. It may also indicate that the caregiver and child may need to be engaging in different processes that may be implemented within differing networks. Prior work suggests that what is happening during the interactions may drive coherence between adults and children moment-to-moment, and that this largely comes down to the effort each person is putting into considering what the other person is doing (Miller et al., 2019; Piazza et al., 2020; Reindl et al., 2018). Coherence can be examined instead in light of sustained behaviours that occur during the interactions, or within both corresponding and non-corresponding pairings. For example, a speech production area may align in the caregiver's brain with an area involved in speech comprehension for the child, or coherence may only occur during intervals of time in which the participants were focused on the same object or when learned words were being named. When the two ideas are combined (that coherence should be linked to relevant behaviour and that coherence could occur in differing but theoretically associated areas of the brain), then this could open up new avenues to explore how and when coherence is occurring at a mechanistic level, rather than assuming that coherence across entire interactions reflects uniformly distributed sharing of meaning and task aims.

An additional consideration is where experimentally relevant coherence should be looked for within the vast amounts of underlying and task-irrelevant brain activity. Studies that have used highly controlled stimuli, have chosen a task-relevant frequency band (such as, corresponding to a period length of approximately 3-13 seconds, Cui et al., 2012) based on the durations of individual cycles of a stimulus presentation and the expected range of participants' response times. Other studies have determined a window in which to look at coherence post-hoc and based on where the greatest differences were found from a resting or non-interactive variant of the task (Hirsch et al., 2017; Liu et al., 2019). For naturalistic interactions, it may not be possible to say exactly at what frequencies coherence that is relevant to the research question can be found, but the problem can be constrained by excluding frequencies that are driven by systemic physiology. As mentioned previously, Kirilina et al. (2013) identified a frequency band within a WTC analysis of individual participants at roughly three seconds corresponding to respiration, and another at roughly ten seconds corresponding to Mayer waves. They further found that removing these frequencies from the intra-brain coherence analysis improved sensitivity to task-relevant brain activity. Future studies could look for optimal coherence within a range of frequencies that may plausibly correspond to the cognitive processing needed to map words with objects. One way to use null baselines, such as scrambled caregiver-child pairs, is to quantitatively identify the frequencies at which coherence is higher specifically for caregivers and children who were interacting across a broad range of possible combinations, and then to separate and analyse those frequencies.

Putting these considerations together, the following constraints in interpreting hyperscanning results can be considered along with the specific limitations of the present study. When possible, hyperscanning should be grounded in behaviour in order to understand why neural coherence occurred. This could be examined via considering the relationship between neural coherence and looking, holding toys, or spatial organization, for examples. Neural synchrony in general should be looked at not just as simultaneous alignment in brain activity but taking into account the possibility that synchrony has a listener- or speaker-led lag due to the importance of mutual prediction in explaining why two people who are in different roles may have similar brain activity. Finally, the frequencies at which neural coherence is analysed are best determined via a combination of theoretical and quantitative exploration that identify those frequencies that mattered for interacting partners. The present study, which compared coherence between actual and scrambled caregiver-child pairs, made a step toward controlling for everything else that is unimportant to what drives genuine experimental results. However, it is limited in terms of the range of methods used to do this out of those that are possible. Additionally, as the above discussion shows, the field is currently grappling with ways to resolve ambiguity in understanding what hyperscanning findings may mean.

5.3.5 Further Design Considerations

The present study investigated caregiver-child interactions within a pseudo-naturalistic, openly structured laboratory task. There were aspects of this design that may be too naturalistic for the level of control expected of most laboratory research, and simultaneously, aspects that may not have been naturalistic enough to ensure full ecological validity. In terms of controls, the main area in which this project was limited was the inability to isolate specific variables from the context of the interaction. For example, brain activity between learned words and words not learned in our task could be influenced by variables such as dialogue and visual joint attention (i.e., viewing the same object) between the caregiver and child, which were not directly measured. Additionally, we were not able to measure the child's level of verbal attention and processing during the task, and whether relevant activation contributed to the differences found between learned words and words not learned. Collectively, these limitations suggest that the study represents a start, but not a complete path, to mapping the neural and behavioural factors that underlie learning in the present task.

In terms of ecological validity, the current design is in some ways artificial compared with learning words in the real world. For example, in our tasks, caregivers were essentially learning the new object names along with their children, thus, they did not have complex experiences from which to draw when teaching them. Additionally, unlike learning in the real world, caregivers likely felt pressured and a great deal more focused on helping their child make word-object mappings based on the fact that their child would immediately be tested on their learning. Thus, some caregivers spent time during the interactions in less natural activities, such as repeatedly testing their child on the object names. Finally, the present design could not capture the full range of variables that are relevant for learning at home. For example, children may frequently map names to objects during interactions where the caregiver multi-tasks or shifts attention between siblings, in interactions with teachers and peers, and in a number of different environments. Therefore, the present study can be viewed as a start toward investigating the neural correlates of interactive word learning within naturalistic environments.

5.4 Theoretical Implications

The present work investigated the behavioural and neural correlates of interactive word-learning between caregivers and children. Within computer-based tasks, previous work had investigated neural correlates in processing of new words (Takashima et al., 2019), concrete nouns (Berl et al., 2014), and structural changes as a result of word learning (Ekerdt et al., 2020), for examples. Collectively, previous research suggested that neural correlates within language-specific and other cognitive neural networks, such as those that implement attention and social processing, may facilitate children's word-learning. In these tasks, cognition required to support word learning includes attending to word-object pairings being presented repeatedly, and also working memory for which words have previously been paired with which objects. Some work specifically suggests that 4-year-old children, among the youngest previously tested in word learning neuroimaging studies, must develop attention in order to meet the demands of intensive word learning training. This was true even to the point where structural changes to the networks that implement attention became the primary result of the word learning training even if the preliminary structural measures of areas that implement language predicted children’s performance (Ekerdt et al., 2020). What this may suggest is that, even in highly controlled laboratory studies, children are ever developing the fundamental cognitive building blocks that support focusing on relevant linguistic stimuli in the world. The present study investigated the neural correlates of word-learning while caregivers taught their children new object names, a naturalistic task in which brain activity associated with word learning had not been previously analysed.

Learning words during a live interaction only increases the demands on perception, attention and working memory because of the need to process naming events in conversation, to resolve ambiguity in naming within dynamic multi-object arrays, and to maintain consistent word-object mappings while objects are moving around. These types of demands, as well as the ways that social interaction facilitates children’s ability to map words to objects, have been extensively investigated at a behavioural level. Collectively, this research shows that coordinated attention and working memory between the adult and child during interactions in which new objects are named, enhances the child’s learning (Cartmill et al., 2013; Kalagher & Yu, 2006; Yu & Smith, 2013). The present neuroimaging findings may shed some light on how and why this occurs in terms of the neural responses to naming. The present event-related findings for children, revealed that changes in attention-related responding occurred when new objects were named in conversation. This is evidenced by the main effect of attention in response to naming and the learning-invariant activation during play. By contrast, there are multiple possible interpretations as to what drove differential responding between learned words and words not learned based on the comparison with prior research. The area that responded more robustly to naming of words not learned, when compared with prior literature, was functionally associated with language processing, and was also robustly associated with social processing (see section 3.3.2). This is generally consistent with prior work that found that areas of the frontal cortex are engaged selectively in semantically challenging tasks (Gabrielli, Poldrack & Desmond, 1998). It isn't possible to know exactly which processes drove the differences between hemodynamic responses to learned words and words not learned based on the neural activation alone in the present study, but one possible interpretation is that when children were learning, something was easier about the social and language processing demands around instances of naming.

This project was the first to test whether caregivers' brain activity differs when successfully, as opposed to unsuccessfully, teaching their children. Though it is difficult to interpret the interaction observed, it may be that even for caregivers, processing is more effortful during times when learning is not taking place. If so, this runs counterintuitive to the common assumption that higher activation levels (as indicated by increased HbO) are associated with successful outcomes.

The findings reveal that within-subjects differences can be detected in the neural encoding of naming events during caregiver-child interactions that reflect learning. Importantly, the present findings also suggest that learning, and the underlying neural responses to naming events that occurred during these interactions, likely relies on differences on the attention demands for social and semantic processing. This combination may reflect the challenges children faced in keeping track of which words went with which referents, given the complexity of the task. For example, when a new name occurred within a speech stream, the child needed to be able to process not only the word-object pairing, but to hone their attention in the moment toward aspects of their caregiver's behaviour that would help them determine the intended referent. The present findings should be interpreted with some caution, as the ultimate standard of measuring whether a new word is processed in a semantically meaningful way, should be found within the language processing network. We did not find evidence, as predicted, that differences between learned words and words not learned could be detected in the child's language network, only that an area that is engaged during language and semantic processing in language tasks within the right temporal cortex was activated differently when children were learning as opposed to not learning from naming events. As discussed prior, this could be for a number of reasons. Barring one study (Ekerdt et al., 2020) there is not enough information from prior fMRI and fNIRS studies to fully map the neural correlates of word learning for children in the preschool years, and, to the best of the author's knowledge, none at all when thinking about openly structured interactions.

Typically, the child’s behaviour and brain activity have been considered as an isolated unit within neuroimaging paradigms of noun processing and of word learning (Berl et al., 2014; Ekerdt et al., 2020; Nora et al., 2017; Rice, Moraczewski et al., 2016; Takashima et al., 2019). The present study investigated the viability of an alternative hypothesis, that the caregiver’s and child brain activity are coordinated when the child is encoding new object names. Broader theoretical support for this hypothesis can be found in social network models of human cognition, which suggest that the behaviour and brain activity of individuals are fundamentally constrained by those of others during an interaction (Hasson et al., 2012). According to the authors, this occurs as a result of non-verbal social information that aligns attention between interacting individuals as well as aligning of their representations through dialogue. Social network models of cognition have been applied to predict speech comprehension (Stephens et al., 2010), word-picture matching (Liu et al., 2019) and have been hypothesized to set the conditions between an adult speaker and child for the child’s language learning (Hasson et al., 2012; Piazza et al., 2020; Wass et al., 2020).

The present findings from examining event-related responses to naming did reveal differences in the neural signatures when caregivers named words their child would learn, compared with words their child would not learn. The findings also revealed differences in the neural responses to naming events when caregivers were teaching 32-month-old, versus 54-month-old, children. While there is evidence suggesting that children draw from multiple processes to make word-object mappings in complex social environments, there is also a possible relationship between the caregiver’s explicit teaching and processing, and whether words were learned or not learned. This is in contrast to Chapter 4, where no evidence was found that interacting caregivers’ and children’s brain activity were temporally correlated, through time, during the word learning interactions. This would normally be predicted within a social network model (Hasson et al., 2012; Stephens et al., 2010; Wass et al., 2020). Prior work has shown that children actively construct their own learning experiences from what is present in the environment (see Gopnik & Wellman, 2012). These include the objects present (Trawick-Smith et al., 2011), socially demonstrated information (Bonawitz et al., 2011; Göksun, Hirsch-Pasek & Golinkoff, 2010; Sobel & Corriveau, 2010; see also Sobel & Kirkham, 2012), and spatial organization (Axelsson et al., 2016; Benitez & Smith, 2012; Samuelson et al., 2011; see also Samuelson et al., 2017). The present results suggest that caregivers may also process what is occurring in the interaction differently, and maybe with less effort, when successfully teaching their children.

In further support of the hypothesis that it's "easier" to keep track of word-object mappings when conversation partners are processing in alignment, the behavioural results suggest that the volume of overlap in objects' locations interfered with children's ability to make and retain specific mappings. Even though the original prediction specified that spatial consistency would be correlated with children's overall performance in the comprehension test, the findings suggested that (a) there were not typically consistent spatial trends in the object locations, across the full duration of the interactions and (b) of the spatial measures examined, only volume of overlap between objects predicted the child's likelihood of learning specific words. These data indicate that reliable behavioural predictors within the caregiver-child coordination can be observed when the child is learning, that appear to break down when the same child is not engaged in the task toward learning. Specifically, children were less likely to learn object names when their locations overlapped more with other objects, aggregated through time. This finding suggests that coordinated behaviour between the caregiver and child may facilitate social and semantic processing. Together, these findings open up a number of potential future directions that could fill the gap between the neural findings, and their grounding in behaviour.

5.4.1 Lay Summary of Theoretical Implications

Previous research had shown that coordination between adults who are teaching and children who are learning facilitates children's language-learning. This may happen via pointing at, talking about and holding toys as they are named (Cartmill et al., 2013; Kalagher & Yu, 2006; Yu & Smith, 2013). Further, the consistency of object locations during interactions between caregivers and children was correlated with how many correct responses children made in a comprehension test (Samuelson et al., 2011). However, no prior work had investigated the neural correlates of word learning during live interactions or how space affected learning during caregiver-child interactions in the preschool years. In terms of brain activity, previous research had only explored the neural correlates of early word learning within tasks in which stimuli were planned and presented to the child on a computer monitor (see Balsamo, Xu & Gaillard, 2006; Berl et al., 2014; Ekerdt et al., 2020; Hirotani et al., 2009; Nora et al., 2017; Romeo, Segaran et al., 2018; Takashima et al., 2019). However, some research with adults had found that language comprehension could also be related to the similarity of brain activity between a speaker and listener, not merely the brain activity of the individual (Liu et al., 2019; Stephens et al., 2010). This could potentially mean that coordinated brain activity between caregivers and children, not just coordination in behaviour, could make teaching and learning more successful. The present study investigated whether there were relationships in behaviour and brain activity between the caregiver and child that could predict whether children learned words and how many words children would learn, as well as how well caregivers and children coordinated the object locations within their shared space during the preschool years.

All of the research questions addressed within this project were investigated within one study in which caregivers taught their 32- and 54-month-old children the names of eight new objects while engaged in unstructured play. Children's and caregivers' behaviour and brain activity were analysed separately between words the child later showed evidence of having learned and words they did not learn during these interactions. From this data, a study of object locations on the table, measured by a machine learning algorithm, revealed that children were less likely to learn a new toy name if that toy frequently overlapped with others. This indicates that the distinctiveness of toy locations mattered for children's learning, informing theories of exactly why spatial predictability matters for early learning (Samuelson et al., 2011; see Samuelson, Kucker & Spencer, 2017).

The present study revealed that children's neural responses are unique during learning. These findings may indicate that children find it easier to keep track of word meanings when they are learning from their caregiver's naming of new toys. Similarly, caregivers process differently when their child was learning from their own naming of the objects. There was a specific difference found in the caregiver's brain activity between naming events of words their child learned and did not learn within the left angular gyrus. By contrast, we did not find continuous relationships through time between caregivers' and children's brain activity throughout the interactions, as had been previously observed in similar studies (Liu et al., 2019; Piazza et al., 2020; Reindl et al., 2018; Stephens et al., 2010). Our findings generally support the theory that learning may be less demanding in terms of the language and social processing that occurs during naming events than not learning, and that this is true for both the caregiver and child. These different types of processes may set the conditions for early learning. Critically, the fNIRS and behavioural data that measured the distinctiveness of object locations during caregiver-child play was associated with differences between learned words and words not learned but did not consistently predict differences in amounts of learning. This may indicate that these measures are better at distinguishing differences between information that results in learning and information that does not, even during the same interaction.

5.5 Future Directions

One compelling future direction from the present work is to explore relationships between brain and behaviour in greater depth. Future research could examine a wider range of behaviours and introduce new analyses to further examine the relationships between these behaviours and the caregiver's and child's brain activity. The current work didn't examine where participants were looking or how they were interacting with the objects. One could assess whether the caregiver and child were more likely to be viewing the same object when caregivers named it for word-object pairings the child later showed evidence of having learned. If a caregiver and child were exploring the same aspects of the environment at the same time, this may also facilitate communication about the objects for various reasons (see Garrod & Pickering, 2009; Tomasello & Farrar, 1986). It may be that the percentage of naming events in which the caregiver and child were jointly viewing the same object was higher for objects the child correctly selected in the comprehension test. This would indicate that the caregiver and child were exploring the same aspects of the environment at the same time, and, further, that shared experiences drove the differences in neural activation to new naming observed in the child's and caregiver's fNIRS data. Other ways that joint attention could be observed between caregivers and children could include the percentage of cases in which the caregiver named an object as the child viewed it or held the object as the child viewed it/vice versa.

Another factor that could be explored in future research was the child's processing of the dialogue during the interactions and how this may have contributed to their heightened attention to new object names. While children's naming of objects during the interactions did not reliably afford enough power to be included in an event-based fNIRS analysis, further analyses of children's naming could establish the extent to which caregivers and children sustained dialogue around the same object as another indicator of sharing of experience. One limitation to this approach, however, is that children often mimicked their caregiver's naming of objects without explicitly demonstrating semantic encoding, for example, by drawing their caregiver's attention to an object or to embedding the object name in a sentence.

Finally, future research could explore the content of caregiver-child dialogues as an indicator of the child's verbal processing during the interactions. Qualitative exploration of the video recordings suggested that children varied in how attuned they were, not only to the caregiver's holding and exploration of objects, but to the content of the caregiver's speech. Assessing the content of caregiver-child dialogue during the interactions along with quantitative metrics such as rates of turn-taking in conversation with similar keywords, may reveal how deeply children processed their caregiver's speech. Further, assessing the content of dialogue could reveal when caregivers actively helped their children form word-object associations via conversing with them about the objects (Han & Neuharth-Pritchett, 2014). It would also assess when caregivers taught their child by testing them, asking them to name object-by-object and providing feedback to their responses. These variables could then be used to predict task-relevant brain activity during the interactions. For example, turn-taking in dialogue between caregiver and child during the interactions may predict brain activity associated with learning, whereas testing the child repeatedly on the object names may not. Beyond the present data and the variables that could be further explored within it, principles from this study could be applied to a broader range of contexts in order to better understand how, for examples, culture, neurodiversity and an increasing dependence on technology shape social influences on word learning (Wei, Leech & Rowe, 2020). Neuroimaging measures could assess how apps provide the social resources needed for children to learn and whether these resources tap similar mechanisms in the child's brain to those observed during real-world interactions with an adult speaker.

5.6 General Conclusion

The present thesis determined that coordination between caregivers and 2- to 5-year-old children during play was associated with which words the child later showed evidence of having learned. Results from an event-based functional near infrared spectroscopy (fNIRS) analysis revealed that, within brain regions associated with both semantic processing and social processing, children and caregivers responded differently to naming of objects children learned (learned words) and those they did not (words not learned). Additional behavioural measures of spatial coordination between the caregiver and child were explored, revealing that the distinctiveness of object locations during the interactions predicted which words children learned. Neither of these sets of findings provide robust evidence that children's learning came down to caregivers' active effort. Rather, the best present evidence suggests that sometimes it was easier for children and caregivers to remain in alignment when naming the objects. However, this conclusion is subject to the methodological limitations of the present study as well as the variables that could feasibly be measured.

The most robust markers of learning were observed within-subjects. While spatial coordination could not predict children's overall amount of learning, distinctiveness of object locations positively predicted which specific words children would learn. Likewise, while no overall cross-correlations were observed between caregivers' and children's brain activity when measured continuously through time, differences were found between the child's and caregiver's neural responses to naming of learned words and words not learned. The current results reveal that the processes children use to learn new words during interactions are similar to those observed in prior research that used computer-based stimuli. Potentially, both language and social processing are especially vital to learning words during openly structured interaction and in dynamic environments, and that the caregiver's neural responses to naming, not merely the child's, can indicate whether learning is taking place. This approach opens avenues for future work that could explore how social networks directly contribute to learning early in development.

References

Aasted, C. M., Yücel, M. A., Cooper, R. J., Dubb, J., Tsuzuki, D., Becerra, L., Petkov, M. P., Borsook, D., Dan, I. & Boas, D. A. (2015). Anatomical guidance for functional near-infrared spectroscopy: AtlasViewer tutorial. *Neurophotonics*, *2*(2), 020801:1-16. https://doi.org/10.1117/1.NPh.2.2.020801

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., & Zheng, X. (2016). TensorFlow: A System for Large-Scale Machine Learning, 265–283. Proceedings of the *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16).* Retrieved from: https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

Abtahi, M., Cay, G., Saikia, M. J., Mankodiya, K. (2016, August). Designing and testing a wearable, wireless fNIRS patch. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 6298-6301). https://doi.org/10.1109/EMBC.2016.7592168

Adank, P., Davis, M. H., & Hagoort, P. (2012). Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia*, *50*(1), 77–84. https://doi.org/10.1016/j.neuropsychologia.2011.10.024

Alloway, T. P., Gathercole, S. E. & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, *77*(6), 1698-1716. https://doi.org/10.1111/j.1467-8624.2006.00968.x

Archibald, L. M., & Joanisse, M. F. (2013). Domain-specific and domain-general constraints on word and sequence learning. *Memory & Cognition*, *41*(2), 268-280. https://doi.org/10.3758/s13421-012-0259-4

Arsalidou, M., Barbeau, E. J., Bayless, S. J., & Taylor, M. J. (2010). Brain responses differ to faces of mothers and fathers. *Brain and Cognition*, *74*(1), 47-51. https://doi.org/10.1016/j.bandc.2010.06.003

Avons, S. E., Wragg, C. A., Cupples, W. L. & Lovegrove, W. J. (1998). Measures of phonological short-term memory and their relationship to vocabulary development. *Applied Psycholinguistics*, *19*(4), 583-601. https://doi.org/10.1017/S0142716400010377

Axelsson, E. L., Perry, L. K., Scott, E. J. & Horst, J. S. (2016). Near or far: The effect of spatial distance and vocabulary knowledge on word learning. *Acta Psychologica*, *163*, 81-87. https://doi.org/10.1016/j.actpsy.2015.11.006

Balsamo, L. M., Xu, B. & Gaillard, W. D. (2006). Language lateralization and the role of the fusiform gyrus in semantic processing in young children. *NeuroImage*, *31*(3), 1306-1314. https://doi.org/10.1016/j.neuroimage.2006.01.027

Bankó, É. M., Gál, V., Körtvélyes, J., Kovács, G., & Vidnyánszky, Z. (2011). Dissociating the effect of noise on sensory processing and overall decision difficulty. *Journal of Neuroscience*, *31*(7), 2663-2674. https://doi.org/10.1523/JNEUROSCI.2725-10.2011

Beier, J. S. & Spelke, E. S. (2012). Infants' developing understanding of social gaze. *Child Development*, *83*(2), 486–496. https://doi.org/10.1111/j.1467-8624.2011.01702.x

Belsky, J., & De Haan, M. (2011). Annual research review: Parenting and children's brain development: The end of the beginning. *Journal of Child Psychology and Psychiatry*, *52*(4), 409-428. https://doi.org/10.1111/j.1469-7610.2010.02281.x

Benitez, V. L. & Smith, L. B. (2012). Predictable locations aid early object name learning. *Cognition*, *125*(3), 339–352. https://doi.org/10.1016/j.cognition.2012.08.006

Berl, M. M., Mayo, J., Parks, E. N., Rosenberger, L. R., VanMeter, J., Ratner, N. B., Vaidya, C. J. & Gaillard, W. D. (2014). Regional differences in the developmental trajectory of lateralization of the language network. *Human Brain Mapping*, *35*(1), 270–284. https://doi.org/10.1002/hbm.22179

Bhat, A., Spencer, J. P. & Samuelson, L. K. (2018). A dynamic neural field model of memory, attention and cross-situational word learning. In *CogSci*. (pp. 142-147). Retrieved from: https://cogsci.mindmodeling.org/2018/papers/0048/0048.pdf

Blackwood, N. J., Bentall, R. P., Simmons, A., Murray, R. M., & Howard, R. J. (2003). Self-responsibility and the self-serving bias: an fMRI investigation of causal attributions. *NeuroImage*, *20*(2), 1076-1085. https://doi.org/10.1016/S1053-8119(03)00331-8

Blondin, F., & Lepage, M. (2005). Decrease and increase in brain activity during visual perceptual priming: An fMRI study on similar but perceptually different complex visual scenes. *Neuropsychologia*, *43*(13), 1887-1900. https://doi.org/10.1016/j.neuropsychologia.2005.03.021

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E. & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322-330. https://doi.org/10.1016/j.cognition.2010.10.001

Braunlich, K., Gomez-Lavin, J., & Seger, C. A. (2015). Frontoparietal networks involved in categorization and item working memory. *NeuroImage*, *107*, 146-162. https://doi.org/10.1016/j.neuroimage.2014.11.051

Brown, M. R., Lebel, R. M., Dolcos, F., Wilman, A. H., Silverstone, P. H., Pazderka, H., Fujiwara, E., Wild. C. T., Carroll, A. M., Hodlevskyy, O. Zedkova, L., Zwaigenbaum, L., Thompson., A. H., Greenshaw, A. J. & Dursun, S. M. (2012). Effects of emotional context on impulse control. *NeuroImage*, *63*(1), 434-446. https://doi.org/10.1016/j.neuroimage.2012.06.056

Burgess, A. P. (2013). On the interpretation of synchronization in EEG hyperscanning studies: A cautionary note. *Frontiers in Human Neuroscience*, *7*(881), 1-17. https://doi.org/10.3389/fnhum.2013.00881

Busch, N.A., J. Dubois, & R. VanRullen (2009). The phase of ongoing EEG oscillations predicts visual perception. *Journal of Neuroscience*, *29*(24), 7869-7876. https://doi.org/10.1523/JNEUROSCI.0113-09.2009

Buss, A. T., Fox, N., Boas, D. A. & Spencer, J. P. (2014). Probing the early development of visual working memory capacity with functional near-infrared spectroscopy. *NeuroImage*, *85*(1), 314-325. https://doi.org/10.1016/j.neuroimage.2013.05.034

Cabell, S. Q., Justice, L. M., McGinty, A. S., DeCoster, J. & Forston, L. D. (2015). Teacher–child conversations in preschool classrooms: Contributions to children's vocabulary development. *Early Childhood Research Quarterly*, *30*(A), 80-92. https://doi.org/10.1016/j.ecresq.2014.09.004

Calenge, C. (2006). The package “adehabitat” for the R software: A tool for the analysis of space and habitat use by animals. *Ecological Modelling*, *197*(3–4), 516–519. https://doi.org/10.1016/j.ecolmodel.2006.03.017

Campagne, A., Fradcourt, B., Pichat, C., Baciu, M., Kauffmann, L., & Peyrin, C. (2016). Cerebral correlates of emotional and action appraisals during visual processing of emotional scenes depending on spatial frequency: A pilot study. *PloS One*, *11*(1), e0144393. https://doi.org/10.1371/journal.pone.0144393

Canessa, N., Gorini, A., Cappa, S. F., Piattelli-Palmarini, M., Danna, M., Fazio, F., & Perani, D. (2005). The effect of social content on deductive reasoning: an fMRI study. *Human Brain Mapping*, *26*(1), 30-43. https://doi.org/10.1002/hbm.20114

Caplan, J. B., Luks, T. L., Simpson, G. V., Glaholt, M., & McIntosh, A. R. (2006). Parallel networks operating across attentional deployment and motion processing: a multi-seed partial least squares fMRI study. *NeuroImage*, *29*(4), 1192-1202. https://doi.org/10.1016/j.neuroimage.2005.09.010

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G. & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), i-174. Retrieved from: https://www.jstor.org/stable/1166214

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N. & Trueswell, J. C. (2013). Quality of early caregiver input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

Chen, Y., Li, W., Sakaridis, C., Dai, D. & Van Gool, L. (2018). Domain adaptive Faster R-CNN for object detection in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3339-3348). Retrieved from: https://openaccess.thecvf.com/content_cvpr_2018/html/Chen_Domain_Adaptive_Faster_CVPR_2018_paper.html

Chou, T. L., Booth, J. R., Burman, D. D., Bitan, T., Bigio, J. D., Lu, D. & Cone, N. E. (2006). Developmental changes in the neural correlates of semantic processing. *NeuroImage*, *29*(4), 1141-1149. https://doi.org/10.1016/j.neuroimage.2005.09.064

Clerkin, E. M., Hart, E., Rehg, J. M., Chen, Y. & Smith, L. B. (2016). How everyday visual experience prepares the way for learning object names. In Joint IEEE International Conference on Development and Learning and Epigenetic Robots, (pp. 126-131). https://doi.org/10.1109/DEVLRN.2016.7846803

Coaster, M., Rogers, B. P., Jones, O. D., Viscusi, W. K., Merkle, K. L., Zald, D. H., & Gore, J. C. (2011). Variables influencing the neural correlates of perceived risk of physical harm. *Cognitive, Affective, & Behavioral Neuroscience*, *11*(4), 494-507. https://doi.org/10.3758/s13415-011-0047-9

Colletta, J. M., Pellenq, C & Guidetti, M. (2010). Age-related changes in co-speech gesture and narrative: Evidence from French children and adults. *Speech Communication, 52*(6), 565-576. https://doi.org/10.1016/j.specom.2010.02.009

Contreras, J. M., Schirmer, J., Banaji, M. R., & Mitchell, J. P. (2013). Common brain regions with distinct patterns of neural responses during mentalizing about groups and individuals. *Journal of Cognitive Neuroscience*, *25*(9), 1406-1417. https://doi.org/10.1162/jocn_a_00403

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., Linenweber, M. R., Peterson, S. E. & Shulman, G. L. (1998). A common network of functional areas for attention and eye movements. *Neuron*, *21*(4), 761-773. https://doi.org/10.1016/S0896-6273(00)80593-0

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C. & Taylor, P. A. (2017). fMRI clustering and false-positive rates. *Proceedings of the National Academy of Sciences*, *114*(17), E3370–E3371. https://doi.org/10.1073/pnas.1614961114

Cui, X., Bryant, D. M. & Reiss, A. L. (2012). NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *NeuroImage*, *59*(3), 2430–2437. https://doi.org/10.1016/j.neuroimage.2011.09.003

Davis, M. H., Di Betta, A. M., Macdonald, M. J. & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, *21*(4), 803-820. https://doi.org/10.1162/jocn.2009.21059

Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: an fMRI investigation. *NeuroImage*, *23*(2), 744-751. https://doi.org/10.1016/j.neuroimage.2004.05.025

Dobbins, I. G., Foley, H., Schacter, D. L., & Wagner, A. D. (2002). Executive control during episodic retrieval: multiple prefrontal processes subserve source memory. *Neuron*, *35*(5), 989-996. https://doi.org/10.1016/S0896-6273(02)00858-9

Dommer, L., Jäger, N., Scholkmann, F., Wolf, M. & Holper, L. (2012). Between-brain coherence during joint n-back task performance: A two-person functional near-infrared spectroscopy study. *Behavioural Brain Research*, *234*(2), 212-222. https://doi.org/10.1016/j.bbr.2012.06.024

Donaldson, D. I., Petersen, S. E. & Buckner, R. L. (2001). Dissociating memory retrieval processes using fMRI: Evidence that priming does not support recognition memory. *Neuron*, *31*(6), 1047-1059. https://doi.org/10.1016/S0896-6273(01)00429-9

Doulgerakis, M., Eggebrecht, A. T. & Dehghani, H. (2019). High-density functional diffuse optical tomography based on frequency-domain measurements improves image quality and spatial resolution. *Neurophotonics*, *6*(3), 35007:1-14. https://doi.org/10.1117/1.NPh.6.3.035007

Dunn, L. M & Dunn, D. M. (2009). *The British picture vocabulary scale.* GL Assessment Limited.

Egorova, N., Shtyrov Y. & Pulvermüller, F. (2016). Brain basis of communicative actions in language. *NeuroImage, 125*, 857-867. https://doi.org/10.1016/j.neuroimage.2015.10.055

Ekerdt, C. E. M., Kühn, C., Anwander, A., Brauer, J. & Friederici, A. D. (2020). Word learning reveals white matter plasticity in preschool children. *Brain Structure and Function*, *225*(2), 607–619. https://doi.org/10.1007/s00429-020-02024-7

Engell, A. D., & McCarthy, G. (2013). Probabilistic atlases for face and biological motion perception: an analysis of their reliability and overlap. *NeuroImage*, *74*, 140-151. https://doi.org/10.1016/j.neuroimage.2013.02.025

Etzel, J. A., Valchev, N., Gazzola, V., & Keysers, C. (2016). Is brain activity during action observation modulated by the perceived fairness of the actor? *PLoS One*, *11*(1), e0145350. https://doi.org/10.1371/journal.pone.0145350

Fang, Q., & Boas, D. A. (2009). Monte Carlo simulation of photon migration in 3D turbid media accelerated by graphics processing units. *Optics express*, *17*(22), 20178-20190. https://doi.org/10.1364/OE.17.020178

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in Psychology*, *5*(335), 1-17. https://doi.org/10.3389/fpsyg.2014.00335

Ferrari, M. & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage*, *63*(2), 921-935. https://doi.org/10.1016/j.neuroimage.2012.03.049

Ferreira, R. A., Göbel, S. M., Hymers, M. & Ellis, A. W. (2015). The neural correlates of semantic richness: Evidence from an fMRI study of word learning. *Brain and Language*, *143*, 69–80. https://doi.org/10.1016/j.bandl.2015.02.005

Fiddick, L., Spampinato, M. V., & Grafman, J. (2005). Social contracts and precautions activate different neurological systems: An fMRI investigation of deontic reasoning. *NeuroImage*, *28*(4), 778-786. https://doi.org/10.1016/j.neuroimage.2005.05.033

Fiebach, C. J. & Friederici, A. D. (2004). Processing concrete words: fMRI evidence against a specific right-hemisphere involvement. *Neuropsychologia*, *42*(1), 62-70. https://doi.org/10.1016/S0028-3932(03)00145-3

Fillmore, P. T., Richards, J. E., Phillips-Meek, M. C., Cryer, A. & Stevens, M. (2015). Stereotaxic magnetic resonance imaging brain atlases for infants from 3 to 12 months. *Developmental Neuroscience*, *37*(6), 515-532. https://doi.org/10.1159/000438749

Fishburn, F. A., Murty, V. P., Hlutkowsky, C. O., MacGillivray, C. E., Bemis, L. M., Murphy, M. E., Huppert, T. J. & Perlman, S. B. (2018). Putting our heads together: Interpersonal neural synchronization as a biological mechanism for shared intentionality. *Social Cognitive and Affective Neuroscience, 13*(8), 841-849. https://doi.org/10.1093/scan/nsy060

Fishburn, F. A., Norr, M. E., Medvedev, A. V. & Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. *Frontiers in Human Neuroscience*, *8*(76), 1-11. https://doi.org/10.3389/fnhum.2014.00076

Forbes, S. H., Wijeakumar, S., Eggebrecht, A. T., Magnotta, V. A., & Spencer, J. P. (2021). Processing pipeline for image reconstructed fNIRS analysis using both MRI templates and individual anatomy. *Neurophotonics*, *8*(2), 025010:1-18. https://doi.org/10.1117/1.NPh.8.2.025010

Fox, S. E., Levitt, P., & Nelson III, C. A. (2010). How the timing and quality of early experiences influence the development of brain architecture. *Child Development*, *81*(1), 28-40. https://doi.org/10.1111/j.1467-8624.2009.01380.x

Frühholz, S., Godde, B., Finke, M., & Herrmann, M. (2011). Spatio-temporal brain dynamics in a combined stimulus–stimulus and stimulus–response conflict task. *NeuroImage*, *54*(1), 622-634. https://doi.org/10.1016/j.neuroimage.2010.07.071

Frühholz, S., Godde, B., Lewicki, P., Herzmann, C., & Herrmann, M. (2011). Face recognition under ambiguous visual stimulation: fMRI correlates of "encoding styles". *Human Brain Mapping*, *32*(10), 1750-1761. https://doi.org/10.1002/hbm.21144

Gabrieli, J. D., Poldrack, R. A., & Desmond, J. E. (1998). The role of left prefrontal cortex in language and memory. *Proceedings of the national Academy of Sciences*, *95*(3), 906-913. https://doi.org/10.1073/pnas.95.3.906

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, *1*(2), 292–304. https://doi.org/10.1111/j.1756-8765.2009.01020.x

Gerstenberger, C., Partanen, N. & Reißler, M. (2017). Instant annotations in ELAN corpora of spoken and written Komi, an endangered language of the Barents Sea region. In *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages, Association for Computational Linguistics*, (pp. 57-66). https://doi.org/10.18653/v1/W17-0109

Goghari, V. M., & MacDonald III, A. W. (2009). The neural basis of cognitive control: Response selection and inhibition. *Brain and Cognition*, *71*(2), 72-83. https://doi.org/10.1016/j.bandc.2009.04.004

Göksun, T., Hirsh-Pasek, K. & Golinkoff, R. M. (2010). How do preschoolers express cause in gesture and speech? *Cognitive Development, 25*(1), 56-68. https://doi.org/10.1016/j.cogdev.2009.11.001

Gopnik, A. & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin, 138*(6), 1085-1108. https://doi.org/10.1037/a0028044

Grassmann, S., Schulze, C. & Tomasello, M. (2015). Children's level of word knowledge predicts their exclusion of familiar objects as referents of novel words. *Frontiers in Psychology*, *6*(1200*)*, 1-8. https://doi.org/10.3389/fpsyg.2015.01200

Grindrod, C. M., Bilenko, N. Y., Myers, E. B. & Blumstein, S. E. (2008). The role of the left inferior frontal gyrus in implicit semantic competition and selection: An event-related fMRI study. *Brain Research*, *1229*, 167-178. https://doi.org/10.1016/j.brainres.2008.07.017

Grinsted, A., Moore, J. C. & Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. Retrieved from: https://hal.archives-ouvertes.fr/hal-00302394

Grosbras, M. H., Leonards, U., Lobel, E., Poline, J. B., LeBihan, D., & Berthoz, A. (2001). Human cortical networks for new and familiar sequences of saccades. *Cerebral Cortex*, *11*(10), 936-945. https://doi.org/10.1093/cercor/11.10.936

Guo, T., Liu, H., Misra, M., & Kroll, J. F. (2011). Local and global inhibition in bilingual word production: fMRI evidence from Chinese–English bilinguals. *NeuroImage*, *56*(4), 2300-2309. https://doi.org/10.1016/j.neuroimage.2011.03.049

Guo, X., Zheng, L., Zhu, L., Li, J., Wang, Q., Dienes, Z., & Yang, Z. (2013). Increased neural responses to unfairness in a loss context. *NeuroImage*, *77*, 246-253. https://doi.org/10.1016/j.neuroimage.2013.03.048

Gupta, P. & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language*, *59*(2), 267-333. https://doi.org/10.1006/brln.1997.1819

Habeck, C., Rakitin, B., Steffener, J., & Stern, Y. (2012). Contrasting visual working memory for verbal and non-verbal material with multivariate analysis of fMRI. *Brain Research*, *1467*, 27-41. https://doi.org/10.1016/j.brainres.2012.05.045

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

Han, S., Huettel, S. A., Raposo, A., Adcock, R. A., & Dobbins, I. G. (2010). Functional significance of striatal responses during episodic decisions: recovery or goal attainment? *Journal of Neuroscience*, *30*(13), 4767-4775. https://doi.org/10.1523/JNEUROSCI.3077-09.2010

Han, J. & Neuharth-Pritchett, S. (2014). Caregivers' interactions with pre-schoolers during shared book reading: Three strategies for promoting quality interactions. *Childhood Education*, *90*(1), 54–60. https://doi.org/10.1080/00094056.2014.872516

Hartwright, C. E., Apperly, I. A., & Hansen, P. C. (2014). Representation, control, or reasoning? Distinct functions for theory of mind within the medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *26*(4), 683-698. https://doi.org/10.1162/jocn_a_00520

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

Henderson, L. M., Yoder, P. J., Yale, M. E. & McDuffie, A. (2002). Getting the point: Electrophysiological correlates of protodeclarative pointing. *International Journal of Developmental Neuroscience*, *20*(3–5), 449–458. https://doi.org/10.1016/S0736-5748(02)00038-2

Hirotani, M., Stets, M., Striano, T. & Friederici, A. D. (2009). Joint attention helps infants learn new words: Event-related potential evidence. *Neuroreport*, *20*(6), 600–605. https://doi.org/10.1097/WNR.0b013e32832a0a7c

Hirsch, J., Zhang, X., Noah, J. A. & Ono, Y. (2017). Frontal temporal and parietal systems synchronize within and across brains during live eye-to-eye contact. *NeuroImage*, *157*, 314–330. https://doi.org/10.1016/j.neuroimage.2017.06.018

Hirshorn, E. A. & Thompson-Schill, S. L. (2006). Role of the left inferior frontal gyrus in covert word retrieval: Neural correlates of switching during verbal fluency. *Neuropsychologia*, *44*(12), 2547-2557. https://doi.org/10.1016/j.neuropsychologia.2006.03.035

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88. https://doi.org/10.1016/j.dr.2005.11.002

Hofstetter, S., Friedmann, N. & Assaf, Y. (2017). Rapid language-related plasticity: Microstructural changes in the cortex after a short session of new word learning. *Brain Structure and Function*, *222*(3), 1231-1241. https://doi.org/10.1007/s00429-016-1273-2

Holper, L., Scholkmann, F. & Wolf, M. (2012). Between-brain connectivity during imitation measured by fNIRS. *NeuroImage*, *63*(1), 212–222. https://doi.org/10.1016/j.neuroimage.2012.06.028

Hooker, C. I., Germine, L. T., Knight, R. T., & D'Esposito, M. (2006). Amygdala response to facial expressions reflects emotional learning. *Journal of Neuroscience*, *26*(35), 8915-8922. https://doi.org/10.1523/JNEUROSCI.3048-05.2006

Horst, J. S. & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research.

Behavior Research Methods, 48(4), 1393-1409. https://doi.org/10.3758/s13428-015-0647-3

Huppert, T. J., Diamond, S. G., Franceschini, M. A. & Boas, D. A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics*, *48*(10), D280-D298. https://doi.org/10.1364/AO.48.00D280

Huppert, T. J., Hoge, R. D., Diamond, S. G. Franceshini, M. A. & Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage, 29*(2), 368-382. https://doi.org/10.1016/j.neuroimage.2005.08.065

Huster, R. J., Eichele, T., Enriquez-Geppert, S., Wollbrink, A., Kugel, H., Konrad, C., & Pantev, C. (2011). Multimodal imaging of functional networks and event-related potentials in performance monitoring. *NeuroImage*, *56*(3), 1588-1597. https://doi.org/10.1016/j.neuroimage.2011.03.039

Iidaka, T., Matsumoto, A., Haneda, K., Okada, T., & Sadato, N. (2006). Hemodynamic and electrophysiological relationship involved in human face processing: Evidence from a combined fMRI–ERP study. *Brain and Cognition*, *60*(2), 176-186. https://doi.org/10.1016/j.bandc.2005.11.004

Jamadar, S. D., Johnson, B. P., Clough, M., Egan, G. F., & Fielding, J. (2015). Behavioral and neural plasticity of ocular motor control: changes in performance and fMRI activity following antisaccade training. *Frontiers in Human Neuroscience*, *9*(653), 1-12. https://doi.org/10.3389/fnhum.2015.00653

Jesse, A. & Johnson, E. K. (2016). Audiovisual alignment of co-speech gestures to speech supports word learning in 2-year-olds. *Journal of Experimental Child Psychology*, *145*, 1–10. https://doi.org/10.1016/j.jecp.2015.12.002

Jiang, J., Chen, C., Dai, B., Shi, G., Ding, G., Liu, L. & Lu, C. (2015). Leader emergence through interpersonal neural synchronization. *Proceedings of the National Academy of Sciences*, *112*(14), 4274–4279. https://doi.org/10.1073/pnas.1422930112

Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L. & Lu, C. (2012). Neural synchronization during face-to-face communication. *Journal of Neuroscience*, *32*(45), 16064–16069. https://doi.org/10.1523/JNEUROSCI.2926-12.2012

Johnson-Frey, S. H., Newman-Norlund, R., & Grafton, S. T. (2005). A distributed left hemisphere network active during planning of everyday tool use skills. *Cerebral Cortex*, *15*(6), 681-695. https://doi.org/10.1093/cercor/bhh169

Kalagher, H. & Yu, C. (2006). The effects of deictic pointing in word learning. In *Proceedings of the 5th International Conference of Development and Learning*. ISBN: 0-9786456-0-X

Kalashnikova, M., Mattock, K. & Monaghan, P. (2015). The effects of linguistic experience on the flexible use of mutual exclusivity in word learning. *Bilingualism*, *18*(4), 626-638. https://doi.org/10.1017/S1366728914000364

Kepinska, O., de Rover, M., Caspers, J., & Schiller, N. O. (2018). Connectivity of the hippocampus and Broca's area during acquisition of a novel grammar. *NeuroImage*, *165*, 1-10. https://doi.org/10.1016/j.neuroimage.2017.09.058

Khader, P. H., Jost, K., Mertens, M., Bien, S., & Rösler, F. (2010). Neural correlates of generating visual nouns and motor verbs in a minimal phrase context. *Brain Research*, *1318*, 122-132. https://doi.org/10.1016/j.brainres.2009.12.082

Kim, J., Yeon, J., Ryu, J., Park, J. Y., Chung, S. C., & Kim, S. P. (2017). Neural activity patterns in the human brain reflect tactile stickiness perception. *Frontiers in Human Neuroscience*, *11*(445), 1-12. https://doi.org/10.3389/fnhum.2017.00445

Kingsbury, L., Huang, S., Wang, J., Gu, K., Golshani, P., Wu, Y. E. & Hang, W. (2019). Correlated neural activity and encoding of behaviour across brains of socially interacting animals. *Cell, 178*(2), 429-446. https://doi.org/10.1016/j.cell.2019.05.022

Kirilina, E., Yu, N., Jelzow, A., Wabnitz, H., Jacobs, A. M. & Tachtsidis, I. (2013). Identifying and quantifying main components of physiological noise in functional near infrared spectroscopy on the prefrontal cortex. *Frontiers in Human Neuroscience*, *7*(864), 1-17. https://doi.org/10.3389/fnhum.2013.00864

Kitada, R., Johnsrude, I. S., Kochiyama, T., & Lederman, S. J. (2009). Functional specialization and convergence in the occipito-temporal cortex supporting haptic and visual identification of human faces and body parts: an fMRI study. *Journal of Cognitive Neuroscience*, *21*(10), 2027-2045. https://doi.org/10.1162/jocn.2009.21115

Koike, T., Tanabe, H. C., Okazaki, S., Nakagawa, E., Sasaki, A. T., Shimada, K., Sugawara, S. K., Takahashi, H. K., Kazufumi, Y., Bosch-Bayard, J. B. & Sadato, N. (2016). Neural substrates of shared attention as social memory: a hyperscanning functional magnetic resonance imaging study. *NeuroImage*, *125*, 401-412. https://doi.org/10.1016/j.neuroimage.2015.09.076

Kruse, B., Bogler, C., Haynes, J. D., & Schütz-Bosbach, S. (2016). Am I seeing myself, my friend or a stranger? The role of personal familiarity in visual distinction of body identities in the human brain. *Cortex*, *83*, 86-100. https://doi.org/10.1016/j.cortex.2016.07.010

Kühn, S., Haggard, P., & Brass, M. (2009). Intentional inhibition: How the “veto-area” exerts control. *Human Brain Mapping*, *30*(9), 2834-2843. https://doi.org/10.1002/hbm.20711

Kyttälä, M., Aunio, P., Lepola, J. & Hautamäki, J. (2014). The role of the working memory and language skills in the prediction of word problem solving in 4- to 7-year-old children. *Educational Psychology, 34*(6), 674–696. https://doi.org/10.1080/01443410.2013.814192

Lancia S., Mammarella S., Bianco D. & Quaresima V. (2018) Is Wireless Functional Near-Infrared Spectroscopy (fNIRS) 3D Neuroimaging Feasible to Map Human Navigation in the Real-World? In: Fogliaroni P., Ballatore

A., Clementini E. (Eds) *Proceedings of Workshops and Posters at the 13th International Conference on Spatial Information Theory (COSIT 2017*). Lecture Notes in Geoinformation and Cartography. Springer, Cham. https://doi.org/10.1007/978-3-319-63946-8_16

Lau, H. C., Rogers, R. D. & Passingham, R. E. (2006). On measuring the perceived onsets of spontaneous actions. *Journal of Neuroscience*, *26*(27), 7265-7271. https://doi.org/10.1523/JNEUROSCI.1138-06.2006

Law, F., Mahr, T., Schneeberg, A. & Edwards, J. (2017). Vocabulary size and auditory word recognition in preschool children. *Applied Psycholinguistics*, *38*(1), 89-125. https://doi.org/10.1017/S0142716416000126

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

Lewis, M., Cristiano, V., Lake, B. M., Kwan, T. & Frank, M. C. (2020). The role of developmental change and linguistic experience in the mutual exclusivity effect. *Cognition*, *198*, 104191. https://doi.org/10.1016/j.cognition.2020.104191

Lewis, P. A., Critchley, H. D., Smith, A. P., & Dolan, R. J. (2005). Brain mechanisms for mood congruent memory facilitation. *NeuroImage*, *25*(4), 1214-1223. https://doi.org/10.1016/j.neuroimage.2004.11.053

Liebeskind, K. G., Piotrowski, J. T., Lapierre, M. A. & Linebarger, D. L. (2014). The home literacy environment: Exploring how media and parent-child interactions are associated with children's language production. *Journal of Early Childhood Literacy, 14*(4), 482-509. https://doi.org/10.1177/1468798413512850

Liljeström, M., Hultén, A., Parkkonen, L. & Salmelin, R. (2009). Comparing MEG and fMRI views to naming actions and objects. *Human Brain Mapping*, *30*(6), 1845-1856. https://doi.org/10.1002/hbm.20785

Liljeström, M., Tarkiainen, A., Parviainen, T., Kujala, J., Numminen, J., Hiltunen, J., Laine, R. & Salmelin, R. (2008). Perceiving and naming actions and objects. *NeuroImage*, *41*(3), 1132-1141. https://doi.org/10.1016/j.neuroimage.2008.03.016

Limanowski, J., Lutti, A., & Blankenburg, F. (2014). The extrastriate body area is involved in illusory limb ownership. *NeuroImage*, *86*, 514-524. https://doi.org/10.1016/j.neuroimage.2013.10.035

Lin, N., Lu, X., Fang, F., Han, Z., & Bi, Y. (2011). Is the semantic category effect in the lateral temporal cortex due to motion property differences? *NeuroImage*, *55*(4), 1853-1864. https://doi.org/10.1016/j.neuroimage.2011.01.039

Liu, W., Branigan, H. P., Zheng, L., Long, Y., Bai, X., Li, K., Zhao. H., Zhou, S., Pickering, M. J. & Lu, C. (2019). Shared neural representations of syntax during online dyadic communication. *NeuroImage*, *198*, 63–72. https://doi.org/10.1016/j.neuroimage.2019.05.035

Liu, N., Mok, C., Witt, E. E., Pradhan, A. H., Chen, J. E. & Reiss, A. L. (2016). NIRS-based hyperscanning reveals inter-brain neural synchronization during cooperative Jenga game with face-to-face communication. *Frontiers in Human Neuroscience*, *10*(82), 1-11. https://doi.org/10.3389/fnhum.2016.00082

Liu, T., Saito, G., Lin, C. & Saito, H. (2017). Inter-brain network underlying turn-based cooperation and competition: A hyperscanning study using near-infrared spectroscopy. *Scientific Reports*, *7*(8684), 1–12. https://doi.org/10.1038/s41598-017-09226-w

Lonigan, C. J. (2007). Vocabulary development and the development of phonological awareness skills in preschool children. In R. K. Wagner, A. E. Muse & K. R. Tenenbaum (Eds.). *Vocabulary Acquisition: Implications for Reading Comprehension*, 15-31. The Guilford Press. ISBN 13: 978-1-59385-338-9

López-Barroso, D., Catani, M., Ripollés, P., Dell'Acqua, F., Rodríguez-Fornells, A. & De Diego-Balaguer, R. (2013). Word learning is mediated by the left arcuate fasciculus. *Proceedings of the National Academy of Sciences*, *110*(32), 13168–13173. https://doi.org/10.1073/pnas.1301696110

Lu, C. M., Zhang, Y. J., Biswal, B. B., Zang, Y. F., Peng, D. L. & Zhu, C. Z. (2010). Use of fNIRS to assess resting state functional connectivity. *Journal of Neuroscience Methods*, *186*(2), 242-249. https://doi.org/10.1016/j.jneumeth.2009.11.010

Majerus, S., Poncelet, M., Van der Linden, M., Albouy, G., Salmon, E., Sterpenich, V., Vandewalle, G., Collette, F. & Maquet, P. (2006). The left intraparietal sulcus and verbal short-term memory: Focus of attention or serial order? *NeuroImage*, *32*(2), 880-891. https://doi.org/10.1016/j.neuroimage.2006.03.048

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M. & Ro, T. (2009). To see or not to see: Prestimulus α phase predicts visual awareness. *Journal of Neuroscience*, *29*(9), 2725-2732. https://doi.org/10.1523/JNEUROSCI.3963-08.2009

Miller, J. G., Vrtička, P., Cui, X., Shrestha, S., Hosseini, S. M. H., Baker, J. M. & Reiss, A. L. (2019). Inter-brain synchrony in mother-child dyads during cooperation: An fNIRS hyperscanning study. *Neuropsychologia*, *124*, 117–124. https://doi.org/10.1016/j.neuropsychologia.2018.12.021

Millspaugh, J. J., Gitzen, R. A., Kernohan, B. J., Larson, M. A. & Clay, C. L. (2004). Comparability of three analytical techniques to assess joint space use. *Wildlife Society Bulletin*, *32*(1), 148–157. https://doi.org/10.2193/0091-7648(2004)32[148:COTATT]2.0.CO;2

Moser, D., Baker, J. M., Sanchez, C. E., Rorden, C., & Fridriksson, J. (2009). Temporal order processing of syllables in the left parietal lobe. *Journal of Neuroscience*, *29*(40), 12568-12573. https://doi.org/10.1523/JNEUROSCI.5934-08.2009

Muehlhaus, J., Heim, S., Altenbach, F., Chatterjee, A., Habel, U., & Sass, K. (2014). Deeper insights into semantic relations: an fMRI study of part-whole and functional associations. *Brain and Language*, *129*, 30-42. https://doi.org/10.1016/j.bandl.2014.01.003

Munakata, Y. & McClelland, J. L. (2003). Connectionist models of development. *Developmental Science*, *6*(4), 413–429. https://doi.org/10.1111/1467-7687.00296

Mundy, P., & Jarrold, W. (2010). Infant joint attention, neural networks and social cognition. *Neural Networks*, *23*(8–9), 985–997. https://doi.org/10.1016/j.neunet.2010.08.009

Napoli, A. R. & Purpura, D. J. (2018). The home literacy and numeracy environment in preschool: Cross-domain relations of parent-child practices and child outcomes. *Journal of Experimental Child Psychology, 166*, 581-603. https://doi.org/10.1016/j.jecp.2017.10.002

Newbury, J., Klee, T., Stokes, S. F. & Moran, C. (2016). Interrelationships between working memory, processing speed, and language development in the age range 2–4 years. *Journal of Speech, Language, and Hearing Research*, *59*(5), 1146-1158. https://doi.org/10.1044/2016_JSLHR-L-15-0322

Newman, A. J., Kenny, S., Saint-Aubin, J., & Klein, R. M. (2013). Can skilled readers perform a second task in parallel? A functional connectivity MRI study. *Brain and Language*, *124*(1), 84-95. https://doi.org/10.1016/j.bandl.2012.11.009

Nieuwland, M. S., Petersson, K. M. & Van Berkum, J. J. (2007). On sense and reference: Examining the functional neuroanatomy of referential processing. *NeuroImage*, *37*(3), 993-1004. https://doi.org/10.1016/j.neuroimage.2007.05.048

Nora, A., Karvonen, L., Renvall, H., Parviainen, T., Kim, J.-Y., & Salmelin, R. (2017). Children show right-lateralized effects of spoken word-form learning. *PloS One*, *12*(2), e0171034. https://doi.org/10.1371/journal.pone.0171034

Otten, L. J., Henson, R. N. A. & Rugg, M. D. (2001). Depth of processing effects on neural correlates of memory encoding: relationship between findings from across-and within-task comparisons. *Brain*, *124*(2), 399–412. https://doi.org/10.1093/brain/124.2.399

Pan, Y., Cheng, X., Zhang, Z., Li, X. & Hu, Y. (2017). Cooperation in lovers: An fNIRS-based hyperscanning study. *Human Brain Mapping*, *38*(2), 831–841. https://doi.org/10.1002/hbm.23421

Pereira, A. F., Smith, L. B. & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review, 21*(1), 178–185. https://doi.org/10.3758/s13423-013-0466-4

Pergola, G., Ranft, A., Mathias, K., & Suchan, B. (2013). The role of the thalamic nuclei in recognition memory accompanied by recall during encoding and retrieval: an fMRI study. *NeuroImage*, *74*, 195-208. https://doi.org/10.1016/j.neuroimage.2013.02.017

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

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S. & Burgess, P. W. (2018). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1464*, 5–29. https://doi.org/10.1111/nyas.13948

Piper, S. K., Krueger, A., Koch, S. P., Mehnert, J., Habermehl, C., Steinbrink, J., Obrig, H. & Schmitz, C. H. (2014). A wearable multi-channel fNIRS system for brain imaging in freely moving subjects. *NeuroImage, 85*(1), 64-71. https://doi.org/10.1016/j.neuroimage.2013.06.062

Prado, J., Van Der Henst, J. B., & Noveck, I. A. (2010). Recomposing a fragmented literature: How conditional and relational arguments engage different neural systems for deductive reasoning. *NeuroImage*, *51*(3), 1213-1221. https://doi.org/10.1016/j.neuroimage.2010.03.026

Quené, H. & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413–425. https://doi.org/10.1016/j.jml.2008.02.002

Quinn, C., Taylor, J. S. H., & Davis, M. H. (2017). Learning and retrieving holistic and componential visual-verbal associations in reading and object naming. *Neuropsychologia*, *98*, 68-84. https://doi.org/10.1016/j.neuropsychologia.2016.09.025

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779-788). Retrieved from: https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Redmon_You_Only_Look_CVPR_2016_paper.pdf

Reindl, V., Gerloff, C., Scharke, W. & Konrad, K. (2018). Brain-to-brain synchrony in caregiver-child dyads and the relationship with emotion regulation revealed by fNIRS-based hyperscanning. *NeuroImage*, *178*, 493–502. https://doi.org/10.1016/j.neuroimage.2018.05.060

Rice, K., Moraczewski, D. & Redcay, E. (2016). Perceived live interaction modulates the developing social brain. *Social Cognitive and Affective Neuroscience*, *11*(9), 1354–1362. https://doi.org/10.1093/scan/nsw060

Rice, K. & Redcay, E. (2016). Interaction matters: A perceived social partner alters the neural processing of human speech. *NeuroImage*, *129*, 480–488. https://doi.org/10.1016/j.neuroimage.2015.11.041

Richards, J. E., Sanchez, C., Phillips-Meek, M. & Xie, W. (2016). A database of age-appropriate average MRI templates. *NeuroImage*, *124*, 1254-1259. https://doi.org/10.1016/j.neuroimage.2015.04.055

Richards, J. E. & Xie, W. (2015). Brains for all the ages: Structural neurodevelopment in infants and children from a life-span perspective. In *Advances in Child Development and Behavior* (Vol. 48, pp. 1-52). https://doi.org/10.1016/bs.acdb.2014.11.001

Richardson, D. C., Dale, R. & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological Science*, *18*(5), 407–413. https://doi.org/10.1111/j.1467-9280.2007.01914.x

Richardson, D. C. & Kirkham, N. Z. (2004). Multimodal events and moving locations: Eye movements of adults and 6-month-olds reveal dynamic spatial indexing. *Journal of Experimental Psychology: General, 133*(1), 46–62. https://doi.org/10.1037/0096-3445.133.1.46

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L. & Gabrieli, J. D. E. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science*, *29*(5), 700–710. https://doi.org/10.1177/0956797617742725

Romeo, R. R., Segaran, J., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Yendiki, A., Rowe, M. L. & Gabrieli, J. D. (2018). Language exposure relates to structural neural connectivity in childhood. *Journal of Neuroscience*, *38*(36), 7870-7877. https://doi.org/10.1523/JNEUROSCI.0484-18.2018

Ruzzoli, M., Torralba, M., Fernández, L. M. & Soto-Faraco, S. (2019). The relevance of alpha phase in human perception. *Cortex*, *120*, 249-268. https://doi.org/10.1016/j.cortex.2019.05.012

Sabb, F. W., Bilder, R. M., Chou, M., & Bookheimer, S. Y. (2007). Working memory effects on semantic processing: priming differences in pars orbitalis. *NeuroImage*, *37*(1), 311-322. https://doi.org/10.1016/j.neuroimage.2007.04.050

Sakaki, M., Niki, K., & Mather, M. (2011). Updating existing emotional memories involves the frontopolar/orbito-frontal cortex in ways that acquiring new emotional memories does not. *Journal of Cognitive Neuroscience*, *23*(11), 3498-3514. https://doi.org/10.1162/jocn_a_00057

Sakaki, M., Niki, K., & Mather, M. (2012). Beyond arousal and valence: The importance of the biological versus social relevance of emotional stimuli. *Cognitive, Affective, & Behavioral Neuroscience*, *12*(1), 115-139. https://doi.org/10.3758/s13415-011-0062-x

Salminen, T., Kühn, S., Frensch, P. A., & Schubert, T. (2016). Transfer after dual n-back training depends on striatal activation change. *Journal of Neuroscience*, *36*(39), 10198-10213. https://doi.org/10.1523/JNEUROSCI.2305-15.2016

Samuelson, L. K., Kucker, S. C. & Spencer, J. P. (2017). Moving word learning to a novel space: A dynamic systems view of referent selection and retention. *Cognitive Science*, *41*(51), 52–72. https://doi.org/10.1111/cogs.12369

Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *PloS One*, *6*(12), e28095. https://doi.org/10.1371/journal.pone.0028095

Sato, W., Kochiyama, T., Uono, S., & Toichi, M. (2016). Neural mechanisms underlying conscious and unconscious attentional shifts triggered by eye gaze. *NeuroImage*, *124*, 118-126. https://doi.org/10.1016/j.neuroimage.2015.08.061

Saxe, R. (2013). The new puzzle of theory of mind development. In M. R. Banaji, S. A. Gelman (Eds.). *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us*, 107-112. Oxford University Press. ISBN 978-0-19-989071-2

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Pavia, J. M., Wolf, U. & Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *NeuroImage*, *85*, 6-27. https://doi.org/10.1016/j.neuroimage.2013.05.004

Scholkmann, F., Holper, L., Wolf, U. & Wolf, M. (2013). A new methodical approach in neuroscience : assessing inter-personal brain coupling using functional near-infrared imaging (fNIRI) hyperscanning. *Frontiers in Human Neuroscience*, *7*(813), 1–6. https://doi.org/10.3389/fnhum.2013.00813

Schulte, T., Müller-Oehring, E. M., Vinco, S., Hoeft, F., Pfefferbaum, A., & Sullivan, E. V. (2009). Double dissociation between action-driven and perception-driven conflict resolution invoking anterior versus posterior brain systems. *NeuroImage*, *48*(2), 381-390. https://doi.org/10.1016/j.neuroimage.2009.06.058

Schreuders, E., Klapwijk, E. T., Will, G. J., & Güroğlu, B. (2018). Friend versus foe: Neural correlates of prosocial decisions for liked and disliked peers. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(1), 127-142. https://doi.org/10.3758/s13415-017-0557-1

Simanova, I., Hagoort, P., Oostenveld, R., & Van Gerven, M. A. (2014). Modality-independent decoding of semantic information from the human brain. *Cerebral Cortex*, *24*(2), 426-434. https://doi.org/10.1093/cercor/bhs324

Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A. & Hasson, U. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature Communications, 7*(12141), 1-13. https://doi.org/10.1038/ncomms12141

Singh, A. K., Okamoto, M., Dan, H., Jurcak, V. & Dan, I. (2005). Spatial registration of multichannel multi-subject fNIRS data to MNI space without MRI. *NeuroImage*, *27*(4), 842-851. https://doi.org/10.1016/j.neuroimage.2005.05.019

Smith, L. B., Jayaraman, S., Clerkin, E. & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, 22(4), 235-336. https://doi.org/10.1016/j.tics.2018.02.004

Snyder, H. R., Banich, M. T. & Munakata, Y. (2011). Choosing our words: retrieval and selection processes recruit shared neural substrates in left ventrolateral prefrontal cortex. *Journal of Cognitive Neuroscience*, *23*(11), 3470-3482. https://doi.org/10.1162/jocn_a_00023

Sobel, D. M. & Buchanan, D. W. (2009). Bridging the gap: Causality-at-a-distance in children's categorization and inferences about internal properties. *Cognitive Development*, *24*(3), 274-283. https://doi.org/10.1016/j.cogdev.2009.03.003

Sobel, D. M. & Corriveau, K. H. (2010). Children monitor individuals' expertise for word learning. *Child Development*, *81*(2), 669-679. https://doi.org/10.1111/j.1467-8624.2009.01422.x

Sobel, D. M. & Kirkham, N. Z. (2012). The influence of social information on children's statistical and causal inferences. In F. Xu & T. Kushner (Eds.) *Advances in Child Development and Behavior* (Vol. 43, pp. 321-350). https://doi.org/10.1016/B978-0-12-397919-3.00012-5

Sobel, D. M. & Sommerville, J. A. (2009). Rationales in children's causal learning from others' actions. *Cognitive Development*, *24*(1), 70-79. https://doi.org/10.1016/j.cogdev.2008.08.003

Sobel, D. & Sommerville, J. (2010). The importance of discovery in children's causal learning from interventions. *Frontiers in Psychology*, *1*(176), 1-7. https://doi.org/10.3389/fpsyg.2010.00176

Soltanlou, M., Sitnikova, M. A., Nuerk, H. C. & Dresler, T. (2018). Applications of functional near-infrared spectroscopy (fNIRS) in studying cognitive development: The case of mathematics and language. *Frontiers in Psychology*, *9*(277), 1-15. https://doi.org/10.3389/fpsyg.2018.00277

Sommet, N. & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using Stata, R, Mplus, and SPSS. *International Review of Social Psychology*, *30*, 203–218. https://doi.org/10.5334/irsp.90

Song, S., Su, M., Kang, C., Liu, H., Zhang, Y., McBride-Chang, C., Tardif, H., Li, H., Liang, W., Zhang, Z. & Shu, H. (2015). Tracing children's vocabulary development from preschool through the school-age years: An 8-year longitudinal study. *Developmental Science*, *18*(1), 119-131. https://doi.org/10.1111/desc.12190

Spreng, R. N., DuPre, E., Selarka, D., Garcia, J., Gojkovic, S., Mildner, J., ... & Turner, G. R. (2014). Goal-congruent default network activity facilitates cognitive control. *Journal of Neuroscience*, *34*(42), 14108-14114. https://doi.org/10.1523/JNEUROSCI.2815-14.2014

Spreng, R. N., Stevens, W. D., Chamberlain, J. P., Gilmore, A. W. & Schacter, D. L. (2010). Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *NeuroImage*, *53*(1), 303-317. https://doi.org/10.1016/j.neuroimage.2010.06.016

Steels, L. & Kaplan, F. (2000). AIBO's first words: The social learning of language and meaning. *Evolution of Communication*, *4*(1), 3-32. https://doi.org/10.1075/eoc.4.1.03ste

Stephens, G. J., Silbert, L. J. & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, *107*(32), 14425–14430. https://doi.org/10.1073/pnas.1008662107

Stokes, S. F. & Klee, T. (2009). Factors that influence vocabulary development in two-year-old children. *Journal of Child Psychology and Psychiatry*, *50*(4), 498-505. https://doi.org/10.1111/j.1469-7610.2008.01991.x

Stokes, S. F., Klee, T., Kornisch, M. & Furlong, L. (2017). Visuospatial and verbal short-term memory correlates of vocabulary ability in preschool children. *Journal of Speech, Language, and Hearing Research*, *60*(8), 2249-2258. https://doi.org/10.1044/2017_JSLHR-L-16-0285

Stolk, A., Noordzij, M. L. Verhagen, L., Volman, I., Schoffelen, J. M., Oostenveld, R., Hagoort, P. & Tone, I. (2014). Cerebral coherence between communicators marks the emergence of meaning. *Proceedings from the National Academy of Sciences, 111*(51), 18183-18188. https://doi.org/10.1073/pnas.1414886111

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

Suanda, S. H., Smith, L. B. & Yu, C. (2016). The multisensory nature of verbal discourse in caregiver–toddler interactions. *Developmental Neuropsychology*, *41*(5–8), 324–341. https://doi.org/10.1080/87565641.2016.1256403

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

Sullivan, L., Mundy, P. & Mastergeorge, A. M. (2015). Joint attention in preschool children: Is it a meaningful measure. *International Journal of School and Cognitive Psychology*, *2*(1), 1000120. https://doi.org/10.4172/1234-3425.1000120

Tak, S. & Ye, J. C. (2014). Statistical analysis of fNIRS data: A comprehensive review. *NeuroImage*, *85*, 72-91. https://doi.org/10.1016/j.neuroimage.2013.06.016

Takashima, A., Bakker-Marshall, I., Van Hell, J. G., McQueen, J. M. & Janzen, G. (2019). Neural correlates of word learning in children. *Developmental Cognitive Neuroscience*, *37*, 100649. https://doi.org/10.1016/j.dcn.2019.100649

Tamir, D. I., & Mitchell, J. P. (2011). The default network distinguishes construals of proximal versus distal events. *Journal of Cognitive Neuroscience*, *23*(10), 2945-2955. https://doi.org/10.1162/jocn_a_00009

Tamis-LeMonda, C. S., Kuchirko, Y. & Song, L. (2014). Why is infant language learning facilitated by caregivers' responsiveness? *Current Directions in Psychological Science*, *23*(2), 121–126. https://doi.org/10.1177/0963721414522813

Tashjian, S. M., Weissman, D. G., Guyer, A. E., & Galván, A. (2018). Neural response to prosocial scenes relates to subsequent giving behavior in adolescents: A pilot study. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(2), 342-352. https://doi.org/10.3758/s13415-018-0573-9

Thompson, J. C., Hardee, J. E., Panayiotou, A., Crewther, D., & Puce, A. (2007). Common and distinct brain activation to viewing dynamic sequences of face and hand movements. *NeuroImage*, *37*(3), 966-973. https://doi.org/10.1016/j.neuroimage.2007.05.058

Tomasello, M. & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*(6), 1454–1463. https://doi.org/10.2307/1130423

Trawick-Smith, J., Russell, H. & Swaminathan, S. (2011). Measuring the effects of toys on the problem-solving, creative and social behaviours of preschool children. *Early Child Development and Care, 181*(7) 909-927. https://doi.org/10.1080/03004430.2010.503892

Van der Laan, L. N., De Ridder, D. T., Viergever, M. A., & Smeets, P. A. (2012). Appearance matters: neural correlates of food choice and packaging aesthetics. *PloS One*, *7*(7), e41738. https://doi.org/10.1371/journal.pone.0041738

Vlach, H. A. & DeBrock, C. A. (2017). Remember dax? Relations between children's cross-situational word learning, memory, and language abilities. *Journal of Memory and Language*, *93*, 217-230. https://doi.org/10.1016/j.jml.2016.10.001

Vlach, H. A. & DeBrock, C. A. (2019). Statistics learned are statistics forgotten: Children's retention and retrieval of cross-situational word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 700-711. https://doi.org/10.1037/xlm0000611

Vogel, A. C., Petersen, S. E., & Schlaggar, B. L. (2012). The left occipitotemporal cortex does not show preferential activity for words. *Cerebral Cortex*, *22*(12), 2715-2732. https://doi.org/10.1093/cercor/bhr295

Vossel, S., Thiel, C. M., & Fink, G. R. (2006). Cue validity modulates the neural correlates of covert endogenous orienting of attention in parietal and frontal cortex. *NeuroImage*, *32*(3), 1257-1264. https://doi.org/10.1016/j.neuroimage.2006.05.019

Walter, W. D., Onorato, D. P. & Fischer, J. W. (2015). Is there a single best estimator? Selection of home range estimators using area-under-the-curve. *Movement Ecology*, *3*(10), 1-11. https://doi.org/10.1186/s40462-015-0039-4

Walter, H., von Kalckreuth, A., Schardt, D., Stephan, A., Goschke, T., & Erk, S. (2009). The temporal dynamics of voluntary emotion regulation. *PLoS One*, *4*(8), e6726. https://doi.org/10.1371/journal.pone.0006726

Wass, S. V, Whitehorn, M., Haresign, I. M., Phillips, E. & Leong, V. (2020). Interpersonal neural entrainment during early social interaction. *Trends in Cognitive Sciences*, 24(4), 329-342. https://doi.org/10.1016/j.tics.2020.01.006

Waugh, C. E., Hamilton, J. P., & Gotlib, I. H. (2010). The neural temporal dynamics of the intensity of emotional experience. *NeuroImage*, *49*(2), 1699-1707. https://doi.org/10.1016/j.neuroimage.2009.10.006

Wei, R., Leech, K. A., & Rowe, M. L. (2020). Decontextualized language use during Chinese and American caregiver-child interactions. *Journal of Applied Developmental Psychology*, *71*(101214), 1-12. https://doi.org/10.1016/j.appdev.2020.101214

Wellman, H. M., Cross, D. & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

Wendelken, C., Chung, D., & Bunge, S. A. (2012). Rostrolateral prefrontal cortex: Domain-general or domain-sensitive? *Human Brain Mapping*, *33*(8), 1952-1963. https://doi.org/10.1002/hbm.21336

White, C. N., Congdon, E., Mumford, J. A., Karlsgodt, K. H., Sabb, F. W., Freimer, N. B., London, E. D., Cannon, T. D., Bilder, R. M. & Poldrack, R. A. (2014). Decomposing decision components in the stop-signal task: a model-based approach to individual differences in inhibitory control. *Journal of Cognitive Neuroscience*, *26*(8), 1601-1614. https://doi.org/10.1162/jocn_a_00567

Whitehead, C., Marchant, J. L., Craik, D. & Frith, C. D. (2009). Neural correlates of observing pretend play in which one object is represented as another. *Social Cognitive and Affective neuroscience*, *4*(4), 369-378. https://doi.org/10.1093/scan/nsp021

Wijeakumar, S., Huppert, T. J., Magnotta, V. A., Buss, A. T. & Spencer, J. P. (2017). Validating an image-based fNIRS approach with fMRI and a working memory task. *NeuroImage*, *147*, 204-218. https://doi.org/10.1016/j.neuroimage.2016.12.007

Wijeakumar, S., Spencer, J. P., Bohache, K., Boas, D. A. & Magnotta, V. A. (2015). Validating a new methodology for optical probe design and image registration in fNIRS studies. *NeuroImage*, *106*, 86-100. https://doi.org/10.1016/j.neuroimage.2014.11.022

Wilcox, T. & Biondi, M. (2015). fNIRS in the developmental sciences. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(3), 263-283. https://doi.org/10.1002/wcs.1343

Williams, J. H., Waiter, G. D., Perra, O., Perrett, D. I., & Whiten, A. (2005). An fMRI study of joint attention experience. *NeuroImage*, *25*(1), 133-140. https://doi.org/10.1016/j.neuroimage.2004.10.047

Wittenburg, P. Brugman, H., Russel, A., Klassmann, A. & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *5th International Conference on Language Resources and Evaluation,* (pp. 1556-

1559). Retrieved from: pure.mpg.de/rest/items/item_60436_2/component/file_60437/content

Wong, F. C., Chandrasekaran, B., Garibaldi, K. & Wong, P. C. (2011). White matter anisotropy in the ventral language pathway predicts sound-to-word learning success. *Journal of Neuroscience*, *31*(24), 8780-8785. https://doi.org/10.1523/JNEUROSCI.0999-11.2011

Wu, R. & Kirkham, N. Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology*, *107*(2), 118-136. https://doi.org/10.1016/j.jecp.2010.04.014

Wutte, M. G., Glasauer, S., Jahn, K., & Flanagin, V. L. (2012). Moving and being moved: Differences in cerebral activation during recollection of whole-body motion. *Behavioural Brain Research*, *227*(1), 21-29. https://doi.org/10.1016/j.bbr.2011.09.042

Xu, Y., Lin, Q., Han, Z., He, Y. & Bi, Y. (2016). Intrinsic functional architecture of human semantic processing: Modules and hubs. *NeuroImage, 132*(15), 542-555. https://doi.org/10.1016/j.neuroimage.2016.03.004

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670. https://doi.org/10.1038/nmeth.1635

Yokoyama, S., Watanabe, J., Iwata, K., Ikuta, N., Haji, T., Usui, N., Taira, M., Miyamoto, T., Nakamura, W., Sato, S., Horie, K. & Kawashima, R. (2007). Is Broca's area involved in the processing of passive sentences? An event-related fMRI study. *Neuropsychologia*, *45*(5), 989-996. https://doi.org/10.1016/j.neuropsychologia.2006.09.003

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

Yu, C. & Smith, L. B. (2013). Joint attention without gaze following : Human infants and their caregivers coordinate visual attention to objects through eye-hand coordination, *PLoS ONE*, *8*(11), e79659. https://doi.org/10.1371/journal.pone.0079659

Yu, C., Smith, L. B., Shen, H., Pereira, A. F. & Smith, T. (2009). Active information selection: Visual attention through the hands. *IEEE Transactions on Autonomous Mental Development*, *1*(2), 141-151. https://doi.org/10.1109/TAMD.2009.2031513

Yücel, M. A., Selb, J., Cooper, R. J. & Boas, D. A. (2014). Targeted principle component analysis: A new motion artifact correction approach for near-infrared spectroscopy. *Journal of Innovative Optical Health Sciences*, *7*(2), 1350066:1-8. https://doi.org/10.1142/S1793545813500661

Zahavi, D. (2015). You, me, and we: The sharing of emotional experiences. *Journal of Consciousness Studies*, *22*(1-2), 84-101.

Zhang, Y., Huang, P., Song, Z., Fang, L., Shen, T., Li, Y., Gong, Q. & Xie, P. (2014). In-context language control with production tasks in bilinguals: An fMRI study. *Brain Research*, *1585*, 131-140. https://doi.org/10.1016/j.brainres.2014.08.027

Zhang, L., Lin, L., Liang, X. & He, K. (2016). Is Faster R-CNN doing well for pedestrian detection? In Leibe, E. Matas, J., Sebe, N. & Welling, M, (Eds.) *Computer Vision-ECCV 2016, Lecture Notes in Computer Science, 9906*, 443–457. Springer, Cham. https://doi.org/10.1007/978-3-319-46475-6_28

Appendix A. – Studies for Regions of Interest

Network	No. Title	Authors	Year	Reference
Visual Working Memory	1. Validating a new methodology for optical probe design and image registration in fNIRS studies	Wijeakumar et al.	2015	Table 2
Word Learning	1. Regional differences in the developmental trajectory of lateralization of the language network	Berl et al.	2014	Table 3, mean group activation
	2. Brain basis of communicative actions in language	Egorova, N., Shtyrov, Y. & Pulvermüller, F.	2016	Table 4, naming > request
	3. Automatic semantic facilitation in anterior temporal cortex revealed through multimodal imaging	Lau, E.F. et al.	2013	Table 1
	4. Word learning is mediated by the left arcuate fasciculus	López-Barroso, D. et al.	2013	Table S5
	5. Localising semantic and syntactic processing in spoken and written language comprehension: An Activation Likelihood Estimation meta-analysis	Rodd, J.M. et al.	2015	Table 3, semantics
	6. A 10-year longitudinal fMRI study of narrative comprehension in children and adolescents	Szaflarski, J.P., et al.	2012	Table 2, decreases with age
Social Interaction	1. The neural bases of social intention understanding: The role of interaction goals	Canessa, N. et al.	2012	Table 1
	2. A frontotemporoparietal network common to initiating and responding to joint attention bids	Caruana, N., Brock, J. & Woolgar, A.	2015	Table 2, c
	3. Theory of mind performance in children correlates with functional specialization of a brain region for thinking about thoughts	Gewon, H. et al.	2012	Table 1, children

4. Imaging social motivation: Distinct brain mechanisms drive effort production during collaboration versus competition	Le Bouc, R. & Pessiglione, M.	2013	Table 1, competition and collaboration
5. Social In social out: How the brain responds to social language with more social language	O'Donnell, M.B. et al.	2015	Table 4 and 6
6. Look into my eyes: Investigating joint attention using interactive eye tracking and fMRI in a developmental sample	Oberwelland et al.	2016	Table 1
7. Perceived live interaction modulates the developing social brain	Rice, K., Moraczewski, D. & Redcay, E.	2016	Table 2.A.1, Live > Social and 2.B. Live > Social
8. Interaction matters: The effect of a social partner on neural processing of speech	Rice, K., Redcay, E.	2013	Table S2
9. Common brain areas engaged in false belief reasoning and visual perspective taking: A meta-analysis of functional brain imaging studies	Schurtz, M. et al.	2013	Table 2
10. Imagining triadic interactions simultaneously activates mirror and mentalizing systems	Trapp, K. et al.	2014	Table 2, conjunction triad/self/other
11. The dorsal medial prefrontal cortex responds preferentially to social interactions during natural viewing	Wagner, D. et al.	2016	Table 1

Table A.1. The studies and review papers that were used to choose the fNIRS cap geometry for our study, arranged by the functional network for visual working memory, word learning and social interaction.

Appendix B. – Regions of Interest

Region of Interest	Hemi	MNI Centre of Mass		
		x	y	z
Prefrontal Cortex (PFC)	Left	-9	60	21
		-3.6	55	21
		-3	52	26
		-16	39	40
	Right	6	44	20
		12	56	28
Dorsolateral PFC	Left	-42	41	20
		33	29	55
Inferior Frontal Gyrus (IFG)	Left	-52	34	-6
		-50	30	7
		-48	30	9
		-31	23	-1
		-51	17	20
		-46	15	32
	Right	41	41	1
		52	36	17
		36	28	-8
		35	22	4
Anterior Cingulate Cortex (ACC)	Left	-2	29	27
Superior Frontal Gyrus (SFG)		-47	25	-8
		-2	16	52
	Right	12	30	62
		30	2	64
Middle Frontal Gyrus (MFG)	Left	-42	22	46
		-52	14	-4
		-43	11	33
		-40	11	51
	Right	53	21	26
		42	4	46

Middle Temporal Gyrus (MTG)	Left	-54	-26	-10
Superior Temporal Gyrus (STG)	Left	-42	-30	8
		-62	-48	18
		-54	-47	7
Superior Temporal Sulcus (STS)	Right	58	-45	16
Temporoparietal Junction (TPJ)	Left	-58	-49	28
		-44	-63	35
		-57	-64	40
	Right	48	-44	42
		62	-45	23
		58	-54	35
		55	-56	31
Inferior Parietal Sulcus (IPS)	Left	-38	-43	44
		-21	-65	46
	Right	40	-36	38
Inferior Parietal Lobe (IPL)	Left	-55	-52	22
Superior Parietal Lobe (SPL)	Left	-22	-66	56

Table A.2. The final selection of regions of interest.

Supplementary References

Canessa, N., Alemanno, F., Riva, F., Zani, A., Proverbio, A. M., Mannara, N., Mannara, N., Perani, D. Cappa, S. F. (2012). The neural bases of social intention understanding: the role of interaction goals. *PLoS One*, *7*(7), e42347. https://doi.org/10.1371/journal.pone.0042347

Caruana, N., Brock, J., & Woolgar, A. (2015). A frontotemporoparietal network common to initiating and responding to joint attention bids. *NeuroImage*, *108*, 34–46. https://doi.org/10.1016/j.neuroimage.2014.12.041

Gweon, H., Dodell-Feder, D., Bedny, M., & Saxe, R. (2012). Theory of mind performance in children correlates with functional specialization of a brain region for thinking about thoughts. *Child Development*, *83*(6), 1853–1868. https://doi.org/10.1111/j.1467-8624.2012.01829.x

Lau, E. F., Gramfort, A., Hämäläinen, M. S., & Kuperberg, G. R. (2013). Automatic semantic facilitation in anterior temporal cortex revealed through multimodal neuroimaging. *Journal of Neuroscience*, *33*(43), 17174–17181. https://doi.org/10.1523/JNEUROSCI.1018-13.2013

Le Bouc, R., & Pessiglione, M. (2013). Imaging social motivation: distinct brain mechanisms drive effort production during collaboration versus competition. *Journal of Neuroscience*, *33*(40), 15894–15902. https://doi.org/10.1523/JNEUROSCI.0143-13.2013

O'Donnell, M. B., Falk, E. B., & Lieberman, M. D. (2015). Social in, social out: How the brain responds to social language with more social language. *Communication Monographs*, *82*(1), 31–63. https://doi.org/10.1080/03637751.2014.990472

Oberwelland, E., Schilbach, L., Barisic, I., Krall, S. C., Vogeley, K., Fink, G. R., Herpertz-Dahlmann, B., Konrad, K. & Schulte-Rüther, M. (2016). Look into my eyes: Investigating joint attention using interactive eye-tracking and fMRI in a developmental sample. *NeuroImage*, *130*, 248–260. https://doi.org/10.1016/j.neuroimage.2016.02.026

Rodd, J. M., Vitello, S., Woollams, A. M., & Adank, P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: an activation likelihood estimation meta-analysis. *Brain and Language*, *141*, 89–102. https://doi.org/10.1016/j.bandl.2014.11.012

Schurz, M., Aichhorn, M., Martin, A., & Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. *Frontiers in Human Neuroscience*, *7*, 712. https://doi.org/10.3389/fnhum.2013.00712

Szaflarski, J. P., Altaye, M., Rajagopal, A., Eaton, K., Meng, X., Plante, E., & Holland, S. K. (2012). A 10-year longitudinal fMRI study of narrative comprehension in children and adolescents. *NeuroImage*, *63*(3), 1188–1195. https://doi.org/10.1016/j.neuroimage.2012.08.049

Trapp, K., Spengler, S., Wüstenberg, T., Wiers, C. E., Busch, N. A., & Bermpohl, F. (2014). Imagining triadic interactions simultaneously activates mirror and mentalizing systems. *NeuroImage*, *98*, 314–323. https://doi.org/10.1016/j.neuroimage.2014.05.003

Wagner, D. D., Kelley, W. M., Haxby, J. V, & Heatherton, T. F. (2016). The dorsal medial prefrontal cortex responds preferentially to social interactions during natural viewing. *Journal of Neuroscience*, *36*(26), 6917–6925. https://doi.org/10.1523/JNEUROSCI.4220-15.2016

Appendix C. -- Comprehension Test Examples for Three and Five Objects