Mechanisms underlying pre-school children's syntactic, morphophonological and referential processing during language production.

Laura Lindsay

A thesis submitted in fulfilment of requirements for the degree of Doctor of Philosophy

to

Department of Psychology
School of Philosophy, Psychology, and Language Sciences, The University of Edinburgh

2021

Declaration

I hereby declare:

(a) that this thesis is of my own composition, and

(b) that the work reported in this thesis has been carried out by myself, except where acknowledgement of the work of others is made in text, and

(c) that the work has not been submitted for any other degree or professional qualification except as specified in text, and

(d) that the included publications are my own composition.

The following chapter of this thesis is based on manuscripts that have been submitted to peer-reviewed journals:

- Chapter 4: Lindsay, L, Hopkins Z. L. & Branigan H.P. (submitted). A horse by any other name: Referential alignment as evidence for flexible perspective-taking in pre-schoolers' language use. *Developmental Psychology.* Authorship details: Lindsay designed Exp 2-4, ran the participants (E2-5), coded and analysed the data, and wrote the original manuscript. Branigan designed Exp 1 and 5 as part of undergraduate dissertations and undergraduate students conducted data collection for all of Exp 1 and part of Exp 5. Hopkins and Branigan acted as supervisors, gave feedback on each of these steps, and contributed to the revision of the manuscript.

Laura Lindsay

Abstract

Much work has focused on how children learn the words and grammar of their language, with the emphasis being on how children learn to understand their native language. Little work has actually considered how children learn to speak that language and become active participants within their linguistic community. How do children formulate utterances? The primary aim of this thesis was to explore the morpho-phonological, syntactic and referential mechanisms underlying children's language production. To do this, I used appropriately adapted experimental methods that have been successfully used to investigate the morpho-phonological, syntactic and referential mechanisms underlying adults' language production.

Experiments 1 and 2 investigated the syntactic and morpho-phonological encoding mechanisms underlying pre-schoolers' (and adult controls') production of simple sentences by tracking participants' eye movements as they described arrays of two pictures of objects, one of which was preceded by a subliminal cue. In each experiment, we investigated syntactic encoding by analysing how participants chose their starting points, and investigated morpho-phonological encoding by analysing their eye movements as they named each picture. Overall, these experiments suggested that pre-schoolers' syntactic and morpho-phonological processes are similar to those of adults in important respects.

With respect to syntactic encoding, we found that pre-schoolers' starting points were affected by our subliminal cue manipulation so that they tended to start their utterance with the cued object. These results suggest that children do not necessarily need to develop a structural plan before they begin lexical selection of a character or object name. With respect to morpho-phonological encoding, we found that pre-schoolers looked at each object in the order they mentioned them (i.e., they looked at the first object prior to naming it and shifted

their gaze to the second object just before they began articulating the first object's name), indicating that they were retrieving each word incrementally. However, pre-schoolers had longer gaze durations for their final word than their first word, indicating that it took longer for them to retrieve this word, whereas adults did not show this same effect. This could indicate that pre-schoolers are less efficient than adults when planning upcoming words whilst speaking.

Finally, Experiments 3-8 explored referential mechanisms and examined how experience with a partner's language use influenced children's choice of perspective and associated referential expressions for objects (e.g., horse vs pony). Previous studies in adults have shown that interlocutors will flexibly adopt the same perspective as their partner, and have suggested that there are a range of mechanisms underlying this behaviour. The results from our experiments showed that children were more likely to use the same perspective (and associated choice of name) for an object if their partner had previously used that perspective, even when using that perspective meant overcoming a strong default preferred perspective (e.g., using pony instead of horse). However, children failed to maintain this tendency to reuse their partner's perspective over time and contexts (E8). We argue that these results show that children's choice of perspective and associated referential expressions can be influenced by their partner's language use, but this effect is not long-lasting. Overall, our pattern of results is consistent with a strong influence of underlying priming mechanisms that facilitate lexical representations and associated perspectives during referential communication.

Taken together, the results from our studies suggest that the underlying mechanisms of pre-schoolers' language production (syntactic, morpho-phonological and referential encoding) are strikingly similar in important ways to those that have been found in adults (using very similar paradigms). However, our pattern of results indicates that pre-schoolers' production abilities may be less efficient than adults, and as a result, their behaviour may be similar to that of adults under processing load. As a result, they may be particularly susceptible to influences of context (linguistic or non-linguistic) that facilitate retrieval processes. For example, they will immediately reuse a perspective their partner has previously used, but this effect will decay over time as the activation of their partner's perspective decreases; or they will begin their utterance with a cued object name because the cue directs attention to one object which facilitates retrieval of that name.

Overall, our results provide novel insight into the mechanisms underlying pre-schoolers' online production.

Lay Summary

Learning to speak is a major developmental milestone, but we don't know much about how children develop the ability to speak. In this thesis, we investigated how pre-schoolers produced utterances by adapting methods that previous studies have successfully used to investigate adult production. We specifically focused on how pre-schoolers developed sentences, how they produced words, and how they ensured their partner would understand what they were talking about.

In Study 1, we investigated how pre-schoolers developed sentences, by investigating the factors that influenced the starting point of their utterance (which influences the form their sentence will take). We found that children's starting points were affected by a cue manipulation, suggesting that children can develop their sentences by beginning their utterance with the most accessible object first and build their sentence around that starting point. In Study 2, we investigated the processes underlying pre-schoolers' production of words. Previous adult studies have found a link between where speakers look at and what they talk about, suggesting there is a relationship between a speakers eye movements and their speech output. We measured pre-schoolers eye movements as they described pictures on a screen, and found this same relationship in pre-schoolers, suggesting that children use the same mechanisms to produce words that adults do. Finally, in Studies 3 and 4, we investigated whether children took their partner's perspective into account when they referred to objects by testing whether pre-schoolers would continue to use a perspective (or name) that their partner had previously used. We found that pre-schoolers were able to take their partner's perspective into and account and used their partner's perspective in their own language use, but this effect was not long lasting.

Overall, our results suggest that pre-schoolers' production system is similar to adults.

Acknowledgements

I could not have completed this thesis without the help and support of so many people. Firstly, I would like to thank my supervisors, Prof. Holly Branigan, Prof. Martin Pickering, and Dr. Hugh Rabagliati, for their support, encouragement, and guidance throughout the last four years. Holly, I cannot thank you enough for all you have done for me, for useful theoretical discussions, for always being enthusiastic about my results, and for offering advice on academic issues in general. Hugh, thank you for encouraging me to think about the big picture, to think theoretically and for helping with any stats issues I had. Martin, thank you for all your helpful comments and advice.

Second, I am very grateful to the Economic and Social Research Council for generously offering me a studentship to fund my PhD (grant number ES/N71737J/1), which allowed me to compensate participants and attend many conferences. I am also grateful to the School of Philosophy, Psychology, and Language Sciences at the University of Edinburgh for granting me several Research Support grants, which allowed me to attend CLS 2019 and compensate participants for Experiment 8.

Third, I would like to thank all the parents and children who participated in my experiments. Without you, I would not have been able to complete any of my data collection. I would also like to thank all the undergraduate students I have had the opportunity to work with who have helped with data collection.

I would not have been able to complete this thesis without the support from all my friends and family. To Granny and Grandad, thank you for being on the other end of the phone and talking me through any of my PhD problems. Mum and Emma (otherwise known as the witches of Livingston), thank you for listening to me moan all the time. Ethan and Oliver, thank you for being the best little brothers in the world and for being great participants. Ruth and Emma, thank you for encouraging me to always eat cake; Amy, thank you for getting me to train and reminding me there is a world outside academia; Ross, thanks for your brutal opinions and being my personal therapist.

Finally, to John and Wilson – I cannot put into words how grateful I am to the both of you. Wilson, you have been the best boy ever. You always let me know when it was time to take a rest and go walks instead. You made me remember that the thesis wasn't the centre of my world. You were always there for cuggles and strokes when I was so tired from writing. I can't wait to give you all the attention you deserve now this is all over. We will go for all the biggest walks now. And John, much like Sam helped Frodo get to Mordor to destroy the one ring, you have been there to help me through thick and thin. I could not have written this without you. When I thought I could not go on anymore, you were there to remind me what I was fighting for; when I was so tired, you'd bring me a cup of tea; when I cried, you told me it would be ok; you always knew when it was time to stop and made sure I still had something that resembled a normal life. I cannot wait to continue our lives together: we make an awesome team. And I promise I will never be this soppy ever again.

I can't wait to have a Star Wars marathon now.

Table of Contents

Figures

Tables

1. Literature Review

Learning to speak is a major developmental milestone as it marks the the time when a child becomes an active participant within their linguistic community. But getting to that point is a significant challenge: a typical child will understand more than fifty words by the time they begin producing their first and can parse complex grammatical sentences by the time they begin putting words together as holographic two-word phrases (e.g., Golinkoff, Ma, Song & Hirsh-Pasek, 2013). This gap between comprehension and production development is a paradox: how is it that infants can learn and understand complex linguistic rules and words at such a young age yet fail to use this linguistic knowledge as toddlers when they begin speaking?

One reason for this may lie in the complex processes involved in production. Language production requires the coordination and overlap of a series of complex cognitive processes. Even producing a simple sentence like “a dog is next to the cat” requires speakers to coordinate retrieving the correct words from their lexicon whilst also imposing a linear order on them, so that their utterance is grammatical whilst also conveying their intended message (Levelt, 1989). Complex models have been developed to explain how speakers coordinate these processes during production of words and sentences in real time (e.g., Bock & Levelt, 1994; Chang, Dell & Bock, 2006; Levelt, 1989). But we don’t know how children learn to coordinate these complex processes. This is surprising as considerable progress has been made to explain how children *learn* the structural properties of their language, such as the words and grammar (funnily enough, by focusing on the utterances children produce).

Chapter 1

How do the mechanisms underlying production develop? The limited number of studies investigating the mechanisms underlying children's sentence production have typically relied on speech error data and have failed to investigate how children manage production processes in real time. This thesis attempts to bridge that gap by investigating the syntactic, morpho-phonological and referential mechanisms underlying pre-schoolers language production. To do this, we adapted experimental methods and paradigms that have successfully been used to investigate language production in adults.

In this chapter, we first provide an overview of the fully developed production system. We describe the general consensus of the architecture of the adult production system before going on to discuss how adults plan their sentences and how they retrieve word forms for their sentences. We then consider how adults produce during conversation, specifically focusing on referential communication. We then review the development of the production system and consider the challenges children face when they begin speaking. We then consider how children plan their sentences before and during production and discuss the limited findings related to children's production planning. Finally, we review the findings related to referential communication in children.

1.1. The Fully Developed Production System

Before looking at how children formulate utterances, we first describe the general architecture of the production system. Despite the complexity of production, there is considerable agreement about what the cognitive architecture of the language production system looks like (Goldrick et al., 2014). We use this framework as a way to evaluate how children formulate their own utterances.

Chapter 1

A common caricature is that language production uses the same processes as comprehension but in reverse. However, production is much more difficult than comprehension: Whilst adults typically recognize the words in their native language quickly and automatically, the actual production of those same words takes much longer, suggesting it requires more cognitive effort to speak than to understand. For example, adult listeners begin to direct their gaze to the referent of a spoken word *before* the articulation of that word has been completed (e.g., Altmann & Kamide, 1999), whereas speakers typically take much longer (approx. 600ms) to produce a single word in picture-naming studies (e.g., Levelt & Indefrey, 2004).

Language production involves a series of complex cognitive processes that transform a non-linguistic representation into a signed, spoken or written output. Most theories agree that production involves at least three stages (Figure 1; Bock & Levelt, 1994): Conceptualization, where speakers decide on the message they want to communicate; formulation, where speakers decide how they are going to express their message; and articulation, where speakers plan and make the articulatory movements necessary to articulate their message (e.g., Bock & Levelt, 1994; Levelt, 1989).

During conceptualisation (or message-encoding), speakers generate a pre-verbal message (Levelt, 1989; also known as interfacing representations; Bock, 1982). This is encoded into different meanings to create an event; these represent the “who did what to whom” of the message and can then be used during grammatical encoding to select and retrieve the best words and structures to convey the message. As you can see from Figure 1, different meanings are encoded during message encoding: Semantic meanings represent the *who*, the *what* and the *whom* themselves; Relational meanings represent how the who, what and whom *relate* to each other; and finally perspective meanings represent which aspects are more or less important.

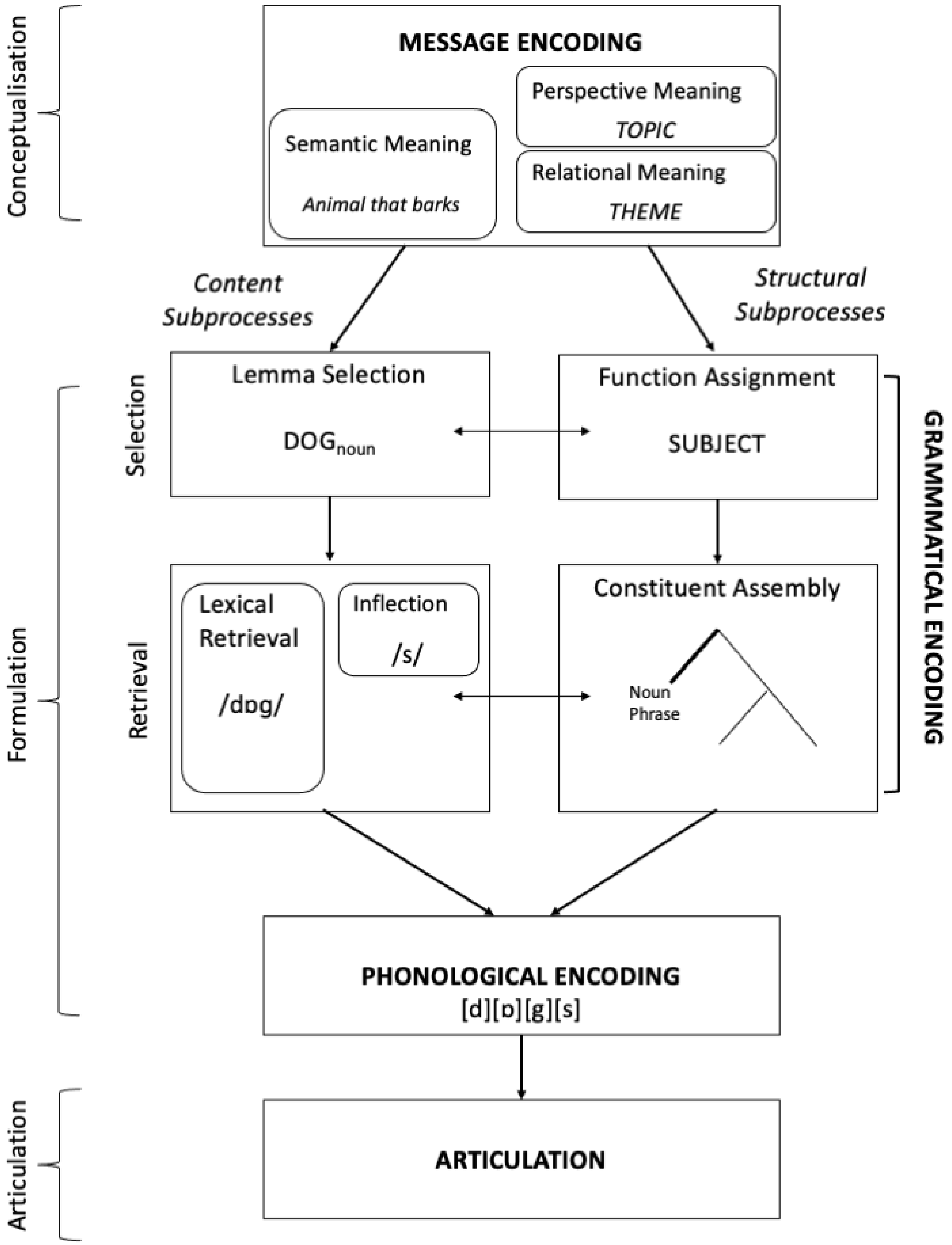

Figure 1. An overview of the general framework for language production processes for retrieving a word (dog) for production in the utterance "dogs are my favourite animal" (Bock

& Levelt, 1994). Arrowheads mark the direction of the primary flow of information as agreed in most models.

For example, to produce the utterance "dogs are my favourite animal," the semantic meanings are an animal that barks ("dog") and the state meaning is that they are my favourite animal. The relational meaning involves attributing the state corresponding to my favourite animal and dogs. And finally, the perspective meaning encodes that dog is the topic of the utterance.

Of central issue to us is formulation. It is typically agreed upon that formulation involves two processes: Grammatical encoding whereby speakers map their intended message onto particular word forms and syntactic structures, and morpho-phonological encoding whereby speakers find the sounds associated with each word and put them together. It is generally agreed that grammatical encoding comprises of two component sets of subprocesses: one that deals with formulating the content (i.e., the words) of an utterance and another that specifies the structure of that utterance. Each subprocess proceeds through two stages (also known as functional and positional processing; Bock & Warren, 1985; Bock & Levelt, 1994; Garrett, 1975), whereby linguistic features are selected and then the properties of those features are retrieved. We discuss each of these subprocesses in turn.

Content subprocesses are generally thought to be responsible for selecting and retrieving the meaning-carrying components – or content words – of a message. The first stage in content processing is lexical selection (Bock & Warren, 1985; Levelt & Bock, 1994). Here, the speaker selects the most appropriate lexical entry – or lemma – that conveys the semantic meaning of the intended message. Lemmas include both semantic and syntactic information, such as its class (e.g., noun, verb) and grammatical gender but crucially do not include phonological information (Levelt, Roelofs & Meyer, 2000; Meyer, 1996).

Phonological information is only accessed and retrieved during lexical retrieval, whereby a speaker accesses the word form – or lexeme – from the mental lexicon. For example, when producing "dogs are my favourite animal," speakers will first select the lemma DOG to describe the message *animal that barks*, before then retrieving the word form /dog/ and adding inflections, such as the suffix "-s."

Similar to content subprocesses, structural subprocesses are responsible for selecting and retrieving syntactic representations necessary to convey the relational and perspective meanings generated during message encoding. The first stage in structural processing is function assignment. Here, the speaker uses the relational meaning to select the grammatical functions for each word – such as subject, direct object, or indirect object, – that best covers the relational meaning in the preverbal message. Once functions have been assigned, the speaker then moves to the next stage – known as constituent assembly - whereby they build the syntactic structure of their utterance. This involves retrieving the constituent structures (e.g., noun phrases) that expresses the grammatical functions selected at function assignment. Again, looking at our example, speakers will assign *dogs* as the subject. As English is a fixed word order language, the subject is typically a noun phrase, and so at constituent assembly, a noun phrase is accessed.

By the end of grammatical encoding, a speaker has a linguistic representation which can be phonologically encoded, whereby speakers select, retrieve, and organise the sounds associated with the planned word (or words).

The overall architecture of the production system relies on a series of complex processes that work together to produce utterances that convey the message the speaker planned to say. In the section that follows, we introduce the incrementality hypothesis, which assumes that speakers do not plan their whole utterance prior to speaking and instead plan small piecemeal chunks at a time. This means that speakers are articulating earlier parts (e.g., the first word) of their utterance whilst still planning later parts (e.g., the final word). This idea has important implications for the development of the production system, as it assumes that speakers can coordinate the sequence and timing of multiple production processes as well as use multiple production processes simultaneously.

1.2. Incremental Planning During Production

One key finding in the production literature is that all stages of adults' language production is incremental (Brown-Schmidt & Konopka, 2014; Brown-Schmidt & Tanenhaus, 2006; Kempen & Hoenkamp, 1987; Levelt, 1989; Griffin & Bock, 2000; Griffin, 2001; Meyer, 2004; Smith & Wheeldon, 1999; see Wheeldon, 2013 for review). Rather than prepare a whole utterance in advance, speakers instead encode small chunks of their utterance whilst simultaneously articulating previously encoded chunks. We now review the findings surrounding incremental planning during sentence planning and lexical retrieval. We then consider the challenges a speaker faces in order to incrementally plan their utterances.

1.2.1. Incrementality during Sentence Planning

In order to incrementally retrieve words, speakers need to decide the order in which those words will appear in their sentence. That is, they must develop a sentence plan that guides the order by which words are incrementally retrieved. A key debate in the sentence planning literature concerns the timing and information speakers use to develop their sentence plan, and specifically, the extent to which content and structural processes interact with each other.

One way researchers have investigated sentence planning is to look at the factors that influence their starting point. The starting point of a sentence constrains both the form and content for the remainder of that utterance (MacWhinney, 1977). That is, it guides the word order for simple phrases like conjoined noun phrases (e.g., *the cat and the dog*), but also constrains syntactic choice when describing more complex scenes, such as transitive events. For example, when describing a picture of a cat chasing a dog, I can choose to begin my description with "the cat." By doing so, I must then access the active construction of the verb "chase" and end my utterance talking about "the dog" (thus producing "*the cat is chasing the dog*").

The structural incrementality account assumes that a speaker develops their sentence plan first before they begin retrieving the words to convey their intended message. Thus, syntactic planning occurs before content planning. This account emphasises top-down control of sentence planning and assumes that word order is determined by a higher-order conceptualisation of an event's semantic relations (i.e., determining who is doing what to whom; Griffin & Bock, 2000). This kind of strategy is consistent with language production models assuming that syntactic structure is generated based on the thematic structure of a scene without reference to lexical representations (e.g., Chang, Dell & Bock, 2006). It argues that a 'gist extraction' mechanism assigns thematic roles to each character and creates an overall structure for the sentence plan, which is imposed before speaking. This plan then guides subsequent eye movements to the first mentioned object (i.e., the starting point) before moving to the next named object. This means that speakers do not begin lexical retrieval until after a sentence plan has been generated (normally after the first 400 ms; e.g., Griffin & Bock, 2000). This account assumes that starting points are influenced by a speaker's implicit biases, such as a preference for agent subjects.

However, the bottom-up lexical incrementality account assumes that syntactic choice and lexical choice interact with each other to determine the form of the utterance. That is, the speaker activates the chosen syntactic structures based on the most active lexical item. For example, take the following sentences:

(1) John gave Wilson the ball

(2) John gave the ball to Wilson

Both (1) and (2) have similar meanings but have different syntactic structures. Thus, when planning an utterance, a speaker's syntactic choice can be determined by whatever word is accessed or activated first. So, if *Wilson* is accessed before *ball,* the speaker will produce an utterance like (1). Yet, if they access *ball* before *Wilson*, they will produce an utterance like (2). And in fact, when speakers have more syntactic choice when planning their utterance, they are faster to begin speaking and produce fewer errors than when their choice is restricted (Ferreira, 1996). Indeed, the lexical incrementality account assumes speakers will begin talking about the most salient or accessible character first and build a sentence plan that is consistent with this starting point (e.g., Gleitman, January, Nappa & Truesewell., 2007; Tomlin, 1997). This kind of strategy is consistent with language production models assuming that visual information immediately activates lexical information prior to the generation of a syntactic structure (e.g., Bock, 1982; Levelt & Bock, 1994). Thus, speakers begin lexical retrieval as soon as they fixate on an item or character, and, as a result, their starting point is not only predicted by their first fixations but can also be influenced by attention manipulations designed to affect these. For example, when describing a picture of a dog chasing a cat, if the speaker fixates on the cat first, they will begin lexical retrieval of the word form *cat*. Upon retrieving the word *cat*, they must then encode the agent of that event as being the dog and then access a passive structure to produce the utterance "*the cat is being chased by the dog*." Overall, this suggests that speakers are not set on the exact form their utterance will take prior to speaking. However, this interaction between syntax and lexis requires some cognitive flexibility: speakers need to be able activate different syntactic structures based on the activation of different lexical items.

Adults can use both sentence-planning strategies when developing their utterance. Much work has shown that adult speakers extract the overall gist before speaking (Bock, Irwin & Davidson, 2004; Griffin & Bock, 2000; Kuchinsky, Bock & Irwin, 2011), but they can switch planning strategy in cases where an event is hard to interpret (van de Velde, Meyer & Konopka, 2014), or after a character has been lexically primed (Konopka & Meyer, 2014). The ability to use both these strategies allows adults to speak fluently, striking a balance between communicating the information that they consider most noteworthy, and that which they can most easily encode.

The structural incrementality and lexical incrementality accounts place different cognitive demands on the speaker at different time windows during planning. The structural incrementality account assumes speakers move from message-encoding processes to syntactic processes to lexical retrieval processes in a serial fashion. That is, a speaker must develop a sentence plan before speaking and maintain this overall structure in working memory, thus increasing load in working memory at the *beginning* of utterance planning. However, the lexical incrementality account assumes message-encoding, syntactic choice, and lexical processes are parallelised to some degree: how the speaker conceptualises an event is affected by the character they began lexical retrieval on first. Thus, this account requires *less* processing resources at the beginning of utterance planning because the speaker only has to focus on lexical retrieval of the first named character or object. However, this strategy also requires the speaker to continue their utterance in such a way that is compatible with their chosen starting point (i.e. talking about the cat first means they have to access either the passive structure or choose a different verb for describing the scene), thus relying on cognitive flexibility in the speaker. As a result, the lexical incrementality strategy demands more processing resources *during* speech, as the speaker must continuously engage in planning both conceptual and linguistic increments *after* speech onset.

1.2.2. Incrementality during Lexical Retrieval

Many eye-tracking studies have shown there is a tight link between a speaker's eye gaze and their speech output Gleitman et al., 2007; Griffin & Bock, 2000; Griffin, 2001; Meyer, Sleiderink & Levelt, 1998; van der Meulen, Meyer & Levelt, 2001; see Meyer, 2004 for review). Eye-tracking studies have shown that when naming a series of objects as "*a dog is next to a cat*", speakers tend to shift their gaze to each new referent a few hundred milliseconds before labelling it (e.g., Meyer et al., 1998; see Meyer, 2004 for review). That is, they will look first to the dog in a scene prior to naming it and then look towards the cat. This suggests that they are using the visual scene to aid retrieval for forthcoming words whilst also articulating previously retrieved words at the same time.

Indeed, the time speakers fixate on an object depends not only on how long it takes for them to identify the object (i.e., visual-conceptual processing; Griffin & Oppenheimer, 2006), but also on the time it takes for them to select the appropriate name (or lemma) and retrieve and generate the phonological word form (Belke & Meyer, 2007; Griffin, 2001; Meyer & van der Meulen, 2000; see also Meyer, Belke, Telling & Humphreys, 2007). For example, Meyer et al (1998) found that speakers' gaze durations were longer for degraded objects than intact objects, but also for low frequency names than high frequency names. Similarly, speakers tended to look longer at objects with long names than at objects with short names (Meyer, Roelofs & Levelt, 2003; Roelofs, 2007; but see Griffin, 2003). Overall, these findings suggest that speakers' gaze remains on an object until they have completed lexical retrieval for that object, and they will only shift their gaze to the next object once lexical retrieval for the previous object has been completed (see also Roelofs, 2007, 2008).

Why do speakers continue fixating on an object for so long? Various accounts have been proposed to explain speakers' late gaze shifts (see Griffin, 2004 for review). Lexical access requires processing resources (e.g., Cook & Meyer, 2008; Ferreira & Pashler, 2002; Roelofs, 2008) and so, under one account, long gaze durations may facilitate not only object identification but also lexical retrieval through a process of spreading activation (e.g., Roelofs, 1992; Wühr & Waszak, 2003; Wühr & Frings, 2008). That is, the conceptual and lexical representations of the fixated-on-object will receive greater activation (thus inhibiting other words), which then facilitates lemma selection and word form retrieval. Furthermore, long gaze durations may reduce the interference of the names of other objects in a scene (e.g., Meyer et al., 1998).

However, whilst speakers may fixate on one object – thus primarily retrieving that object name – they can also engage in secondary, extra-foveal processing of the next object name. That is, they can begin retrieving the next object's name whilst they are still retrieving the first object's name (i.e., they are engaging in parallel processing). For example, when producing "*a dog is next to a cat,*" speakers must access two different content words *dog* and *cat*. Thus, later stages of lexical access (e.g., phonetic encoding) would be occurring for the first word (*cat*) while earlier stages of lexical access (e.g., conceptual processes) will be occurring for the second word (*dog*). Such parallel planning serves fluency as if a speaker encodes each next word from scratch after completing the current word's articulation, then speech would appear disconnected and disfluent. Indeed, in typical multiple-object naming studies, adult speakers can still identify and process objects extrafoveally (meaning that they don't necessarily have to directly focus their attention on an object to process it; Morgan &

Meyer, 2005; Meyer, Ouellet & Hacker, 2008), suggesting there is some temporal overlap in the processing of successive objects' names. However, the ability to retrieve names of the fixated-on and extrafoveal objects in parallel is dependent on the difficulty of processing the fixated-on object (Malpass & Meyer, 2010), which implies that processing load may be relevant to parallel planning.

1.2.3. Challenges of Incremental Planning

Incremental planning is advantageous to the speaker for several reasons. Firstly, it involves planning smaller units of an utterance, which means speakers can begin articulating chunks of an utterance as soon as they have been formulated (Levelt, 1989). This means speakers can begin speaking quickly, which is especially advantageous during dialogue, as it ensures short turn-intervals (Corps et al., 2018; Corps et al., 2020). Secondly, it reduces the need for speakers to buffer units of their utterance in working memory, thus uses less processing resources (Ferreira, 1991; Levelt, 1989). Finally, incremental planning allows the speaker to continuously update their utterance, which may be prompted by changes in their communicative intent or listener feedback (Ferreira, 1996; Brown-Schmidt & Konopka, 2015).

However, incremental planning also poses various challenges and requirements of a speaker. Firstly, incremental planning requires processing resources: Speakers need to keep track of their message as well as use production processes concurrently. Secondly, it requires cognitive flexibility: Speakers must carefully balance how much of an utterance they plan in advance of production. If planning lags behind articulation speakers may produce more errors or disfluencies (e.g., repetitions and/or producing fillers like *uhm*) whereas if encoding is faster than articulation, it may result in greater demands to working memory as the production system has to buffer future articulatory plans. In the sections that follow, we discuss each of these challenges in turn.

1.2.3.1 Incremental Planning Requires Cognitive Flexibility

One central concern in the language production literature is determining how far in advance speakers plan their utterance (i.e., what are the planning units) and whether advance planning differs at different representational levels (e.g., Allum & Wheeldon, 2007, 2009; Brown-Schmidt & Tanenhaus, 2006; Costa & Caramazza, 2002; Damian & Dumay, 2007; Dell & O'Seadhgha, 1992; F. Ferreira, 1991; Ford & Holmes, 1978; Garrett, 1980; Griffin & Bock, 2000; Lindsley, 1975; Martin, Miller, & Vu, 2004; Meyer, 1996; Oppermann, Jescheniak, & Schriefers, 2008; Schnur, Costa, & Caramazza, 2006; Schriefers, Teruel, & Meinshausen, 1998; Smith & Wheeldon, 1999, 2004; Wheeldon & Lahiri, 2002). The radical incrementality approach proposes that speakers only need to plan one word at a time (e.g., Griffin 2003; 2004; see also Chang, Dell & Bock, 2006) whereas the advanced planning approach argues that speakers must plan at least the first phrasal unit of an utterance before articulation (e.g., Smith & Wheeldon, 1999; see also Garrett, 1980, for evidence that a clause is the planning unit).

Evidence for a tightly incremental word-by-word approach comes from eye-tracking studies where speakers name pictures of objects (e.g., "the A and B are above the C," Griffin & Bock 2000; Griffin, 2001). In these studies, speakers tend to fixate on the objects in order of mention and shift their gaze to the next object prior to articulation (e.g., Griffin & Bock, 2000; Griffin & Speiler, 2006; Gleitman, January, Nappa & Trueswell, 2007; see Meyer, 2004 for review). If participants plan more than one word at a time, the time delay between gazing and speaking should be shorter for words occurring later in the utterance. Instead, the delay between fixation and onset of articulation is similar, irrespective of a word's position in an utterance (Griffin & Bock, 2000). Furthermore, Griffin (2001) showed that a speaker's initiation of the first word in an utterance isn't affected by either the frequency or codeability (the number of alternative names an object has) of the second word in an utterance, suggesting that no planning of the second object's name occurs before articulation of the first word.

However, other studies suggest that speakers can also plan at least the first phrase of their utterance before articulation (e.g., Allum & Wheeldon 2007; Damian & Dumay, 2007; F. Ferreira, 1991; Martin et al., 2010; Meyer, 1996; Smith & Wheeldon, 1999; Wheeldon, 2013). For example, Smith and Wheeldon (1999) showed that response latencies were longer when participants produced sentences beginning with a complex phrase (e.g., *"the dog and the kite moves above the house"*) compared to a simple one (e.g., *"the dog moves above the kite and the house"*). As the overall utterance length was controlled, this result suggests that more processing time was required for producing the first phrase because it was more difficult and thus provides evidence for a phrasal planning scope (see also Allum & Wheeldon, 2007; 2009).

Studies using the picture-word interference paradigm (e.g., Schriefers, Meyer & Levelt, 1991) provide further support for an advanced planning scope (e.g., Meyer, 1996; Smith & Wheeldon, 2004; Schnurr et al., 2006). For example, Meyer (1996) presented participants with pairs of pictures, which they named using either noun-phrase conjunctions (e.g., "the arrow and the bag") or locative sentences ("the arrow is next to the bag"). Whilst speakers planned their utterance, they were presented with an auditory distractor which was either semantically-related, phonologically-related or unrelated to the first or second noun. She found that speakers took longer to begin speaking when the distractor was semantically-related to either noun (the first or second one) compared to an unrelated distractor for both noun-phrase conjunctions and locative sentences. This suggests that both target lemmas were accessed prior to speaking – even though in locative phrases, each word is in a different phrase. Thus, this study suggests that lexical selection has a clausal planning scope. However, Meyer (1996) only found a phonological facilitation effect for the first word. That is, participants were faster to begin speaking when the distractor word was phonologically-related to the first word only compared to an unrelated distractor. This suggests that speakers had only retrieved the word form (or lexeme) of the first word and not the second word. Overall, this study suggests that there are different scopes of planning at different levels of grammatical encoding: At lexical selection, where speakers select the lemmas of their utterance, speakers have a clausal planning scope, yet at the level of lexical retrieval, speakers only retrieve one word form at a time (see also Wheeldon, Ohlson, Ashby & Gator, 2013).

The studies discussed so far provide an inconsistent picture. On the one hand, there is evidence to suggest that speakers plan a word at a time, yet other studies suggest that the planning unit is larger. This discrepancy of results has led some researchers to question whether a speaker's planning scope is in fact a rigid planning unit and in fact, more recent studies suggest that speakers are flexible in the amount of advanced planning they do, in a context-dependent way. For example, speakers can alter how much of an utterance they plan in advance depending on variations of time pressure (Ferreira & Swets, 2002; Swets, Jacovina & Gerrig, 2013), cognitive load (Wagner, Jescheniak & Schriefers, 2010), working memory load (Slevc, 2011) or even experience with certain sentence structures (Konopka, 2012).

This suggests that speakers use different strategies to plan their response depending on the demands of the task (Swets et al., 2013; Van de Velde et al., 2014). That is, the scope of planning is flexible and speakers can make use of multiple sources when planning their utterance. The ability to flexibly alter the scope of planning is an especially useful aspect of the production system, as it means a speaker can adjust when and how much they plan according to task demands. For example, during conversation, a speaker may not be able to begin planning until later in their utterance and so it would make sense to plan less of an utterance in advance as it would mean the speaker would have less complex an utterance to begin planning (e.g., Boiteau et al., 2014; Sjerps & Meyer, 2015). Yet, in cases where speakers can anticipate how a turn will end, they could plan more of their utterance (e.g., Bögels, Magyari & Levinson, 2015; Corps et al., 2018). Flexibility in advance planning may also be a useful strategy whilst speaking to ensure that speakers balance between planning too much of their utterance and planning too little. However, adapting the scope of advanced planning to task demands requires cognitive flexibility and some sort of meta-awareness of how difficult a task is going to be.

1.2.3.2 Incremental Planning Requires Processing Resources

As an adult, speaking feels "easy" to do. What makes speaking feel intuitively easy is that it is a highly automatized process and as a result we are not consciously aware of the amount of mental effort required for production. Yet, humans have a limited mental capacity. That is, there are limits to the number of processes a person can complete at any given time (Just & Carpenter, 1992; Kail & Salthouse, 1994). As a result, there are only so many resources an individual can bring to a certain task: Performance in any task depends on the relationship between costs and available resources (Just & Carpenter, 1992).

Indeed, we can see that a speaker's performance during production can be affected by the processing demands of a task. For example, speech errors and disfluencies, such as pauses, repairs or rephrasing of utterances, are likely due to the cognitive demands of utterance planning (e.g., Clark & Wasow, 1998; Goldman-Eisler, 1968; Maclay & Osgood, 1959; see also Kubose, Bock, Dell, Garnsey, Kramer & Mayhugh.,2006; Strayer & Drews, 2007). And in fact, researchers have shown there are capacity limits to production by using dual-task paradigms, whereby participants are asked to speak whilst completing a secondary task. These studies have shown that language performance interferes with performance in the secondary task (Almor, 2008; Cook & Meyer, 2008; Ferreira & Pashler, 2002; Kunar, Carter, Cohen & Horowitz, 2008; Sjerps & Meyer, 2015; see Roelefs & Piai, 2011 for review on word naming). For example, Almor (2008) measured participants' performance on a vision task whilst participants either listened, spoke, or planned to speak. They found that performance on the vision task was slowed down when participants were speaking or when they were planning to speak, suggesting that the processes involved in production require more processing resources than listening does. Similarly, the secondary task can also affect language performance, with participants producing shorter utterances, more disfluencies, and more grammatical errors when they complete a secondary task as well compared to when they do not (Hartsuiker & Barkhuysen, 2006).

Overall, these studies show that there are capacity limits to production – even in adults who are experts in their native language. Whilst planning incrementally reduces the need for the speaker to buffer their whole utterance in working memory, it still requires processing resources as speakers need to coordinate the sequence and timing of production processes in real time, but also have them running in parallel, with each activity working on different parts of the utterance to ensure that their results become available when the speaker requires them (Dell, Burger & Svec, 1997; Levelt, Roelofs & Meyer, 1999). Thus, a speaker might begin producing the initial parts of their message as soon as part of it has been fully planned, whilst continuing to plan later parts of the message (Brown-Schmidt & Konopka, 2015), or they can begin articulating the first word in a sentence whilst still retrieving the final one (Meyer, 1996). This means that speakers are concurrently planning their message, activating and retrieving lexical items, activating and retrieving syntactic structures, accessing the sounds associated with each word, and speaking. That is, they are performing multiple language processing activities simultaneously.

From a limited capacity perspective, parallel processing like this should be labour intensive and cognitively demanding. As such, you might expect to find greater demands on the processer when there are multiple processes happening simultaneously – such as at the beginning of an utterance whereby speakers are encoding the message, accessing words and syntactic structures and sounds (remember that one assumption of incremental planning is unidirectional processing). However, towards the end of an utterance, where utterance planning is slowly coming to an end and speakers are typically accessing final words and sounds, then there are more resources available. And indeed, dual task studies suggest that this is the case. For example, Sjerps and Meyer (2015) used a dual task paradigm to determine when speakers began preparing their response during conversation. Participants were instructed to complete a finger tapping exercise whilst taking turns describing pictures with an interlocutor. Unsurprisingly, speakers' performance in the finger tapping exercise decreased as they began preparing and initialising their response. However, as their turn came to an end (i.e., later parts of the utterance), performance started to increase. This could suggest that as less processes are happening in parallel towards the end of their utterance, participants had more processing resources available to them, resulting in better performance in the secondary task. Likewise, Power (1986) found that speakers who were completing a tracing task produced more errors whilst speaking over the first clause of two-clause sentences, then decreased over the second clause, again suggesting that more processing resources were required at earlier parts of utterance planning, when more processes are required to work in parallel (see also Clark & Wasow, 1998).

We can also make assumptions about how parallel planning will affect speakers' subsequent utterances. For example, if a speaker commits resources to a more complex operation (e.g., by accessing a more complex syntactic structure), resources may be limited for other aspects of the utterance. The speaker may reduce or sacrifice processing elsewhere, which may result in them omitting certain pieces of information or reducing the overall complexity of their utterance (Bock, 1982; 1995; Just & Carpenter, 1992).

1.3. Referential Communication

When we speak, we tend to do so with a conversation partner. For communication to be successful, speakers must ensure they can be understood. One way speakers can make themselves understood is through their referential choice (i.e., how they choose to describe a referent). When describing an object, speakers must take a certain perspective on that object to ensure their partner will understand them. But we can think about objects in different ways depending on the context, and these different perspectives are reflected in our referential choices. For example, when a speaker describes her pet dog, she can describe him as *my dog*, *my pet*, *my little sausage*, or *Wilson*, depending on their conversational partner and other aspects of the context (e.g., other potential referents).

Adult speakers flexibly adopt different perspectives when conversing, as reflected in their referential choices. For example, they refer to the same referent using a basic-level term (*dog*) or a more specific subordinate (*labrador*) depending on whether other category exemplars are present (Brennan & Clark, 1996), and using a proper name (*UN building)* or a description (*the building with all the flags)* depending on their partner's presumed knowledge (Brennan & Clark, 1996; Isaacs & Clark, 1987). Critically for our interest, they are also responsive to their conversational partners' perspectives, so that they show a robust tendency to spontaneously adopt the perspective, and hence referring expression, that a partner has previously used (e.g., using *pony* after their partner has previously used *pony*). That is, adult speakers tend to *referentially align* (or entrain) with a partner's perspective, and so maintain the *referential precedent* that a partner has set (and adult comprehenders correspondingly expect speakers to maintain precedents once established; Metzing & Brennan, 2003; Shintel & Keysar, 2007). This tendency occurs both for novel objects without conventional names (e.g., *floor* vs. *row* for a configuration of points; Garrod & Anderson, 1987) and familiar objects with conventional names (e.g., *basket* vs. *hamper;* Branigan, Pickering, Pearson, McLean, & Brown, 2011). Importantly, speakers align with a partner's perspective even when this entails adopting a perspective that they would not usually favour (e.g., using *pony* where they would normally strongly favor using *horse;* Branigan et al., 2011; Tobar, Rabagliati, & Branigan, 2020). Such alignment is associated with successful and satisfying communication (Fusaroli et al., 2012).

1.3.1. Accounts Underlying Referential Alignment

Adults' referential alignment has been attributed to a range of underlying mechanisms which importantly may not be mutually exclusive (Branigan et al., 2011). Note that each account assumes that experience with another speaker plays a different role in the reasons why speakers align (or entrain) with their partners.

The collaborative account assumes referential alignment is driven by speakers' motivation to be understood by their addressee (i.e. audience design; Clark, 1992; 1996). According to this account, speakers consider what their addressee is likely to know when deciding how to refer to an object. For example, these judgments can be based on either the speaker's beliefs about the cultural community their interlocutor belongs to and the kinds of information people from that community are likely to know (e.g. Isaacs & Clark, 1987) or their prior experience with that interlocutor (e.g. Clark & Marshall, 1981; Galati & Brennan, 2010; Wilkes-Gibbs & Clark, 1992). For example, if a speaker had previously heard her addressee use the term *pony* to describe a horse, thus creating a referential pact, the speaker is likely to reuse this term in her own productions as she can be sure her addressee will understand it. The collaborative account implicates an important role for theory of mind (ToM) – the ability to attribute mental states to others and to understand that others have mental states different from our own (Wimmer & Perner, 1983). Thus, speakers must model their own knowledge against that of the listeners when planning their utterance. This account assumes that experience with a partner means you'll model their knowledge on to your own, and as a result, you'll use the referential choice as them. Thus, looking at referential choice in children is interesting because it assumes perspective-taking mechanisms (e.g., theory of mind) need to have developed before children will even align with their partner

However, the memory-based account proposes that the collaborative approach is computationally burdensome, and instead suggests that referential alignment is due to domain-general memory processes (Horton & Gerrig, 2005; Horton, 2007; cf. Horton & Gerrig, 2016). Referential expressions are remembered and associated with different partners and contexts. Thus, the context (e.g., partner) acts as a memory cue. As a result, when a speaker talks about the same object with the same partner, the referential expressions used before will be activated and pre-empt the use of other terms. Crucially, such activation is due to ordinary, episodic memory representations and not special-purpose memory representations for common ground. The memory-based account assumes you associate a certain referential choice with a specific partner, and so that activates a specific referential choice. Both this account and collaborative account assume partner-specificity effects.

However, non-goal-directed accounts of alignment, such as the Interactive Alignment Model (IAM; Pickering & Garrod, 2004), assume that alignment is a simple response to an interlocutor's linguistic behaviour (i.e. it is due to a simple priming mechanism). There is evidence that all levels of linguistic processing (lexical, syntactic, semantic, etc.) are strongly influenced by prior processing of linguistic material (e.g., Bock, 1986; Branigan, Pickering & Cleland, 2000; Pickering & Branigan, 1998). For example, exposure to the term *pony* activates the word form /pony/ in the speaker's mental lexicon, which in turn activates the PONY perspective. This inhibits other perspectives (e.g., HORSE), and as such, facilitates subsequent production of their partner's perspective (pony). Thus, referential alignment is the by-product of an automatic, non-goal orientated, priming mechanism. However, such an account cannot explain the partner-specificity of referential pacts. The IAM assumes that experience hearing a word influences the subsequent production of that word. As a result, there would be no partner-specificity effects.

Interim Summary

There is considerable agreement about how the cognitive architecture of production is organised (Goldrick et al., 2014). Speakers must first decide on a message, transform that message into a linguistic representation during grammatical and phonological encoding, and then begin the process of articulating that utterance. Furthermore, it is generally agreed that speakers plan their utterance incrementally (i.e., in a piecemeal fashion), although there is still some debate on exactly how speakers do this. But in order to plan incrementally, speakers must have the have the cognitive flexibility to be able to adapt the planning unit based on task demands as well as processing resources available to have multiple production processes running in parallel.

We also considered how adults produce language during conversation, specifically focusing on how they ensure their partner understands them. Adults tend to reuse the same referential choice that their conversation partner previously used, and this reflects an ability to adopt their partner's perspective. Adults' lexical alignment has been attributed to a range of underlying mechanisms, including higher-order perspective-taking processes and memory processes.

In the sections that follow, we discuss the developing production system.

1.4. The Developing Production System

The stages of production development have generally been well documented. From birth, infants typically coo in response to their caregiver. This kind of turn-taking behaviour, also known as proto-conversation, is thought to play a significant role for infants' early communicative development as well as providing an important context to starting to acquire language (Clark, 2007; S. A. Gelman, 2009; Hirsh-Pasek et al., 2015; Jaffe, Beebe, Feldstein, Crown, & Jasnow, 2001; Murray & Trevarthen, 1986). At around 6-8 months of age, infants begin babbling. This typically begins with singular consonant-vowel productions (e.g., *ba, da, ga*) before progressing on to combinations of CV-CV syllables (e.g., *bababa, dadadada, gagagaga*; Locke, 1986). By the end of their first year, infants begin to produce their first words. Their productive vocabulary is generally quite limited at first, yet goes through a growth spurt during their second year (e.g., Hart & Risley, 1995). This increase also coincides with the time children begin putting words together and produce holistic two-word phrases such as *give juice*. By the age of two, they begin producing multi-word utterances and throughout their third year, the number of syntactic constructions children can produce increases.

Whilst the stages of production development are well documented, little work has investigated exactly how the mechanisms underlying production actually develop. In the sections that follow, we first consider why it is important to investigate the development of the mechanisms underlying production, before going on to discuss the methods that researchers have used to investigate production mechanisms in children. We then consider how children develop the abilities to produce words and put those words together whilst also discussing the challenges children face when they produce words and sentences, before considering the mechanisms underlying incremental production. Finally, we discuss the findings surrounding children's referential communication, and the mechanisms underlying these abilities.

1.4.1. Why study the mechanisms underlying children's production?

Speaking is something that, intuitively, feels easy to do. Yet, as we have seen, it involves a complex set of mental activities. When we speak, we must first access real-world conceptual knowledge to create a message to be expressed, access words from our lexicon that convey our message, assign grammatical roles to those words, and build a syntactic structure which is then translated into a phonological code ready for articulation (Bock & Levelt, 1994; Levelt, 1989). The goal of language production research is to explain how speakers in real time retrieve and assemble elements of language in order to communicate their message and so the issues in language production centre on how the system is organised, how and when the production system retrieves different kinds of lexical knowledge (e.g., accessing lexical items and syntactic structures), how the system coordinates different processes, and how the system is constrained by human cognitive capacities.

Yet these issues have largely been ignored in developmental research (see Charest & Johnston, 2011 for review). Our understanding of how children learn to use and coordinate production processes is limited as most developmental research has focused on *what* children are able to say rather than on *how* children are able to say it. Analyses of children's utterances – and the errors they produce – has given us insights into their linguistic and communicative knowledge in areas such as phonology (e.g., Gerken, 1996), morpho-syntax (e.g., McKee, McDaniel & Snedeker, 1998), and pragmatic ability (e.g., Clark, 2014; Matthews, Lieven & Tomasello, 2010). But this work rarely focuses on the development of the mechanisms underlying children's production. Even when developmental studies use paradigms that are used to investigate production mechanisms in adult psycholinguistic studies, they are often asking questions about children's competence: For example, syntactic priming methods are used to investigate whether children's syntactic representations are abstract or not rather than on how these representations are structured or accessed (e.g., Branigan & McLean, 2016; Rowland, Chang, Ambridge, Pine & Lieven, 2012).

By assuming children's errors or omissions during production reflect gaps in their linguistic knowledge (e.g., Bloom, 1990), this competency-based approach to language development fails to take into account the effects a *developing* production system has on children's utterances. Indeed, pressures on the developing system may in fact lead to errors that children themselves would deem as ungrammatical and so it is necessary to be cautious when interpreting children's production data, as it is the output of a developing performance system and not a direct measure of children's linguistic knowledge.

Investigating the developing production system and the mechanisms underlying production is important in its own right. The production system requires the coordination of multiple processes, which may present significant challenges to children (Charest & Johnston, 2011). First, the resources required for producing may exceed children's processing capacities, especially when they produce more complex utterances. Second, the costs associated with different production components vary independently (i.e., costs associated with syntactic formulation differ from costs associated with lexical retrieval). Finally, the cognitive load from one domain of language production may affect performance in another domain. By neglecting to investigate how the mechanisms underlying production develop, we are unable to establish how children deal with the real-time challenges of constructing an utterance.

1.4.2. Methodological Considerations When Investigating Production Development

Much of what we know about the developing production system comes from observations of spontaneously produced or naturally occurring speech, although more recent studies have begun using experimentally elicited speech. Whilst spontaneously produced speech has the advantage of being generated naturally, and therefore has high ecological validity, it is difficult to know if what was produced is exactly what the child intended to say. However, experimentally elicited speech has the advantage of being produced in a controlled environment, and therefore we know exactly what children were intending to say. From speech observations, we get different kinds of data. Most used in developmental research is the analysis of speech errors and disfluency patterns. These types of data can tell us about different stages of the production process.

Speech errors were often used to understand the architecture of the production system in adults 40 years ago (e.g., Garrett, 1980). For example, stranding errors whereby words do not appear in the correct position in a phrase (e.g., *where wings take dream* when intended to say *where dreams take wing*) suggest that there is a process whereby an abstract template is created and words slot into that template (i.e., constituent assembly). This kind of analysis has also been popular in developmental research (e.g., Fromkin, 1971; Jaegar, 2005; Stemberger, 1989; Wijnen, 1990, 1992). However, it is important to bear in mind that when analysing children’s speech errors, it is difficult to distinguish between errors that adults would also make (and therefore reflect production processes) and errors that reflect limitations in linguistic competence. For example, if a child produces a morphological tense error such as "*I had never connect them,*" it is unclear if this is an error because there was a problem during grammatical encoding, or if it is showing that the child has not yet mastered the past tense (Stemberger, 1989), unless you have other evidence that they know it.

During production, speakers are rarely fluent and instead produce disfluencies throughout their utterance, such as pauses, fillers or repetitions. Studies often look at where these disfluencies appear in a phrase (e.g., within a phrase, before a word, etc.) and these are often thought to reflect processing difficulties, such as problems during lexical retrieval or increased load during sentence planning. Such disfluencies are also interpreted in this way in the child literature (McDaniel, McKee & Garrett, 2010; Rispoli, Hadley & Holt 2008; Wijnen, 1990).

The reliance on speech error data to investigate the developing production system repeats how language production was studied in adult psycholinguistics thirty years ago. Furthermore, the analysis of speech errors and disfluencies can only tell us about processing when the production system goes wrong and fails to tell us about what is happening during successful production. Indeed, speech errors are only ever produced very occasionally and thus do not reflect the output of most of children and adult's spontaneously produced utterances (Garrett, 1985; Jaegar, 2005; Stemberger, 1989; Wijnen, 1990).

In the adult literature, important theoretical advancements were made with the development of experimental paradigms using more robust techniques, such as chronometric (i.e., response times: Smith & Wheeldon, 1999; Meyer, 1996) and eye-tracking methods (Griffin & Bock, 2000; Meyer et al., 1998; see Meyer, 2004 for review). By using these methods, researchers were able to isolate specific aspects of the production process, resulting in the establishment of various models explaining how the production system operates during successful language use. But these methods are rarely applied to developmental work, perhaps because testing young children using these methods is practically challenging to do: the participants are stubborn and may not follow instructions, and the data are too noisy. Indeed, pre-schoolers' reaction times are often slower, show more variance and show a greater skew than adults'. This is problematic for more traditional analyses that focus on differences between means, as high variances can mask shifts in the mean or can increase the chances of statistical tests showing a false positive or false negative. However, recent advancements in statistical techniques mean we can now use these methods to investigate production processes in children (see Lindsay, Gambi & Rabagliati, 2019 for an example using Bayesian distributional regression analyses on children's production data). Furthermore, the use of Bayes Factors can allow us to determine whether effects from statistical models are truly null effects or not.

1.4.2.1. Using Eyetracking to Measure Children's Language Production

One method of particular interest to us is eye-tracking. Eye-tracking methods in adult production studies have revealed how the production system operates during *successful* language use (e.g., Griffin & Bock, 2000; Meyer et al., 1998). These methods have shown that there is a tight link between speakers' eye movements and their speech output, and thus can tell us about the online processes involved in language production (i.e., how are they constructing their utterance in real time). For example, by using eye-tracking methods we can analyse how participants inspect the scene while speaking, revealing the processes of incremental planning, but also how they inspect the scene prior to speaking, revealing strategies that guide their formulation.

Whilst many studies have used eye-tracking paradigms to investigate how children comprehend language online (e.g., Fernald, Pinto, Swingley, Weinberg & McRoberts, 1998;

Mani & Huettig, 2012; Snedeker & Trueswell, 2004), little work has actually gone on to use these paradigms to investigate how children plan and structure their own utterances. However, more recent work has begun investigating different aspects of production using eyetracking methods (Bunger, Trueswell & Papafragou, 2012; Norbury, 2014; Rabagliati & Robertson, 2017). For example, Bunger and colleagues (2012) investigated differences between English-speaking 4-year-olds and adults during message encoding (see also Bunger, Skordos, Trueswell & Papafragou, 2016). They asked participants to watch motion events as their eye movements were monitored. Half of the participants in each age group were then asked to describe the event (e.g., *the boy skated into the net*) whilst the other half completed a memory test about the event. In their analysis of eye movements, they found that adults and children attended to both manner and path aspects of the event with similar timing regardless of task, suggesting that both age groups encoded the event in similar ways. However, children tended to omit one event component (i.e., manner or path) in their utterances, despite being able to successfully determine manner and path components in the memory task. Overall, these results suggests that pre-schoolers struggled with the linguistic element of the task (i.e., turning the non-linguistic representation into a linguistic one). However, as Bunger and colleagues only measured participants' eye movements as they watched the video, and did not continue measuring eye movements prior to and during speech, it is unclear exactly what process caused children to omit one aspect of the event.

Similarly, Norbury (2014) measured eye-movements of typical and atypical children as they described pictures (note that their data only looked at one typical child's eye movements (to compare against atypically developing children's eye movements). However, unlike Bunger and colleagues, she measured fixations to various scene elements (e.g., agent, patient, event core, and background) prior to speaking, before speakers mentioned the subject, and after speakers mentioned the subject. In their analysis, their typically-developing speaker (aged 10.5 years) looked more to the agent prior to speaking, looked equally towards the agent and event core prior to naming the subject, and looked more to the patient after naming the subject, whereas their atypical sample (aged between 8 and 14 years) tended to focus on the background, regardless of what they were talking about, suggesting that atypically developing children were affected by poorer visual attention control. This study in particular highlights the difficulty of investigating production processes in young children and raises the question of whether young children, who have limited attention spans, will be able to focus on aspects of a scene whilst also producing an utterance.

However, a more recent and successful example of using eye-tracking as a method with pre-schoolers comes from Rabagliati and Robertson (2017), who investigated the monitoring processes of 3-5-year-olds as they dealt with referential ambiguity. They measured children's and adults' eye movements before and after they described pictures in ambiguous scenes (e.g., *a yellow car* in ambiguous scenes where there are two cars vs *a car* in scene where there is only one car among two other objects). They analysed participants' saccades between the to-be-described picture (e.g., a yellow car) and a distractor foil picture (e.g., a red car). Adults tended to produce more saccades before and after naming, suggesting they were pro-actively monitoring the environment for potential ambiguity prior to speaking, as well as monitoring their utterance for potential ambiguity. Children also showed a similar pattern when they produced informative utterances (e.g., "*a yellow car*"), but failed to show this effect when they produced uninformative utterances (e.g., *"a car"*). Overall, these results suggest that monitoring processes are more error-prone in children. However, they suggest that children have the ability to pro-actively monitor their environment for potential ambiguity – something that would have been difficult to distinguish without using eye-tracking methods.

Overall, these studies show that researchers can successfully use eye-tracking paradigms to investigate production processes in children. The next step in developmental research using these methods is to investigate how children construct their utterances in real time, as the first eye-tracking studies did in the adult psycholinguistics literature 20 years ago. This thesis attempts to provide some insight into how children plan and produce their utterance by measuring children's eye movements before they begin speaking and whilst they are speaking.

1.5. The Architecture of the Developing Production System

Studies investigating the mechanisms underlying children's production are limited, but as we shall see, where available they typically suggest that the overall architecture of the production system is surprisingly adult-like. Here, we focus on findings related to grammatical encoding. Specifically, we discuss the evidence underlying content subprocesses - namely lexical retrieval – and structural subprocesses – namely syntactic encoding. Finally, we consider whether children can plan their utterances incrementally or not.

1.5.1. Producing Words

Children typically begin producing their first words early in their second year. Following the onset of word naming, the rate of word acquisition is generally quite slow, with children learning a few new words a month and by 18 months of age, infants are typically able to produce 50 words. However, towards the end of their second year, children demonstrate a sudden spurt in their expressive vocabularies as new words are acquired at a rapid rate (Bates, Bretherton, & Snyder, 1988; Benedict, 1979; Bloom, 1973; Corrigan, 1978; Dromi, 1987; Fenson, Dale, Reznick, Bates, Thal, Reilly & Harthung, 1990; Gentner, 1983; Goldfield & Reznick, 1990; Goldin-Meadow, Seligman, & Gelman, 1976; Gopnik &

Meltzoff, 1987; Halliday, 1975; Lifter & Bloom, 1989; McShane, 1980; Nelson, 1973). This vocabulary spurt is often coupled with qualitative changes in children's linguistic abilities, and over the years, it has been linked to the development of children's representational, semantic, and conceptual skills (Bloom, 1973; Golinkoff, Mervis, & Hirsh-Pasek, 1994; Gopnik & Meltzoff, 1987, 1992; Mervis & Bertrand, 1994).

But children can understand more words than they can produce: Why can't children produce all the words they have already learned when they first start speaking? To answer this question, we need to think about the difference between comprehending a word and producing a word: Comprehension simply requires recognition of word sounds and recall what that word refers to. Production, on the other hand, requires retrieval of the pattern of sounds associated with a given meaning. The discrepancy between production and comprehension – for words at least – must depend on the processes of retrieving this information. If this is the case, then this would suggest that the vocabulary growth spurt is the outcome of children having perfected these retrieval processes.

Dapretto and Bjork (2000) considered this hypothesis by investigating at the relationship between the vocabulary growth spurt and children's word retrieval abilities. Children who were either pre-vocabulary growth spurt, undergoing a spurt, or post-spurt watched an experimenter hide objects they could already name into one of two boxes. The box either had pictures of the object on it or it did not. Once the object was hidden, the experimenter would ask children "*what* is in the box?" She then removed the picture cue and asked "*where* is the X?" All three groups of children were able to successfully answer the *where* question, showing that they were able to understand the object label. However, children with limited vocabularies (i.e., the pre-spurt children) were less likely to produce the labels of the hidden object names than were children with larger vocabularies (i.e., the post-spurt children). This is despite them being able to name that object. In addition, pre-spurt children were more likely to name the hidden object when there was a picture cue on the box than when there was no cue. Overall, these results suggest that lexical retrieval is challenging at the early stages of lexical development and relies on the availability of contextual cues (Bloom, 1993). The reliance on cues to aid word retrieval would limit the number of words a child could produce as well as constrain the range of situations in which those words could be produced. That is, children may be able to produce the word *dog* in the presence of a dog but would not be able to do so at times when a dog is not present. Indeed, this finding could implicate improved word retrieval processes as underlying children's ability to talk about things that are not physically present (Ingram, 1978).

Another way researchers have investigated children's lexical retrieval abilities is to focus on the naming errors they produce (Dapretto & Bjork, 2000; Gershkoff-Stowe & Smith, 1997; Gershkoff-Stowe, 2001, 2002;). Naming errors have often been attributed to overgeneralizations (e.g., Clark, 1973), which assumes that children's naming errors are therefore a linguistic knowledge issue: children are overextending the label of *dog* to a different category because they thought the word extended further than it actually does or because they do not yet have a better word to describe an object in their lexicon, and so using that word is the best approximation they can make. However, a popular assumption is to view naming errors as retrieval errors (Bloom, 1973; Gershkoff-Stowe & Smith, 1997). According to this assumption, children's errors instead reflect a disruption in their ability to retrieve the correct word from their lexicon. This may be due to the lexical representation being weakly represented, which may be the case when a word is newly learned, or not being represented at all, as is the case when learning a novel word. And indeed, children's naming errors are often the result of interference of a previously retrieved word (Gershkoff-Stowe, 2001; Gershkoff-Stowe & Smith, 1997; Gershkoff-Stowe, Connell & Smith, 2006).

Chapter 1

Researchers have typically found a u-shaped curve in the frequency of naming errors in the child's second year: Children rarely produce naming errors when they have limited vocabularies (i.e. pre-vocabulary spurt), yet during the vocabulary spurt, children's show a dramatic increase in the number of naming errors they produce (Gershkoff-Stowe, 2001; Gershkoff-Stowe & Smith, 1997; see also Dapretto & Bjork, 2000). Importantly, the frequency of these naming errors typically decreases post-vocabulary spurt, even though children's vocabulary growth and amount of talking continued to rise. Overall, this suggest important changes to children's lexical retrieval abilities during the growth spurt. But what exactly are these changes due to?

One explanation lies in the strength of children's lexical representations. Thus, when children first begin speaking, the lexicon is immature with weak lexical representations, that don't allow lexical retrieval processes to function properly or in an efficient manner. Indeed, evidence suggests that children's early lexical representations lack fine phonetic detail pre-vocabulary growth spurt (Stager & Werker, 1997; see also Swingley & Aslin, 2002). During the growth spurt, newly acquired words would have weak representations in an ever-expanding lexicon. Thus, activation to these weaker words would be low and as a result would be weaker to resist interference from words that have stronger representations or more activation. As a result, children may are more likely to produce naming errors. Yet, more practice retrieving words results in stronger representations, allowing for more resistance to competitors and better retrieval and fewer naming errors.

Overall, these results suggest that as children develop, their lexical retrieval abilities become more accurate and less prone to interference (although it should be noted that factors that interfere with children's lexical retrieval also affect adult's retrieval processes during lexical priming; Gershkoff-Stowe, 2001; 2002; Gershkoff-Stowe et al., 2006).

We now consider the stages of lexical retrieval. As discussed previously, adults show a two-stage retrieval process whereby they first select a lemma and then retrieve the corresponding word form associated with that. Analysis of children's errors suggest that they too go through this two-stage process. For example, their word substitution errors clearly show a distinction between form-driven and meaning-driven retrieval errors (see Table 1).

Table 1: Examples of word substitution errors from Jaegar (2005).

Error Type	**Error**	**Target Word**	**Age (page no.)**
a. Phonologically-based	Hey, *Collection* is on!	Contraption	2;7 (635)
b. Semantically-based	*Daddy*, sit.	Mommy	1;9 (629)
c. Mixed (semantically/phonologically-based)	Thank you for the *cookie*.	candy	3;0 (636)

However, there are important differences in the frequency of these kind of errors in children's production. Firstly, children typically produce more mixed errors like (c) than adults do (Jaeger, 2005). For adults, the majority of their word substitution errors are either form-based or meaning-based, with mixed errors only making up a small minority of errors produced (e.g., Fay & Cutler, 1977). Yet in Jaeger's corpus, these kinds of errors are fairly common. Secondly, the majority of children's substitution errors are semantically-based in some way (i.e., errors such as (b) or (c) in Table 1). Finally, young children rarely produce purely phonologically-based errors like (a), but as they get older, children begin to produce significantly more of these kinds of errors. Note that this is different from semantic and mixed errors, which do not show this kind of trajectory. This pattern of errors suggests that during lemma selection, semantic competitors receive higher levels of activation, resulting in more interference. However, the fact that children produce more mixed errors rather than pure phonological-based errors could suggest that phonological activation can increase activation to semantically-related lemmas, resulting in greater interference, but phonological activation on its own rarely triggers an error on its own. This could suggest that children's phonological representations of a word are weaker than their semantic representations.

Overall, the results discussed suggest that children find lexical retrieval challenging when they first begin speaking, but once they go through their vocabulary growth spurt, lexical retrieval processes seem functionally adult-like.

1.5.2. Producing Multi-Word Utterances

Children begin putting words together shortly after they begin producing their first words. This typically occurs around about the same time they go through their vocabulary growth spurt. However, these multi-word utterances, also known as holo-phrases, are typically very short, and do not include any kind of function words. For example, if a child sees a ball on the table, they may produce an utterance like "*ball*" to describe this scene. At around this time, many of children's multiword utterances also show a more systematic pattern in that they use one word to structure their utterance in the sense that it determines the speech act, with the other word(s) slotting into the gap. For example, they may use a word like "*more*" with a wide variety of object labels (e.g., *more milk, more juice, more food*). Similarly, within their second year, children also begin to use verbs in their utterances (e.g., *give ball*). However, children's ability to use verbs and different syntactic structures in a flexible way seems limited as certain syntactic structures seem to be lexically tied (Akhtar & Tomasello, 1997; Pine, Lieven & Rowland, 1998; Tomasello, 1992, 2000). That is, children may be able to use the verb *give* in a simple way (e.g., *give ball, give cake*), but will be able to use another verb, such as *draw*, using more complex syntactic structures and frames (e.g., *Draw dog on that, draw dog for mummy*). Children can also show this same pattern with nouns and will only have certain nouns appearing as only a subject or only an object (e.g., "mummy" only appears as a subject and "doggie" only as an object in children's productions; Pine et al., 1998). By the pre-school years, children typically begin producing a variety of abstract utterance-level constructions, including such things as transitives, intransitives, ditransitives, passives, imperatives, reflexives, locatives, causatives, and various kinds of question constructions. These kinds of findings have led some researchers to propose that children's early syntactic representations are item-specific, which they eventually generalize to create abstract syntactic representations (Bloom, 1998; Tomasello, 2000).

However, children appear to understand more complex syntactic structures before they can produce them (Bates, McWhinney, Caselli, Devescovi Natale & Venza, 1984; DeVilliers & DeVilliers, 1973; Hirsh-Pasek & Golinkoff, 1991, 1996). One potential explanation for this discrepancy between comprehension and production is that children's syntactic representations are not strong enough to retrieve yet for production (as with lexical retrieval). An alternative explanation is that the developing production system places limits on the kinds of utterances children are able to produce. It is possible that certain structural subprocesses are not yet fully developed, resulting in omissions, errors, disfluencies, and ungrammatical utterances that the child themselves would deem as being incorrect.

Indeed, Wijnen (1990) suggested that children's inability to produce grammatical utterances is due to the fact they have not yet developed the mechanism underlying constituent assembly during positional planning. He measured the number of disfluencies a two-year-old child produced over the course of a seven-month period, which captured the transition from 'pre-grammatical' (i.e. two-word holophrases) to grammatical language. He found that the frequency of disfluencies showed a u-shaped curve: in the first half of the observation period, the rate of disfluencies increased and then declined in the latter half. As well as this change in the frequency of disfluencies, the distributional pattern also changed: in the first half of the observation period, dysfluencies were distributed relatively randomly over sentences, yet in the second half they tended to appear at function words and sentence-initial words (thus showing a similar hesitation pattern as adults). Wijnen argued that this pattern of results indicates that the mechanism underlying positional planning, whereby the child is able to slot words into various syntactic frames, is not yet fully developed in the 'pre-grammatical' stages of production. As such, children are unable to produce fully grammatical utterances until this mechanism develops.

Similarly, there may be differences in the way children access syntactic elements during production, resulting in omissions or errors. For example, children's omission of functional morphology (e.g., function words) in early speech has often been regarded as reflecting limited syntactic knowledge. However, findings showing that even two-year-olds can comprehend functional morphology before they reliably produce it may implicate the developing production system as the reason underlying these omissions and errors (e.g., Gerken & McIntosh, 1993; Petretic & Tweney, 1977; Shipley, Smith, & Gleitman, 1969). It is possible there are qualitative differences in the way children and adults retrieve function words. Indeed, Wijnen (1992) found that two and three-year olds children produce more sound errors on function words than adults do. He suggests that this is due to young children not yet developing the specialised retrieval procedure needed for function words. Instead, they are retrieving function words in the same way as content words. This may be because function words have not yet formed a closed class in young children as their lexicon is still expanding (but see McKee & Iwasaki, 2001 for an alternative explanation whereby children's omission of function morphemes is due to their late realization).

The results discussed so far suggest that some mechanisms necessary for producing grammatical multi-word utterances are not yet fully developed when children first begin speaking. However, by the latter half of their second year, the overall architecture of the production system is similar to that of adults. Indeed, the analysis of children's speech errors suggest that the planning processes underlying multi-word production are adult-like by this time. For example, word exchange errors like those seen in Table 2 suggest that children can complete multiple production processes simultaneously. Errors like (a) occur in both children and adults, suggesting that lexical retrieval and phrasal planning occur concurrently for both age groups (Jaeger, 2005). Similarly, in errors like (b), the interacting elements (*water* and *head*) go across two adjacent phrases, as well as being a part of the same grammatical category. Errors like this have been interpreted in the adult data as reflecting early syntactic planning, whereby speakers begin planning aspects of later phrases whilst planning the first. Thus, errors of this sort suggest that, at least by the age of 5, children's production processes appear to be adult-like (although note that just because children produce similar errors as adults, it does not necessarily mean that the underlying mechanisms underlying these errors are the same).

Table 2: Examples of word exchange errors seen in children aged 2-5 years from Jaegar (2005).

Error Type	**Error**	**Target**	**Age (Page no.)**
a. Word exchange with morpheme stranding	Her *run* is *nos*ing.	Her nose is running	2;7 (706)
b. Cross-phrase word exchange	I can go in the deep part where the *head*'s over my *water*.	I can go in the deep part where the water's over my head.	5;11 (708)

Similarly, children also produce anticipation and perseveration errors like adults do (see Table 3). Anticipation errors indicate advance planning and that lexical retrieval processes are running faster than phrasal planning, whereas perseveration errors could indicate lexical retrieval is slower than phrasal planning. Indeed, for anticipatory errors like (a) to be triggered, the content of the to-be-spoken word (in this case *I*) must be planned at some level in the linguistic representation prior to speaking.

Table 3. Examples of anticipations and perseverations in pre-schoolers from Jaegar (2005)

Error Type	Error	Target	Age (Page no.)
a. Word anticipation	Yeah, *it* likes it ... I like it.	Yeah, I like it	3;8 (687)
b. Word perseveration	Daddy, me watching *Daddy* cooking ... no ... Mommy's cooking	Daddy, me watching Mommy's cooking	2;4 (695)

Whilst children do produce both kinds of errors, there are some differences in the frequency of these kinds of errors in children and adults: Children produce more perseverations than anticipations (Jaeger, 2005; Wijnen, 1992), whereas adults produce more anticipatory errors. This has been taken as evidence that children are planning less of their utterance in advance than adults. However, the fact that children produce more perseverations could reflect limited inhibitory skills, as the previously produced (and semantically-related) word in (b) could inhibit the target word, resulting in them repeating what they have just said (Gershkoff-Stowe & Smith, 1997). The fact that children produce fewer anticipatory errors than adults could reflect a performance issue in children's planning of multi-word utterances: advanced planning requires the speaker to hold a plan of the to-be-spoken utterance in a buffer and thus, relies on a speaker's working memory capacity. This would suggest that children, who have limited working memory capacity compared to adults, would struggle with advanced planning.

Overall, the findings discussed suggest that the overall architecture of the production system is similar to that of adults by the latter half of children's second year. However, there are indications that there are limits in children's ability to fully implement the full resources of the production system (e.g., limitations in their ability to plan multi-word utterances in advance). What is missing in the production literature is an exact understanding of how children plan their utterances. There is only so much that error data can tell us about how the developing production system operates. It is also important to remember that it is unclear exactly what children's production errors reflect: on the one hand, errors could reflect difficulties planning, yet other errors may also reflect limited linguistic knowledge. As discussed earlier (Section 1.4.2.), much of our current understanding of how the adult production system operates came with the development of more precise methods, such as eye-tracking and chronometric methods, which could measure how speakers produce during *successful* production. We argue that now is the time to use these more precise methods to further order to understand exactly how children plan their utterances and to investigate the development of the production system more accurately.

1.5.3. Can Children Plan Incrementally?

As discussed previously (Section 1.2.), a hallmark of adult production is that it is incremental. Rather than prepare a whole sentence in advance, speakers instead encode small chunks of their utterance whilst simultaneously articulating previously encoded chunks. This kind of planning removes some of the cognitive load associated with production, as speakers do not have to hold their whole utterance in a buffer. It also allows the speaker to continuously update their message with new information. However, incremental planning also poses some challenges for the speaker as it requires them to coordinate the sequence and timing of various production processes. This means that various production processes are working in parallel with each other and as such could affect the processing resources available to the speaker during utterance planning. Adults have automatized the processes involved in production and so therefore have the mental capacity to plan in this way. However, young children, whose planning processes may be more susceptible to the processing demands of production, may find incremental planning challenging.

In the following sections, we consider whether pre-schoolers can use the sentence planning strategies that adults can use (namely, structural incrementality and lexical incrementality) to develop a plan to incrementally retrieve words. We then consider whether pre-schoolers can incrementally retrieve word forms like adults.

1.5.3.1. How do pre-schoolers develop a sentence plan?

Adults can develop a sentence plan by either construing the overall gist of a scene which guides the order by which words are incrementally retrieved (i.e., structural incrementality) or they can begin retrieving the most accessible or salient word first which then determines the form their sentence will take (i.e., lexical incrementality).

On the one hand, children may prefer to use a structural incrementality strategy like adults do, as this kind of planning removes parallel processing between conceptualisation and formulation. In adults, shifts between lexical and structural incrementality planning strategies occur when linguistic encoding of individual sentences is harder or easier to execute (Konopka & Meyer, 2014): that is, adults will use a lexical incremental strategy when linguistic encoding is easier but a structural incrementality strategy when linguistic encoding is harder (see also Konopka, Meyer & Forest, 2019). This suggests that adult speakers prefer to complete easier processes before harder processes. We can assume that this will also be the case with children. Message-level encoding may be easier to complete for children than linguistic encoding which requires parallel planning (Bunger et al., 2012). Thus, children may prefer to encode the message first (the "easier" process) and postpone linguistic encoding (the "harder" process). This is exactly what non-native adult speakers do: L2 adult speakers typically prefer to plan their utterances in a structural incremental way (Konopka et al., 2019), presumably because this way of planning reduces the amount of cognitive load required once the sentence plan has been generated (i.e., after 400ms) as the speaker can concentrate wholly on the linguistic encoding of that plan.

However, there are reasons to believe that children will rely more on the bottom-up lexically incremental strategy. First, the lexical incrementality account assumes that lexical retrieval is the key to sentence planning and is what drives a sentence's syntactic structure. This assumption is consistent with proposals that attribute a central role to the lexicon in young children's early representations of syntax (e.g., Tomasello, 2000). For example, two-year-olds may be able to produce a transitive construction with some verbs but not all the verbs they know, suggesting that the syntactic structure has not yet been generalised across all verbs and is instead lexically specific, i.e., linked only to specific verbs (Tomasello, 2000). One might therefore expect that young children would rely on lexical retrieval when determining syntactic structure during sentence production, i.e., they would first retrieve the relevant lexical entry before accessing the syntactic structure associated with it. In addition, the lexical incrementality account predicts that speakers will begin talking about whatever grabs their attention first, thus assuming attention and language production are linked. Indeed, the interplay between attention and language has been intensely investigated in psycholinguistics (e.g., Henderson & Ferreira, 2004) and many developmental studies have shown that during language learning, infants' make use of various attentional cues, such as caregivers' gaze or pointing, to match names to referent objects (Baldwin, 1995; Carpenter, Nagell, Butterworth & Moore, 1998). Similarly, during language comprehension, two-year-olds attend to each referent in the order that they have been auditorily described in a sentence (e.g., Yuan & Fisher, 2009). Thus, it seems plausible that pre-schoolers will not only attend to objects in the order they mention them in, but their utterances may also be influenced by external influences that capture their attention. Indeed, attentional cues may make sentence planning easier for children, as it highlights one object or character, which should therefore become more accessible, resulting in facilitated lexical retrieval.

Recent studies have investigated the planning strategies that children use by measuring whether attention manipulations affect children's word order choice (Brough, Branigan, Gambi & Rabagliati, 2018; Ibbotson, Lieven & Tomasello, 2013). If children use a lexically incremental planning strategy, the beginning of their utterance (i.e., their starting point) should be influenced by an attention manipulation. However, if children use a structurally incremental planning strategy, starting points should be influenced by linguistic biases, such as a preference to begin a sentence with an agent. Ibbotson and colleagues (2013) had pre-schoolers and adults describe transitive events whereby the experimenter either directed their gaze to either the agent or patient of a scene (test condition), thus directing participants' attention to a character, or they looked directly towards the participant (control condition), and measured whether children would use an active or passive construction (i.e., would they begin their sentence with either the agent or the patient). The experimenter's eye gaze influenced sentence construction for four-year-olds but not three-year olds: four-year-olds were more likely to produce a passive if the experimenter looked at the patient compared to when the experimenter looked at the agent, and vice versa when the experimenter looked at the agent, suggesting that the four-year-olds' starting points were influenced by the attentional manipulation. However, three-year-olds' sentence construction was influenced by eye gaze only when the experimenter looked at the agent, and not at the patient. That is, they were more likely to produce an active if the experimenter looked at the agent but were not more likely to produce passives when the experimenter looked at the patient. More recently, Brough, Branigan and Rabagliati (in prep) tracked three- and four-year-olds' eye movements as they described pictures of events. Importantly, they investigated the influence that children's attention and their linguistic biases had on starting points by manipulating the animacy (e.g., human vs. shape) and saliency (e.g., grayscale vs. colourful) of the characters depicted in an event. They found that whilst children fixated on the most salient character first, they tended to begin their utterance with the animate character in subject position (e.g., *the man is chasing the triangle*). Overall, these results suggest that pre-schoolers use a planning strategy that favours linguistic biases (i.e., structural incrementality) over one that assumes speakers develop a sentence plan based on what they look at first (i.e., lexical incrementality).

However, both these studies asked participants to describe pictures of events, meaning they had to generate semantic roles as well as determine grammatical relations and produce a complex constituent structure. As a result, there is tension between the attentional manipulation and other constraints linked to grammatical relations and constituents, which may increase cognitive load. Thus, it may be difficult to detect an influence of attention on word order choice in these studies. It may only be possible to find effects of attentional manipulations on word order if the to-be-produced utterance does not conflate syntactic structure with word order choice.

1.5.3.2. Can pre-schoolers incrementally retrieve word forms?

We now consider whether children can retrieve words incrementally, or whether they plan all aspects of their utterance before they begin speaking. No study to date (to the author's knowledge) has investigated whether children do indeed retrieve lexical items incrementally. However, if we assume a primary issue for children is that they have limited processing resources and as a result, struggle with the processing demands involved in production, then we can make some assumptions about whether they can plan incrementally or whether they instead plan their whole utterance prior to speaking (i.e., holistic planning).

On the one hand, we might assume children fail to plan incrementally and instead plan their whole utterance in a serial-like fashion. That is, they retrieve all lexical items prior to speaking, put them in the correct order, and then begin the process of articulation. This kind of planning would mean children have to hold their prepared utterance in a buffer of some sort. Cross-phrase word exchanges as seen in Table 2(b) suggest that children are planning across phrases, suggesting that they are planning in a holistic way rather than an incremental way.

Planning in this holistic way would require more cognitive load for children before speaking as children have to plan their whole utterance and hold that utterance in a buffer. However, once they begin speaking, cognitive load would be more evenly distributed as they are not engaging in parallel planning as they do not have to plan later parts of their utterance whilst speaking. If children do plan in this holistic way, then it might explain why children typically produce shorter utterances than adults (Jaeger, 2005). By producing shorter utterances, children would have less prepared speech to hold in a buffer, resulting in less cognitive load. Holistic planning in this way may also affect the final output of an utterance. For example, we can assume that a buffer will have a certain capacity. Thus, earlier prepared words (such as the subject) may be more affected by later prepared words. Indeed, English-speaking two-year-olds will omit subjects from their utterances (Bloom, 1990; Valien, 1991), and this is more likely to happen when they produce longer sentences than shorter ones (Valian, Hoeffner & Aubrey, 1996; Valian & Aubrey, 2005). We can explain this finding under this hypothesis: as the buffer has a certain capacity, once that capacity is reached, previously encoded chunks are pushed out to make room for newly encoded chunks. Thus, in the case of null subjects, as children plan later part of their (longer) utterance, the subject is pushed out of the encoded utterance because there is no more room in the buffer for it. Thus, the child produces an utterance without a subject. However, holistic planning places greater demands on children's already limited working memory. Incremental planning requires children to only plan small chunks of their utterance and thus requires less working memory capacity. However, it does require children to coordinate production processes so that the results of those processes are available to them at the correct time.

We can make some assumptions about whether children incrementally retrieve words or not, by looking at the errors they produce. As discussed previously, children and adults differ in the number of perseverations (e.g., Daddy, me watching *Daddy* cooking ... no ... *Mommy's* cooking) and anticipatory errors (e.g., Yeah, *it* likes it ... *I* like it.) that they make: adults produce significantly more anticipatory errors than children do, who instead produce substantially more perseverations (Jaeger, 2005; Wijnen, 1990). Anticipatory errors are thought to reflect advance planning, as in order to produce an anticipatory error, a speaker must have already encoded later elements of the upcoming speech which then triggers the error. Thus, if young children produce fewer anticipatory errors, it may mean they engage in less advance planning than adults do, and instead incrementally retrieve lexical items whilst speaking. Children's perseverations may also reflect incremental planning: if children have not yet retrieved the correct word form required for the next part of their utterance, they may repeat a word form they have just used as a placeholder so they can continue their utterance. Evidence of parallel planning in children comes from word exchange errors (e.g., *her run is nosing*): children, like adults, produce word exchange errors, suggesting that lexical retrieval and phrasal planning occur concurrently for both age groups (Jaeger, 2005). Taken together, these findings suggest that not only are children able to plan incrementally, but the planning process is parallelized, just like in adults.

As with adults, incremental planning requires the speaker to be strategic in the amount of advance planning that they undertake. If lexical retrieval lags behind articulation, speakers will become disfluent and may produce more filled pauses or restarts. However, if lexical retrieval is faster than articulation, more words are stored in a buffer, resulting in greater demands to working memory. Speakers must learn to balance between these processes. The evidence suggests that this is something children struggle with. For example, McDaniel and colleagues (2010) asked younger (three- to five-year-olds) and older children (six- to eight-year-olds) and adults to complete an elicited production task which encouraged the production of different relative clauses. They found that children produced more filled pauses and restarts than adults did. Furthermore, children typically produced a relativizer (e.g., who) and complementizer (e.g., that) more than adults and paused after producing one, suggesting that children were using these components of their utterance as a stalling device so they could begin planning of the next part of their utterance. Based on these findings, the authors propose that adults can plan further ahead than children (e.g., they can encode a whole phrase whilst speaking), whereas children need to plan more frequently and more locally than adults do. That is, children need to stop to plan each phrase as it comes, resulting in more disfluencies and pauses, whereas adults can continue planning the next phrase as they are articulating the current one, resulting in fluent speech. Findings such as these could suggest that children do plan their utterances in small chunks (i.e., incrementally), but struggle with parallel planning. That is, they do plan incrementally but not in the same way that adults do: they encode small parts of their utterance, articulate what they have already encoded, and then begin encoding of the next part of their utterance, resulting in more pauses and disfluencies in their actual output. This could suggest that pre-schoolers are planning incrementally but they are not engaging in parallel planning.

To conclude this section, incremental planning is a hallmark of adult production, yet it may be something children struggle with as they have limited processing resources. The results discussed so far present a mixed picture for whether children can engage in incremental planning or not. On the one hand, we find evidence of incremental planning: children produce errors that indicate they plan less of their utterance in advance than adults do. Yet, children also tend to produce shorter utterances than adults which could reflect holistic planning. In addition, children's disfluencies and speech rate analyses suggest that they struggle with parallel processing – a requirement for successful incremental planning. Importantly, error data fails to tell us the moment-by-moment mechanisms involved in children's planning. Eye-tracking methods in adult production studies have revealed how the production system operates during *successful* language use (e.g., Griffin & Bock, 2000; Meyer, Sleiderink & Levelt, 1998). Developmental researchers must also begin using these methods in order to truly grasp how children plan their utterances.

1.6. Communicative Abilities

Our discussion so far has focused on children's ability to produce words and put those words together to produce multi-word utterances. However, children must also learn to produce utterances in an effective way that can be easily understood by their addressee. In particular, they must be able to produce utterances that are both unambiguous and informative. Yet, it is well established that young children find this especially difficult to do. Even young school age children, who have mastered producing most of the structural features of their language, frequently produce unclear and ineffective descriptions when they are trying to describe a specific referent to an addressee (e.g., describing an object as *the red one* even though there are other objects of the same colour present; Deutsch & Pechmann, 1982; Glucksberg & Krauss, 1967; Glucksberg, Krauss & Weiserg, 1966; Matthews, Lieven & Tomasello, 2007). Whilst children's difficulty in referential communication is well established, it is still unclear why this difficulty persists so late in development.

One explanation for this has focused on perspective-taking: In order to produce an informative referential expression, a speaker must monitor their partner's perspective against their own and use this information to plan their utterances (Clark, 1992; 1996). Historically, children under the age of 7 were assumed to be egocentric and thus incapable of tailoring their communicative acts to the mental states of their partner, resulting in utterances that appear to not take their partner's perspective into account (Deutsch & Pechmann, 1982).

However, despite their tendency to produce egocentric utterances, study after study has shown that infants are able to reason about the mental states – the knowledge and beliefs – of others and adapt their communicative acts in accordance with this (Akhtar, Carpenter, & Tomasello, 1996; Liebal, Carpenter & Tomasello, 2010; Moll, Richter, Carpenter & Tomasello, 2008; Perner & Leekham, 1986). Findings such as this have been taken as evidence that young children are able to recognise that their partner has perspectives different from their own, but they struggle with integrating this knowledge into their own messages for production (Nadig & Sedivy, 2002).

More recent theories have argued that children's general cognitive limitations, such as their developing inhibitory skills, can explain children's inability to use their perspective-taking skills whilst speaking. This account assumes that both children and adults have egocentric biases when producing and understanding referential expressions, but adults can inhibit these biases in a much more efficient manner than children can, and as a result will communicate and process referring expressions in a way that is consistent with their partner's perspective (e.g., Brown-Schmidt, 2009; de Cat, 2015; Epley, Morewedge & Keysar, 2004a; Nilsen & Graham, 2009; Wardlow, 2013). Indeed, during comprehension, both children and adults show automatic egocentric interpretations of their partner's utterances, but adults are much quicker and more effective at correcting their interpretation than children are (e.g., Epley et al., 2004a). Likewise in production, adults are more likely to produce egocentric utterances in cases where they are unable to override their biases, such as when they are under cognitive load or time pressure (e.g., Epley, Keysar, Van Boven & Gilovich, 2004; Ferreira, Slevc & Rogers, 2005). However, it is unclear to what degree inhibitory skills play a role in children's ability to produce perspective-taking utterances. Whilst Nilsen and colleagues (2009) have shown that 4-5-year-old children who have better inhibitory skills are more likely to override their egocentric biases when *interpreting* their partner's statements, they have been unable to find strong evidence for this relationship in children's *production* of sentences (Nilsen & Graham, 2009; Nilsen, Varghese, Xu & Fecica, 2015).

Perspective-taking explanations of children's poor referential skills ignore the effect the developing production system may have on children's abilities to produce informative utterances. It seems plausible that skills required for successful production, such as inhibition, working memory or monitoring, may play important roles in children's success at referential communication. For example, inhibitory skills play an important role in adults' sentence production and are necessary for production processes such as lexical retrieval: speakers must select the correct word and inhibit competing alternatives (e.g., Schriefers, Meyer & Levelt, 1990). Indeed, typical referential communication paradigms, whereby participants describe a target object amongst distractor objects that are identical to the target apart from on one dimension (e.g., a red car is the target, and a yellow car is the distractor), require all of these skills. To succeed in these tasks speakers must monitor the context for any potential ambiguity (i.e., pro-active monitoring), identify and distinguish the target object from the distractors, and encode any distinguishing features into their chosen referring expression, all whilst inhibiting prepotent and more frequent under-informative expressions (e.g., "the car" in multiple car contexts). Thus, it seems likely that pre-schoolers – whose inhibitory and monitoring skills are still developing – would struggle to consistently produce informative utterances.

For example, pre-schoolers may be more likely to struggle to inhibit highly frequent words. As discussed earlier, young children's lexical retrieval processes can often be affected by interference, resulting in errors (Gershkoff-Stowe & Smith, 1997). Thus, when children produce utterances that appear uninformative and ambiguous, it may in fact be failures to inhibit more dominant expressions. Indeed, we can explain classic findings demonstrating children's inability to produce informative referential expressions with this hypothesis. For example, Pechmann and Deutsch (1982) asked children (2- to 9-year-olds) to describe an object they liked best out of an array of 8 objects. In order to uniquely identify a specific referent, children had to identify multiple dimensions (e.g., size, shape, colour), thus giving children more choice in deciding how to describe the referent as well as increasing the complexity of formulation processes. They found that in this task, 6-year-olds produced inadequate descriptions (i.e., descriptions that would not specify a unique referent) 50% of the time, whilst 2-year-olds produced such descriptions 94% of the time (Pechmann & Deutsch, 1982). However, this task requires children to not only inhibit responses associated with the other objects in the array (e.g., *the small green spoon*), but also inhibit more dominant labels (e.g., *big yellow spoon* could simply be described as *spoon* because spoon is the more typical label). Indeed, children can produce more informative responses in cases where there are fewer distractor items, and thus fewer responses to inhibit (Davies & Kreysa, 2018; Nadig & Sedivy, 2002).

Similarly, children's developing monitoring skills can also affect their ability to produce informative utterances. Recent studies combining referential communication tasks with eye-tracking methods have shown that pre-schoolers typically fail to monitor the context for potential ambiguity prior to speaking, resulting in uninformative responses (Rabagliati & Robertson, 2017). That is, when describing a car to a partner, children may not check that there are no other cars in a scene before they begin speaking. Thus, they may produce an uninformative utterance like *car*. However, in cases where children do produce informative responses, they do look at both the target and distractor objects prior to speaking (Davies & Kreysa, 2018; Rabagliati & Robertson, 2017).

This pattern of results suggests that even pre-schoolers have the ability to pro-actively monitor their environment and adjust their utterance accordingly, but often fail to do so. Thus, rather than learning to inhibit the perspectives of others during utterance planning, these results suggest that children instead must learn to look for ambiguity within their environment to aid utterance planning. Rabagliati and Robertson (2017) also found that pre-schoolers often showed evidence of monitoring their own descriptions: they tended to look back to the distractor object after naming the target although failed to correct their utterance, suggesting they had a limited ability to incorporate this feedback back into production.

The previous sections (Section 1.5.1 & 1.5.2) have shown that by the latter half of their third year, the overall processes involved in production are seemingly adult-like. However, when we look at children's abilities to communicate in an effective way, we can see that even older children find such tasks difficult. This may be due to difficulties in integrating other developing skills with production processes. We now consider whether children can adopt their partner's perspective and then use this perspective in their own language use.

1.6.1. Can pre-schoolers adopt their partner's perspective during conversation?

As discussed previously (Section 1.3.), adult speakers can spontaneously, flexibly and responsively adopt different perspectives on objects during dialogue, and this is reflected in their referential choice. However, young children's ability to recognize and accommodate multiple perspectives in their language use remains contested. Early claims for egocentricity and inflexibility in young children's referential communication (e.g., Deutsch & Pechmann, 1982; Dickson, 1982; Krauss & Glucksberg, 1969; Whitehurst & Sonnenschein, 1981) are supported by findings that young children reject alternative perspectives - and therefore alternative names - for objects during comprehension. For example, three-year-olds deem a partner's use of an object name to be wrong if this is different from their own preferred perspective (Doherty & Perner, 1998; Doherty, 2000; Perner, Stummer, Sprung & Doherty, 2002). This failure may reflect difficulties in perspective taking and an inability to model another's knowledge against their own – skills necessary for alignment according to the collaborative account. In fact, it is only when children begin passing traditional false belief tasks that they also pass alternative naming tasks (Doherty & Perner, 1998).

However, more recent evidence suggests that young children can recognize and take different perspectives on objects during comprehension. For example, Matthew et al (2010) showed that three- and five-year-olds are flexibly responsive to different conversational partners' use of different perspectives within the same pragmatic context. In their study, children moved objects with two possible names (e.g., *horse* vs. *pony*) in response to instructions. Some children heard the same experimenter give instructions in two blocks, using the same name in both blocks (*horse-horse;* i.e., maintaining the same perspective) or switching names between blocks (*horse-pony;* i.e., changing to a new perspective); other children heard different experimenters in each block. Both three- and five-year-old children were able to switch perspectives across blocks. However, they were slower to do so in the different name/same experimenter condition than in the different name/different experimenter condition, suggesting that they associated a particular perspective with a particular partner. Four-year-olds also demonstrate a similar ability to switch perspectives (and associate particular perspectives with particular partners) when comprehending modified noun phrases (e.g., *striped ball* vs. *purple ball;* Graham, Sedivy, & Khu, 2014). Moreover, they can rapidly shift between multiple speakers' perspectives during comprehension (Khu, Chambers, & Graham, 2020).

However, preschoolers' capacity to flexibly take alternative perspectives on objects appears more limited during production. Three-year-olds can recognize and use different perspectives when given pragmatically appropriate contexts and/or cues that explicitly elicit multiple perspectives (Clark, 1997; Deák & Maratsos, 1998; Mervis, Golinkoff, & Bertrand, 1994); for example, pre-schoolers can call a dinosaur-shaped crayon both *a dinosaur* and *a crayon* when probed by an experimenter to use alternative perspectives (Deák, Yen & Pettit, 2001). However, the ability to spontaneously switch between perspectives (i.e., to abandon an established precedent and adopt a new perspective) seems to emerge later in development. In Deák et al.'s study, three-year-olds were less likely to adopt a particular perspective on an object after they already taken another perspective on it than when they had not.

However, the ability to adopt multiple perspectives during production seems do develop on a gradient: Both four- and six-year-old speakers can spontaneously switch from using a proper noun (e.g., *Peter,* an individual-level perspective) to a common noun (e.g., *boy,* a category-level perspective) in response to changes in the pragmatic context (e.g.., when they speak to a different partner who does not know the name of a character), but only six-year-olds similarly switch from a subordinate-level perspective (*woman's shoe*, in the context of more than one shoe) to a category-level perspective (*shoe*, in the context of a single shoe; Koymen, Schmerse, Lieven, & Tomasello, 2014; see also Birch & Bloom, 2002; Diesendruck, 2005). These results suggest that pre-schoolers understand that their partner may not have the exact same knowledge that they have (e.g., the names of other people), but they struggle to extend this understanding to alternative perspectives of common nouns.

Existing evidence does not discriminate between the mechanisms underlying referential communication (namely, the collaborative account, cue-based account, and priming-based accounts). Most of the findings regarding children's sensitivity to a partner's referential choice have been explained in terms of children's ability to explicitly model a partner's perspective, but they are also compatible with other explanations. For example, preschoolers' rejection of others' names for objects might reflect an impaired ability to model others' perspectives (Doherty & Perner, 1998; Doherty, 2000; Perner, Stummer, Sprung & Doherty, 2002), but could equally reflect the cognitive demands of the experimental task (i.e., the need to maintain two perspectives simultaneously; Clark, 1997). Similarly, three- and five-year-olds slowed comprehension when speakers fail to maintain referential precedents (Graham et al., 2014; Matthews et al., 2010) might arise because they explicitly model and accommodate partner-specific common ground (i.e., perspective-taking), but is also compatible with automatic memory-based associations between particular contexts, partners, objects, and expressions (Horton & Gerrig, 2005). Indeed, children in Matthews et al.'s study were slowed whenever they encountered an alternative perspective, irrespective of the speaker, suggesting a basic advantage for previously processed words that is compatible with lexical priming (see Ostashchenko, Deliens, Geelhand, Bertels, & Kissine, 2019).

One way to investigate the underlying mechanisms is to examine how pre-schoolers choose perspectives in response to a partner's previous referential expressions, and specifically whether they spontaneously adopt the same perspective as a conversational partner - in other words, whether they referentially align. Indeed, recent work has shown that older school-aged children can flexibly adopt their partner's perspective in their own language use (Branigan, Tosi & Gillespie-Smith, 2016). Furthermore, autistic children show referential alignment to the same extent as typically developing children and their partner's perspective was not predicted by performance on Theory of Mind tests (Branigan et al., 2016; Hopkins, Yuill & Branigan, 2017). The fact that performance on theory of mind tests did not influence autistic children's likelihood of aligning with their partner – thus their likelihood of using the same perspective –suggests that referential alignment is not necessarily reliant on mechanisms that implicate theory of mind. However, it is possible that different mechanisms underlie autistic and neurotypical children's alignment. For example, recent work has shown that neurotypical children's tendency to align with their partner (and use the same perspective) can be affected by manipulations designed to instil social exclusion, suggesting that older children's social affiliative mechanisms affect referential alignment (Hopkins & Branigan, 2020). However, autistic children do not show this same effect, suggesting that they did not rely on social-affective mechanisms to the same extent as neurotypical children (Hopkins, Yuill & Branigan, 2021).

Overall, these results suggest that different mechanisms underlie older neurotypical and autistic children's referential alignment. It is possible that pre-schoolers can also differentially use multiple mechanism to adopt their partners' perspective, and referentially align with them.

1.7. Summary of Literature Review and Research Questions

Much work has focused on how children learn the words and grammar of their language, with the emphasis being primarily on how children learn to *understand* their native language. Similarly, complex models have been developed to explain how adult speakers produce words and sentences. But very little work has combined these two fields to investigate how the mechanisms underlying children's production system develops: how do children learn to *speak* their native language and become active participants within their linguistic community.

Studies investigating the mechanisms underlying sentence planning in children are limited and, where available, have typically relied upon speech error data. These findings suggest that, overall, the architecture of the developing language production system is similar to adults. What is different, is the way children behave: children tend to produce more errors and disfluencies as the utterance they produce becomes more complex. For example, when it comes to produce multi-word utterances, children produce fewer anticipatory errors and more perseverations than adults do. This indicates difficulties in advanced planning of utterances (although it should be noted that this work has typically relied on the errors children produce, rather than on online processing). In addition, children often produce utterances that are difficult to interpret and understand when communicating with a partner, and such difficulties could be due to them struggling to encode the partner's perspective during planning processes or difficulties integrating developing skills with production.

Why is this the case? Do these errors reflect immature production mechanisms or is there something qualitatively different about the way children produce sentences? In this thesis, we examine these questions and ask generally how do pre-schoolers formulate simple sentences? Specifically, we explore the syntactic, morpho-phonological, and referential mechanisms underlying pre-schoolers language production by adapting experimental methods and tasks that have been successfully used to investigate mechanisms underlying adults' language production.

Chapter 1

In Study 1 (Experiments 1a and 2a), we investigate the mechanisms underlying pre-schoolers' sentence planning. Previous studies have shown that pre-schoolers can use linguistic biases when they describe events, indicating they use a structurally incremental planning strategy. But do pre-schoolers sometimes use bottom-up factors to build a sentence plan (i.e., can they plan in a lexically incremental way)? To do this, we focused on children's starting points, and ask whether bottom-up factors can influence how children will begin their utterance for simple sentences that require minimal semantic roles.

In Study 2 (Experiment 1b and 2b), we investigate the mechanisms underlying morpho-phonological encoding and measure whether pre-schoolers' eye gaze, like adults eye-gaze, is coordinated with their speech output. Such coordination would be indicative of incremental lexical retrieval. To do this, we measured children's eye movements as they named objects and compared this to adult controls.

In Studies 3-4 (Experiments 3 to 9), we investigate the mechanisms underlying pre-schoolers' referential communication. Specifically, we focus on referential alignment and examine whether pre-schoolers' choice of perspective and their associated referential expressions were affected by the perspective their partner previously used. To do this, we use an adapted referential communicative task that has successfully been used in studies investigating referential alignment in older children (e.g., Branigan et al., 2016; Hopkins et al., 2020), and tested whether children will reuse their partner's referential choice (which would show they can adopt their partner's perspective).

Chapter 2

2. Study 1. Mechanisms Underlying Pre-Schoolers Sentence Planning: Can Pre-Schoolers Use Bottom-Up Factors To Guide Word Order Choice?

2.1. Introduction

In order to produce a sentence, speakers need to decide the order in which the different elements will appear. To do this, they must develop a sentence plan that guides the order in which words are retrieved. One way researchers have investigated this is to focus on how speakers begin their utterance. Under the uncontroversial assumption that sentences are produced incrementally, the *starting point* of a sentence constrains both the form and content for the remainder of that utterance (MacWhinney, 1977). Not only do starting points guide word order for simple phrases like conjoined noun phrases (CNPs), but they also constrain syntactic choice when expressing more complex messages, such as locative descriptions. For example, when describing a picture of a cat next to some trees, a speaker can choose to begin their description with *the cat*. By doing so, they then construct a locative construction with a singular form of the verb *be* and end their utterance talking about *the trees* (producing *the cat is next to the trees*). In contrast, if they choose to begin their utterance with *the trees*, they must then construct an active construction with a plural form of the verb *be* and end their utterance talking about *the cat* (producing *the trees are next to the cat*).

A variety of factors can influence adults' starting points, such as the speaker's visual attention (e.g., Gleitman, January, Nappa & Truesewell, 2007), and investigations of these have informed us on how they produce language. Indeed, there is a link between what adults look at and their starting point (see Bock, Irwin & Davidson, 2004 for review). Here, we report two experiments that examine how pre-schoolers choose their starting points.

Specifically, we investigate 1) whether there is a link between what children look at and how they begin their utterances and 2) whether we can manipulate their starting points and discuss what these findings can tell us about how they plan their utterances.

Examination of how adults choose their starting points has revealed evidence for two distinct planning strategies underlying their language production, which adults can flexibly switch between using, depending on the task (e.g., Konopka & Meyer, 2014; van de Velde, Meyer & Konopka, 2014). The *structural incrementality* account emphasises top-down control of utterance planning and assumes that a speaker develops an overall structure for the sentence plan, which is imposed before speaking (Griffin & Bock, 2000; see Papafragou & Grigoroglou, 2019). According to this account, a speaker's starting point is driven by linguistic biases that influence the development of an overall structural plan, such as a bias to prioritise agents (likely to be animate objects) over patients (likely to be inanimate objects) in an event. These biases may then affect which object they attend to first during linguistic encoding (e.g., attending to animate object first rather than inanimate object; attending to an agent first rather than patient), which then affects the order a speaker fixates on objects in a scene. That is, the word order reflects the order of fixation, but the order of fixation is driven top-down by a structural plan. That is, speakers' first fixations may predict what they will talk about first – but those first fixations are themselves driven (hence predicted by) higher level factors (e.g., agency or spatial configuration; Bock et al., 2004; Kuchinsky, Bock & Irwin, 2011). Importantly, speakers do not begin lexical retrieval until *after* a sentence plan has been generated (proposed to be after the first 400ms; e.g., Griffin & Bock, 2000).

Adults can also use a *lexically incremental* strategy to plan their utterance whereby speakers begin talking about the most salient or accessible character first and build a sentence plan that is consistent with this starting point. According to this account, starting points are driven by bottom-up influences, such as visual saliency based on low level perceptual features (e.g., more colourful objects; Coco & Keller, 2014; 2015; Coco, Malcolm & Keller, 2014) and exogenous cues (e.g., Gleitman et al, 2007; Myachykov, Tomlin & Posner, 2005; Tomlin, 1997), as well as perceptual biases (e.g., bias to parse visual scenes from left to right; Christman & Pinger, 1997). This kind of strategy is consistent with language production models assuming that visual information immediately activates lexical information prior to the generation of a syntactic structure (Bock, 1982; Bock & Levelt, 1994). Thus, speakers begin lexical retrieval as soon as they fixate on an item or character, and, as a result, their starting point is not only predicted by their first fixations but can also be influenced by attention manipulations designed to affect these.

However, the strength by which such perceptual factors influence adults' starting points varies. For example, the left bias is strong in adults. Indeed, when producing simple conjoined noun phrases or locative phrases, adults often name objects in a left-to-right order, without explicit instruction to do so (e.g., Meyer, 1996; Meyer, Sleiderink & Levelt, 1998). Exogenous cues can also affect adults' starting points. For example, Gleitmen et al (2007) asked participants describe a range of events (including transitive events, but also scenes that could be described as a conjoined NP; *a dog and a* cat). Importantly, they subliminally cued one object over the other. They found that adults' first fixations were influenced by the cue, but so were their starting points (see also Myachykov, Tomlin & Posner, 2005; Tomlin, 1997). However, more recent work suggests that exogenous cues have relatively weak effects on adult' starting points, despite the fact that they affect first fixations (Kuchinsky & Bock, 2010; Hwang & Kaiser, 2009; Myachykov, Garrod & Scheepers, 2010; van de Velde et al., 2014). In fact, much work has shown that adult speakers extract the overall gist before speaking (Bock, Irwin & Davidson, 2004; Griffin & Bock, 2000; Kuchinsky, Bock & Irwin, 2011), but adults are more likely to begin their utterance with a cued character than an uncued character (and more generally use a lexically incremental rather than a structurally incremental planning strategy) in cases where an event is hard to interpret rather than when an event is easy to interpret (Kuchinsky & Bock, 2010; van de Velde et al., 2014; see also Konopka & Meyer, 2014), indicating that they can switch planning strategies depending on the task at hand. The ability to use both these strategies allows adults to speak fluently, striking a balance between communicating the information that they consider most noteworthy, and that which they can most easily linguistically encode.

In this paper, we are interested in how pre-schoolers choose their starting points and what this can tell us about how they plan their utterances. There are reasons to believe children will be more affected by bottom-up influences during production (and therefore, plan in a lexically incremental way). For example, infants often make use of various attentional cues, such as caregivers' gaze or pointing, to match names to referent objects during word learning tasks (Baldwin, 1995; Carpenter, Nagell, Tomasello, Butterworth & Moore, 1998)[1]. Furthermore, visual attention is tied to language comprehension. By the age of two, children focus on each referent in the order a sentence has been auditorily described (e.g., Yuan & Fisher, 2009). Thus, it seems plausible that pre-schoolers' production may be influenced by external influences that capture their attention. Indeed, an attentional cue may make sentence planning easier for children, as it highlights one object or character, which should therefore become more accessible, resulting in facilitated lexical retrieval. However, lexical incrementality requires cognitive flexibility: a speaker must be able to continue their

[1] Note that these findings have also been interpreted at a higher level, in that by attending to their caregiver, infants assume that that information is most important and so should be highlighted. This would therefore be consistent with structural incrementality.

utterance in a way that is congruent with their chosen starting point. This is likely to be more difficult for more complex utterances than simple utterances.

Studies investigating the link between attention and children's word order choice have failed to find bottom-up effects. For example, Ibbotson, Lieven and Tomasello (2013) asked children (three- and four-year-olds) and adults to describe scenes whereby the experimenter either directed their gaze to either the agent or patient of a scene (test condition), thus directing participants' attention to a character, or they looked directly towards the participant (control condition), and measured whether children used an active or passive construction (i.e., began their sentence with the agent or the patient). Interestingly, experimenter's eye gaze influenced sentence construction for four-year-olds but not three-year olds: Four-year-olds were more likely to produce a passive (and begin their utterance with the patient) if the experimenter looked at the patient compared to when the experimenter looked at the agent, and were more likely to produce an active (thus beginning their utterance with the agent) when the experimenter looked at the agent, suggesting that the four-year-olds' starting points were influenced by the attentional manipulation. However, three-year-olds' sentence construction was influenced by eye gaze only when the experimenter looked at the agent, and not the patient. That is, they were more likely to produce an active if the experimenter looked at the agent, but were not more likely to produce a passive when the experimenter looked at the patient. One interpretation of these findings is that older and younger children are affected by attention manipulations differently when planning their utterances: Older children are able to begin lexical retrieval regardless of the syntactic structure that they will consequently produce, but younger children's utterances are influenced by attention manipulations only when this cues an entity that they would prefer to be first-mentioned (e.g., a bias to begin their utterance with the agent) and/or to be subject (e.g., a bias to make the agent the subject). This could mean that older and younger children are using different planning strategies: four-year-olds are preferentially using a lexically incremental strategy whereby they begin lexical retrieval for wherever their gaze is directed, whereas three-year-olds are preferentially using a structurally incremental strategy that relies on biases.

However, Ibbotson et al (2013) failed to measure whether their attention manipulation was effective in capturing children's – especially younger children's – attention. It is possible that three-year-olds' attention was directed to the agent of the scene due to their own linguistic bias to attend to agents first, regardless of where the experimenter directed their gaze, resulting in them talking about the agent first in both the test and control conditions. Studies investigating adults' planning strategies track participants' eye movements from picture onset, and as a result, can measure the efficacy of any attention manipulation (e.g., Gleitman et al., 2007; Van de Velde et al., 2014). More recently, Brough, Branigan, Gambi & Rabagliati (2018) tracked three-to-four-year-olds' eye movements as they described pictures of events in order to investigate the influence of children's attention and linguistic biases on their gaze and starting points. They manipulated the animacy (e.g., human vs. shape) and visual saliency (e.g., grayscale vs. colourful) of the characters depicted in the event and found that although children fixated on the most salient character first, they nevertheless tended to begin their utterance with the animate character as sentence-initial subject (e.g., the man is chasing the triangle), suggesting that children used a planning strategy that favoured linguistic biases (i.e., structural incrementality) over one in which speakers choose a starting point based on whatever they look at first (i.e., lexical incrementality). However, it is possible that pre-schoolers are more susceptible to planning strategies that rely on linguistic biases when they describe events (and therefore have to generate semantic roles, determine grammatical relations and produce a complex constituent structure). Thus, it may be difficult to detect an influence of attention on word order choice in these studies. It may only be possible to find effects of attentional manipulations on word order in the to-be-produced utterance for simple utterances that have minimal semantic roles.

2.1.1. The Current Study

Here, we investigate whether children's starting points are determined by their pattern of visual attention, and specifically whether their first fixations predicts their word order choice. In Experiment 1, we do this for stimuli that do not involve any semantic roles and require a minimal syntactic structure, specifically a conjoined noun phrase like *a car and a doll*. In Experiment 2, participants described the relative location of the objects (e.g. *a car is next to a doll*), requiring them to determine minimal semantic roles (i.e., figure and ground) and associated grammatical relations (subject, oblique object), and generate a more complex locative structure (NP VP[V PP])[2]. To further determine whether any tendency for their starting points to reflect their visual attention is driven by bottom-up factors, specifically visual properties of an object, rather than top-down factors, we used an attention manipulation which cued one object over another (as previous studies have successfully used in adults; Gleitman et al., 2007). By using a cue manipulation, we were able to causally test whether their starting points were influenced by attention allocation. That is, we are not just tracking where they fixate first, but manipulating where they fixate first. If children can plan in a bottom-up lexically incremental way, children's first fixations should predict their starting point, and more importantly, the cue should also predict their starting point, so that the same visual stimulus would elicit different starting points as a function of the visual cue.

We also test adult controls. However, as exogenous cue manipulations are known to have a relatively weak effect on starting points in adults (in contrast to a strong left bias), we

[2] Note that we present graphs for both experiments in the same figure to save space

do not expect them to show the same pattern of behaviour as children. In particular, we expect to find a weaker relationship between cue and word order choice, and instead a stronger left-right bias that might also influence first fixations.

2.2. Experiment 1a

2.2.1 Method

Participants: Participants were 30 adult native English speakers (M= 23.13years, range = 19-32, 17 female) & 30 native English-speaking children aged between 3 and 4 (M=46.33months, range = 36-58, 13 female)[3]. We recruited most children from private nursery schools around Edinburgh, and the rest from a database of interested families. Ethnicity and SES were not recorded, but were representative of the area (almost entirely white, predominantly from middle-class Scottish families).

Materials The experimental items comprised 30 pairs of pictures and an associated subliminal cue (a red star 445 x 445 pixels). To create our 30 picture pairs, we chose 60 words with an average age of acquisition (AOA) of less than 3 years (Johnston, Dent, Humphreys & Barry, 2010; mean AOA = 2.31) and paired each word with another that had a similar AOA (e.g., *car* and *doll* had the same AOA of 1.83 and were paired together). We then chose child-friendly pictures of each word. Each picture ranged in size from 155 x 480 pixels (e.g., carrot) to 480 x 480 pixels (e.g., clock). These pictures were then spliced together, with one object appearing on the left and the other on the right; for each pair, we created two versions, so that the same pictures appeared in reversed orientations (e.g., LEFT: car/RIGHT: doll; LEFT: doll/RIGHT: car). Cue position was counterbalanced across items so

[3] Note that this is the same sample as used in the study reported in Chapter 3; Experiment 1b.

that the cue appeared on the left in 15 items and on the right in 15 items. We created two lists, each containing one orientation of each picture pair (e.g., for the item car-doll/LEFT CUE, in list one the car appeared on the left of the screen and was cued; in list two, the doll appeared on the left of the screen and was cued); this was to avoid any confound between individual objects and cue locations.

Procedure Prior to the study, participants were told that they would see pictures on the screen, and were instructed to name each picture using a conjoined noun phrase as quickly as possible (*a car and a doll*)[4]. The study began with 4 practice trials. If participants didn't know what to say or they made a mistake, the experimenter would produce a correct version of the utterance. Participants then completed 30 trials, with a short break after 15 trials. During the experimental session, the experimenter would prompt children to look at the screen if they became distracted by saying *What do you see here? Can you tell me what these pictures are?*

Each trial began with a central fixation point, and then a cue (a red star) was presented on either the left or right side of the screen for 75ms (Fig 1). Participants then saw two pictures on the left and right side of the screen, and named them as instructed. The experimenter ended each trial once the participant had named the pictures. If participants did not know the name for the items, the experimenter moved on to the next trial. Order of trials was individually randomized.

---

[4] We asked pre-schooler participants to name the pictures like "*a cat and a dog.*" During the practice rounds, if pre-schoolers did not produce the NP Conjunctions (e.g., "*a cat dog*"), the experimenter would repeat the pictures to the child as a NP Conjunction (correct version) and ask them to talk about the pictures in the same way that they did.

Stimuli were presented using SMI's Experiment Centre on a laptop fitted with a REDn Scientific eye-tracker (SensoMotoric Instruments GmbH, www.smivision.com). The tracker was calibrated once at the start of the session and once after 15 trials using a 2-point grid. It recorded fixations binocularly at 30Hz, but we only analysed right-eye fixations. Sessions lasted approximately 10-15 minutes.

Coding. Participants' utterances were transcribed, and scored as to whether they named the left or right item first. Trials were discarded if participants did not produce a response (Adults: 0.003%, Children: 5%). We also coded the time at which participants initiated their utterance (defined as when the participant began naming the pictures, including hesitations such *uhm*), the onset of the first determiner and onset of the first named object. We removed trials where participants took more than 5000ms to begin speaking or were 1.5 standard deviations below the age-appropriate mean first named object onset time (<1% of data points for each age group).

We analysed participants' eye movements for the time period in which the images were presented on screen. Using BeGaze software (Version 3.6), we established two areas of interest that covered the entire left- and right-hand sides of the screen, respectively. We coded whether fixations were directed to one of those AOIs, offscreen, or could not be measured due to trackloss. We removed observations on which no gaze was recorded (due to looking away or tracking ratio < 60: Adults: 1.0%; Children: 20.7%)[5].

[5] Tracking ratio is a measure the eye-tracker calculates and determines how precise the eyetracker was at tracking a participant's eyes. Greater tracking ratio implies that the eye-tracking results are more accurate (due to greater calibration, participant not moving their head, etc).

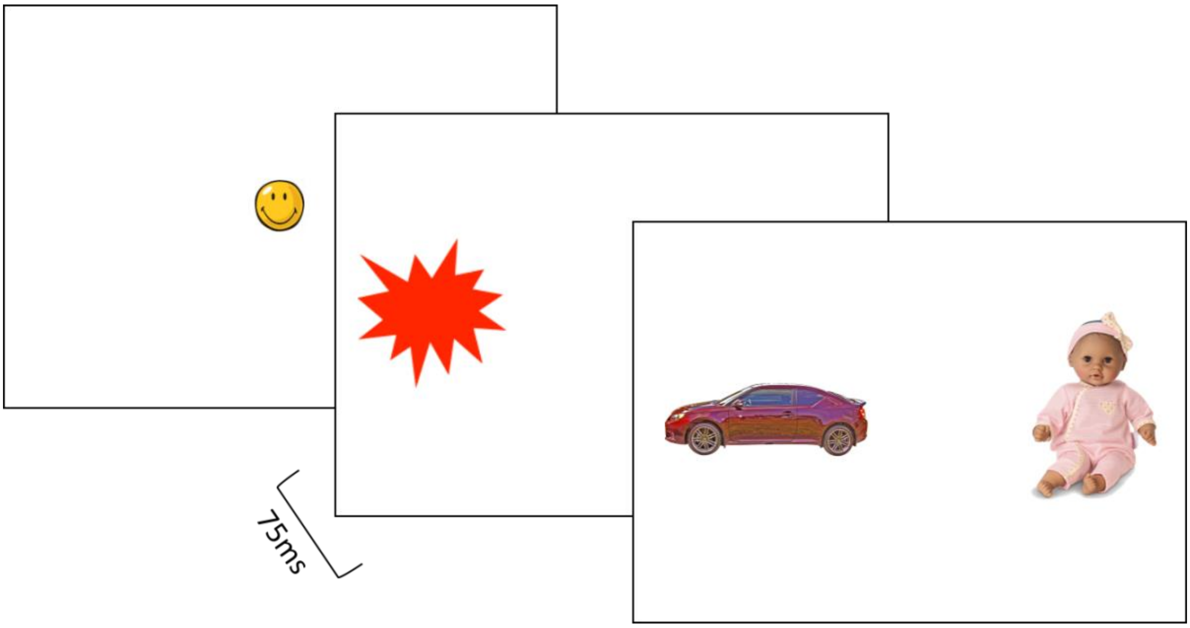

Figure 2. Example trial for Experiments 1 and 2. The cue appeared on either the left or right of the screen an equal number of times throughout the experiment. Participants were asked to describe the pictures as quickly as they could. In Experiment 1, participants described the images as a conjoined noun phrase (a car and a doll"); in Experiment 2, they produced a locative sentence (the car is next to the doll).

Data Analysis. We analysed participants' utterances and eye movements over the trial. We used logistic mixed-effects regressions to analyse binomial data, such as the picture that participants first fixated and the picture that participants first named. We also analysed the (log-transformed) response time to name the first item, using linear mixed effects models. Specific details of each model are given in the relevant section, using lme4 syntax. All models included participants and items as random effects (Baayen, Davidson & Bates, 2008), and incorporated maximal random effects structure (Barr, Levy, Scheepers & Tily, 2013), except when otherwise specified. In cases where the model would not converge, correlations among random effects were fixed to zero (Bates, Kliegl, Vasishth, & Baayen, 2015), and in cases of singularity, we used a partially Bayesian approach using the *blme* package – an extension for *lme4* (Chung, Rabe-Hesketh, Dorie, Gelman & Liu, 2013). Unless otherwise specified, all predictors were contrast coded (-1, 1). For the GLMM models, we report coefficient estimates (*b*), standard errors (*SE*) and *z* and *p* values for each predictor; 95% confidence intervals (CI) are from the *confint* function (method=Wald). For LMM models, we report coefficient estimates (*b*), standard errors (*SE*) and *t* values, *p* values are from log-likelihood ratio tests and so we also report χ^2 values; 95% confidence intervals (CI) are from the *confint* function (method=Wald).

A notable feature of this data is that many adult participants were at ceiling level in looking to the left object first in some cells of our experimental design, which resulted in separation in our data. Separation can make it difficult for logistic models to estimate an effect due to their being so little variability within the data. To counteract this, we also ran Bayesian mixed models fitted with a Bernoulli distribution using the *brms* package alongside our glmer models (Version 2.12.0; Bürkner, 2016). We ran four chains per model, each for 2,000 iterations, with a warm-up period of 1,000 iterations and initial parameter values set to zero[6]. For each parameter, we report an estimate (*b*), estimated error (EE), and the 95% credible interval (CrI). If zero lies outside the CrI, then we conclude that there is sufficient evidence to suggest the estimate is different from zero (i.e., there is an effect). We also calculated Bayes Factors (BF) by comparing the null model (i.e., that without the predictor) with the full model. Following Dienes (2014), we interpret a Bayes Factor of greater than 3 as strong evidence for the alternative hypothesis over the null, less than 1/3 as strong

[6] For all Bayesian mixed models reported in this chapter, we set normal priors for all priors; for our standard deviations of random effects, we set a mean of 0 and sd of 1 (normal(0, 1) and for our coefficients, we set a mean of 0 and an sd of 2.5 (normal(0, 2.5).

evidence for the null hypothesis over the alternative, and anything between these values as support for neither the alternative nor the null.

2.2.2. Results

Manipulation Check. We first analysed whether the subliminal cue influenced which picture participants first looked at (and also whether there was an overall left-right bias in where participants looked), by testing whether participants were more likely to first fixate the left item when it was cued. We recorded which of the two areas of interest participants fixated first (defined as the first eye tracked sample within the first 500ms; Left Object vs., Right Object). Figure 2A shows the proportion of looks to the left object as a function of cue in both children and adults.

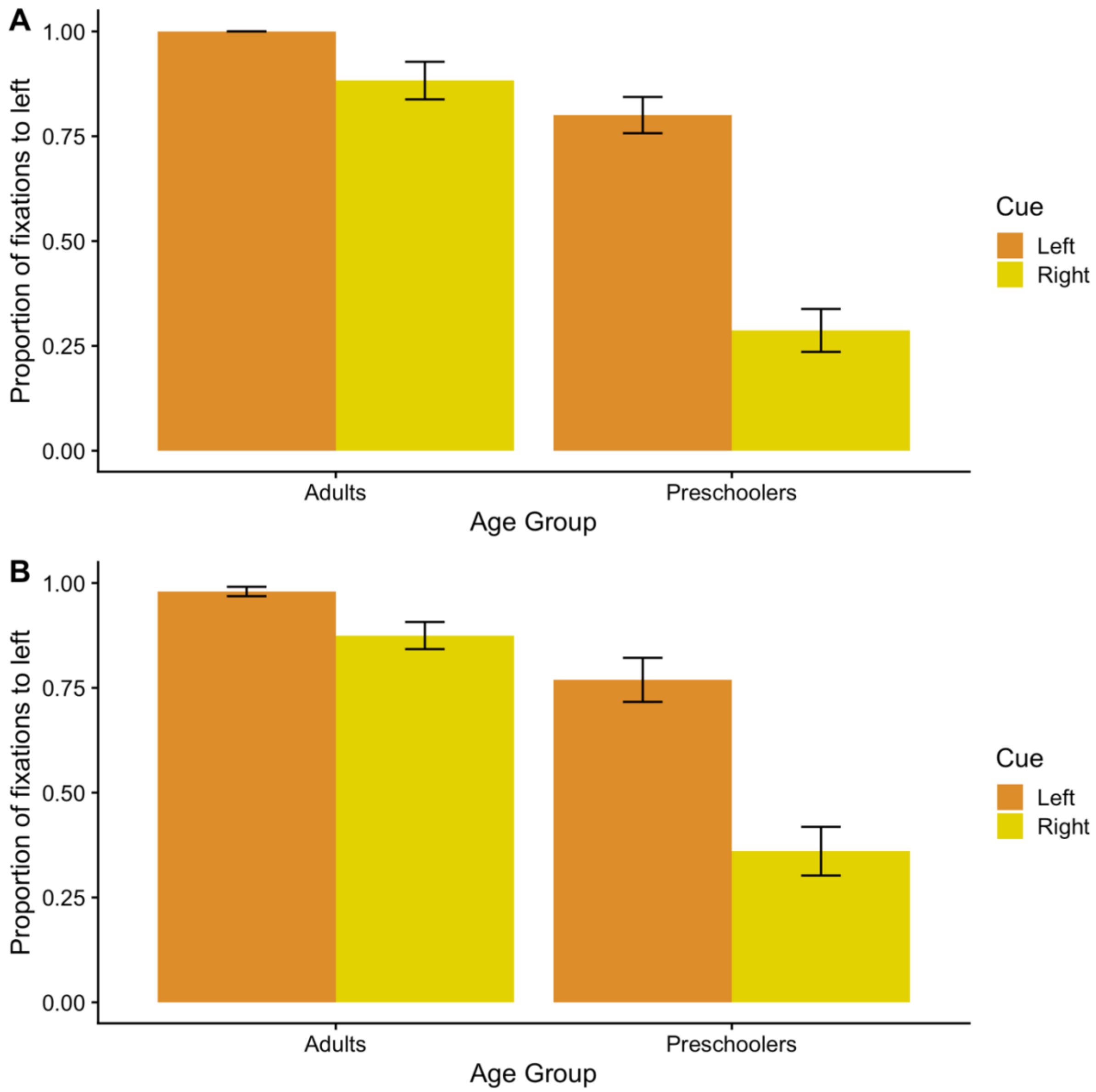

Figure 3. Mean proportion of looks to the left item within the first 500ms of picture onset in Experiment 1 (A) and Experiment 2 (B) as a function of whether the cue appeared on the left or the right of the screen for each age group. Bars indicate ±1 standard error.

Table 4 reports the logistic model output for each of our analyses and Table 5 reports the Bayesian mixed model output. For all our analyses, we first tested whether pre-schoolers first fixations were predicted by our cue manipulation (Left First Fixations ~ Cue) and then ran a combined model to looking at age differences (i.e., do children behave differently from adults; model form: Left Fixations ~ Cue * Age Group).

Overall, children were more likely to fixated on the left object first after the cue appeared on the left of the screen than the right (90% [SD = .24] vs. 29% [SD = .28]), suggesting that their first fixations were influenced by our manipulation cue. Furthermore, our combined Bayesian analysis found a main effect of cue, suggesting that, overall, participants were more likely to first fixate on the left item after a left cue than a right cue (90% [SD=.19] vs. 59% [SD=.40]), and a main effect of age, suggesting that adults were more likely to fixate on the left item first than children were (adults: 94% [SD=.12] vs. children: 56% [SD=.23]). Furthermore, we did not find a reliable interaction between age and cue, suggesting that both adults and children's first fixations were affected by the cue.[7]

[7] Note that our GLMER model did not converge and so we only report our Bayesian Regression model

Table 4 Logistic Mixed Effects Model output for children and adults' first fixations in both experiments. Note we could not converge the combined analysis model in Experiment 1, and so we do not have results for this.

		Experiment 1					Experiment 2				
	Predictor	**B**	**SE**	**Z**	**CI**	***p***	**B**	**SE**	**Z**	**CI**	***p***
Child Model	Cue	1.55	.18	8.67	[1.19, 1.89]	<.001	1.23	.21	5.90	[.82, 1.64]	<.001
Combined Analysis	Cue	-	-	-	-	-	1.47	.34	4.31	[.80, 2.13]	<.001
	Age Group	-	-	-	-	-	1.89	.29	6.36	[1.31, 2.47]	<.001
	Cue * Age Group	-	-	-	-	-	0.10	.26	.39	[-.40, .60]	.70

Table 5. Bayesian Regression Model ouput for children and adults' first fixations in both experiments. All models occurred with no divergent transitions (all $\hat{R}$ s ≤ 1.01).

		Experiment 1				Experiment 2			
	Predictor	**b**	**EE**	**CrI**	**BF**	**b**	**EE**	**CrI**	**BF**
Child Model	Cue	1.60	.22	[1.16, 2.06]	8824250	1.19	.20	[.82, 1.62]	45075
Combined Analysis	Cue	2.57	.71	[1.43, 4.31]	1147911	2.34	.40	[1.60, 3.17]	8004.01
	Age Group	3.19	.59	[2.22, 4.58]	2738654930981	3.38	.48	[2.49, 4.38]	106617896
	Cue * Age Group	.69	.55	[-.17, 2.06]	.15	-.04	.67	[-1.30, 1.29]	.03

Choice of Word Order. Our next analysis focused on the relationship between children's gaze and the word order of their utterances. We first tested whether participants were more likely to name the object that they fixated on first (i.e., their first fixation within the first 500ms of a trial). Table 6 reports the results from our logistic model output for our model testing whether pre-schoolers were more likely to name the object they fixated furst (model form: Left Named First ~ Left First Fixations) and our combined analysis looking at age group differences (model form: Left Named First ~ Left First Fixation * Age Group), and Table 7 reports the output from the Bayesian mixed models. Figure 4 shows the proportion of trials where participants named the left object first when they fixated on the left object first versus when they fixated on the right.

Both our logistic and Bayesian models indicated that pre-schoolers were more likely to talk about the left item first if they fixated on the left item first than if they did not fixate on the left item first (73% [SD=.25] vs. 36% [SD=.28]). Furthermore, both our logistic and Bayesian models from our combined analysis found a main effect of first fixations to the left item, suggesting that, overall, participants were more likely to name the left item first if they fixated on the left item first than if they fixated on the right item first (85% [SD=.22] vs. 37% [SD=.34]). Again, the model indicated a main effect of age, suggesting that adults were more likely to name the left item first than children were (94% [SD=.13] vs. 53% [SD=.21]). Finally, we found a significant interaction between fixations to the left item and age: When participants fixated on the left item first, adults were more likely to name the left item first than children were (B =1.71, SE= .23, z= 7.52, CI= [1.27, 2.16], p<.001). However, when participants did not fixate on the left item first, there were no age differences in the likelihood of naming the left item first (B =.14, SE=.31, z= .42, CI= [-.49, .76], p=.67). Overall, this analysis suggests that both children and adults tended to name the object they fixated on first.

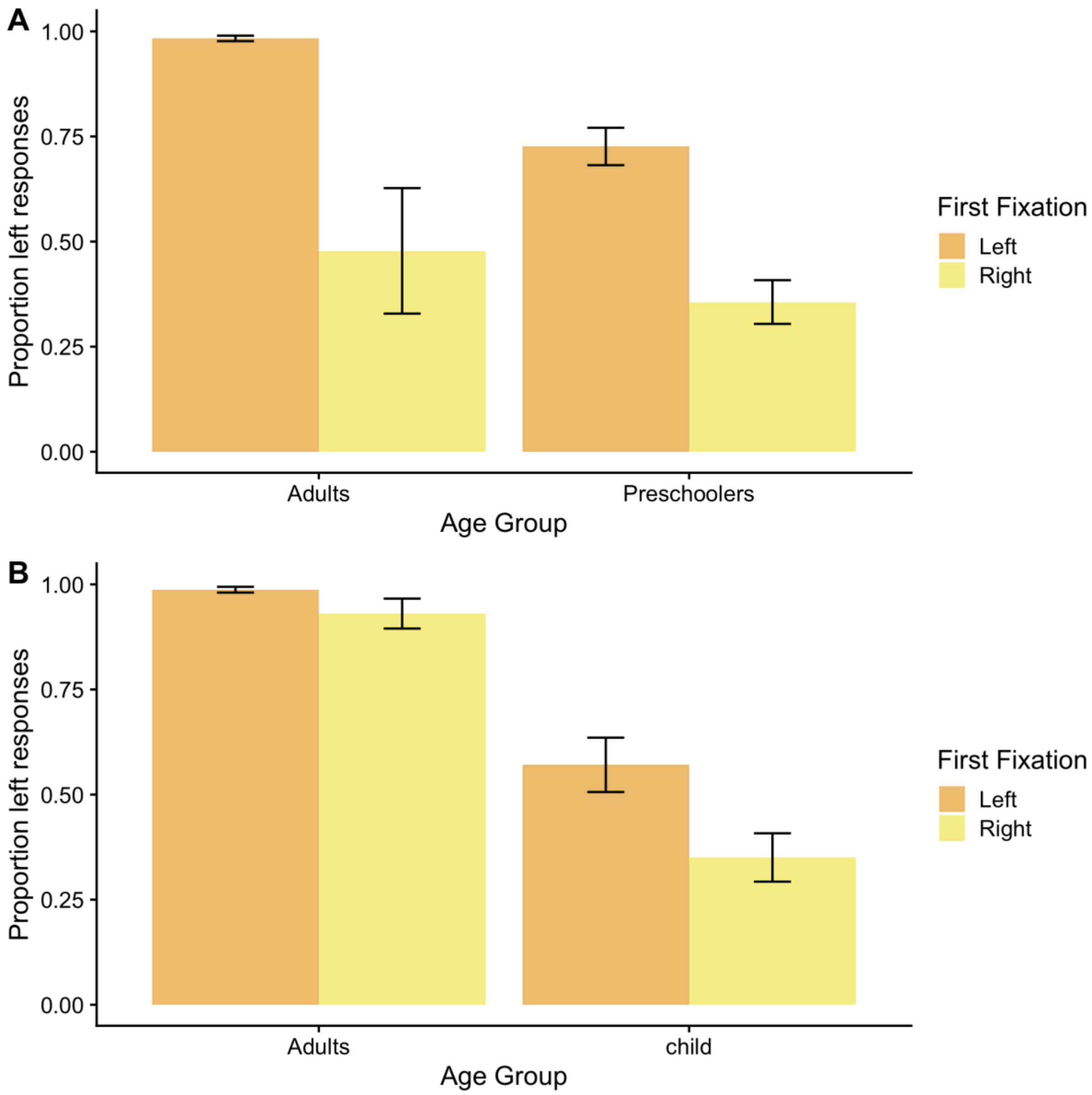

Figure 4. Mean proportion of responses beginning with the left object in Experiment 1 (A) and Experiment 2 (B) as a function of whether participants' first fixation (within the first 500ms) was to the the left or the right of the screen for each age group. Bars indicate ±1 standard error.

Our second analysis focused on whether children's starting points can be driven bottom-up by low-level perceptual factors, and not inherent semantic properties of objects or an overall perceptual bias for talking about scenes left-to-right. We tested whether participants were more likely to name the left item first after a left cue than a right cue. Table 8 reports the output for our logistic model for our model looking at pre-schoolers (model form: Left Named First ~ Cue) and our combined analysis looking at whether pre-schoolers behaved differently from adults (model form: Left Named First ~ Cue * Age Group), and Table 9 reports the output from our Bayesian models (following the same model form). Figure 5 shows the proportion of trials where participants named the left object first when the cue appeared on the left versus the right.

For both models, children were more likely to name the left item first after a left cue than a right cue (64% vs. 41%), suggesting that the cue did affect children's starting points. Furthermore, our combined analysis found both a main effect of cue and age on participants' starting points; overall, participants were more likely to name the left item first after a left cue than a right cue (81% [SD=.24] vs. 65% [SD=.34]); and adults were more likely to name the left item first than children were. Finally, there was no interaction between cue and age suggesting that the cue affected both children and adults' starting points.

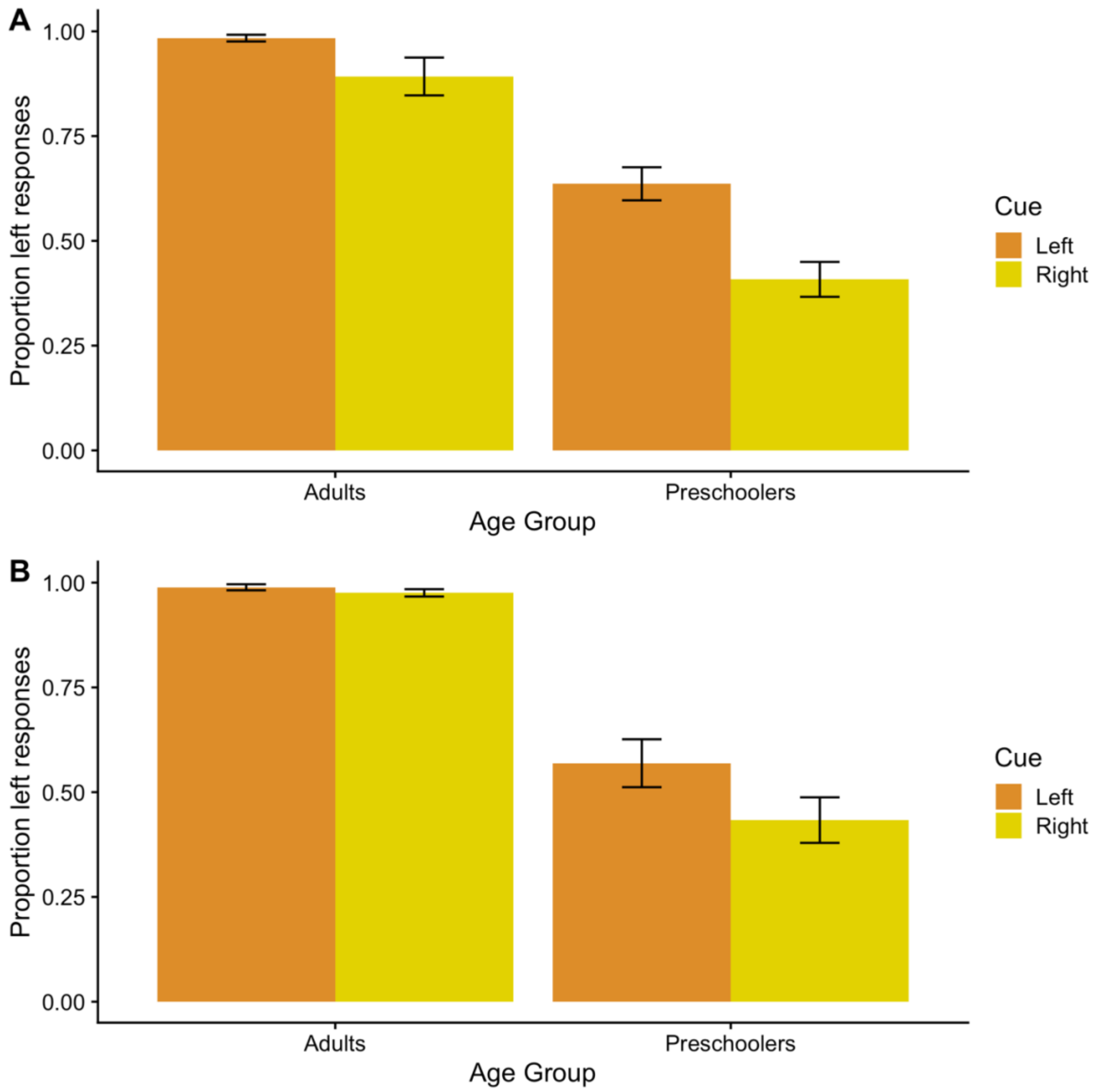

Figure 5. Mean proportion of responses beginning with the left object in Experiment 1 (A) and Experiment 2 (B) as a function of whether the cue appeared on the left or the right of the screen for each age group. Bars indicate ±1 standard error.

Table 6. Logistic Mixed Effects Model output for children and adults' likelihood of naming the left item first in both experiments as a function of participants' first fixations.

	Predictor	Experiment 1					Experiment 2				
		B	**SE**	**Z**	**CI**	***p***	**B**	**SE**	**Z**	**CI**	***p***
Child Model	Left First Fixation	1.12	.18	6.37	[.77, 1.46]	<.001	.51	.19	2.68	[.14, .88]	<.01
Combined Analysis	Left First Fixation	2.57	.43	6.00	[1.52, 2.35]	<.001	1.04	.28	3.70	[.49, 1.60]	<.001
	Age Group	2.03	.48	4.25	[.96, 1.78]	<.001	2.47	.32	7.70	[1.84, 3.09]	<.001
	Left First Fixation * Age	1.12	.33	3.37	[.36, 1.02]	<.001	.42	.24	1.79	[-.04, .88]	.07

Table 7. Bayesian Mixed Effects Model output for children and adults' likelihood of naming the left item first in both experiments as a function of participants' first fixations. All models occurred with no divergent transitions (all $\hat{R}$s ≤ 1.01).

	Predictor	Experiment 1				Experiment 2			
		b	**EE**	**CrI**	**BF**	**b**	**EE**	**CrI**	**BF**
Child Model	Left First Fixation	1.06	.16	[.75, 1.40]	553937	.49	.19	[.12, .87]	5.37
Combined Analysis	Left First Fixation	1.94	.24	[1.49, 2.45]	286635337	.84	.28	[.27, 1.38]	2039.33
	Age Group	1.42	.22	[1.03, 1.90]	1249353717	2.32	.30	[1.76, 2.92]	16368564180
	Left First Fixation * Age Group	.65	.20	[.28, 1.04]	50.33	.23	.23	[-.23, .67]	0.70

Table 8. Logistic Mixed Effects Model output for children and adults' likelihood of naming the left item first in both experiments as a function of our attention manipulation.

		Experiment 1					Experiment 2				
	Predictor	**B**	**SE**	**Z**	**CI**	***p***	**B**	**SE**	**Z**	**CI**	***p***
Child Model	Cue	.59	.10	6.08	[.40, .78]	<.001	.41	.14	3.06	[.15, .68]	<.01
Combined Analysis	Cue	.93	.34	2.77	[.27, 1.02]	<.01	.54	.23	2.30	[.08, 1.00]	<.05
	Age Group	2.49	.38	6.55	[1.66, 2.75]	<.001	2.67	.32	8.43	[2.05, 3.29]	<.001
	Cue * Age Group	.28	.31	.91	[-.34, .37]	.36	.12	.21	.56	[-.29, .52]	.58

Table 9. Bayesian Mixed Effects Model output for children and adults' likelihood of naming the left item first in both experiments as a function of our attention manipulation. All models occurred with no divergent transitions (all $\hat{R}$s ≤ 1.01).

		Experiment 1				Experiment 2			
	Predictor	**b**	**EE**	**CrI**	**BF**	**b**	**EE**	**CrI**	**BF**
Child Model	Cue	.59	.10	[.40, .79]	43143	.40	.13	[.15, .66]	5.99
Combined Analysis	Cue	.73	.19	[.36, 1.12]	140.04	.50	.22	[.09, .95]	4.57
	Age	2.17	.27	[1.68, 2.73]	81671398900707	2.63	.31	[2.07, 3.26]	1249434969383057
	Cue * Age	.10	.18	[-.25, .45]	.39	.09	.19	[-.28, .47]	1.09

Speech Onsets. Our next analyses focused on when our participants began naming the first item. For our first analysis, we tested whether first fixating on the named object facilitated naming times. If participants begin lexical retrieval when they fixate on the first named image, then they should be faster to begin speaking when they name the first fixated item first. This finding could support a lexical incrementality account. However, as a structural incrementality account predicts that speakers do not begin lexical retrieval until after they have developed a sentence plan, participants' first fixations should not affect their onset times. To test this, we first analysed pre-schoolers' onset times (with the form First Item Onset ~ First Fixated Item * Left Item Named First) followed by a combined analysis (First Item Onset ~ Left First Fixations * Left Item Named First * Age; see Table 10 for all models' output). Figure 6 shows the mean time for participants to begin naming the first item as a function of what they fixated on first.

Children were faster to begin naming the first item when they fixated on the left item first than when they fixated on the right item first (1755.72ms [SD=304.87] vs. 1897.66ms [SD=644.52]). We did not find a main effect of naming the left item first (left item named first = 1869.97 [SD=407.85] vs. right item named first = 1795.21 [SD=589.97]), but there was a significant interaction between naming the left item first and left first fixations: when children named the left item first, they were faster to begin naming when they fixated on the left item first than when they fixated on the right item first (1714.18ms [SD=387.68] vs. 2198.76ms [SD=907.30]: B = -.13, SE= .04, t= -3.62, CI=[-.20, -.06], $\chi^2(1)$= 10.74, p<.01), and when children named the right item first, they were slower to begin naming when they fixated on the left item first than the right item (1926.38ms [SD=562.47] vs. 1652.19ms [SD=559.58]: B = .07, SE= .03, t=2.55, CI=[.02, .13], $\chi^2(1)$= 6.46, p<.05).

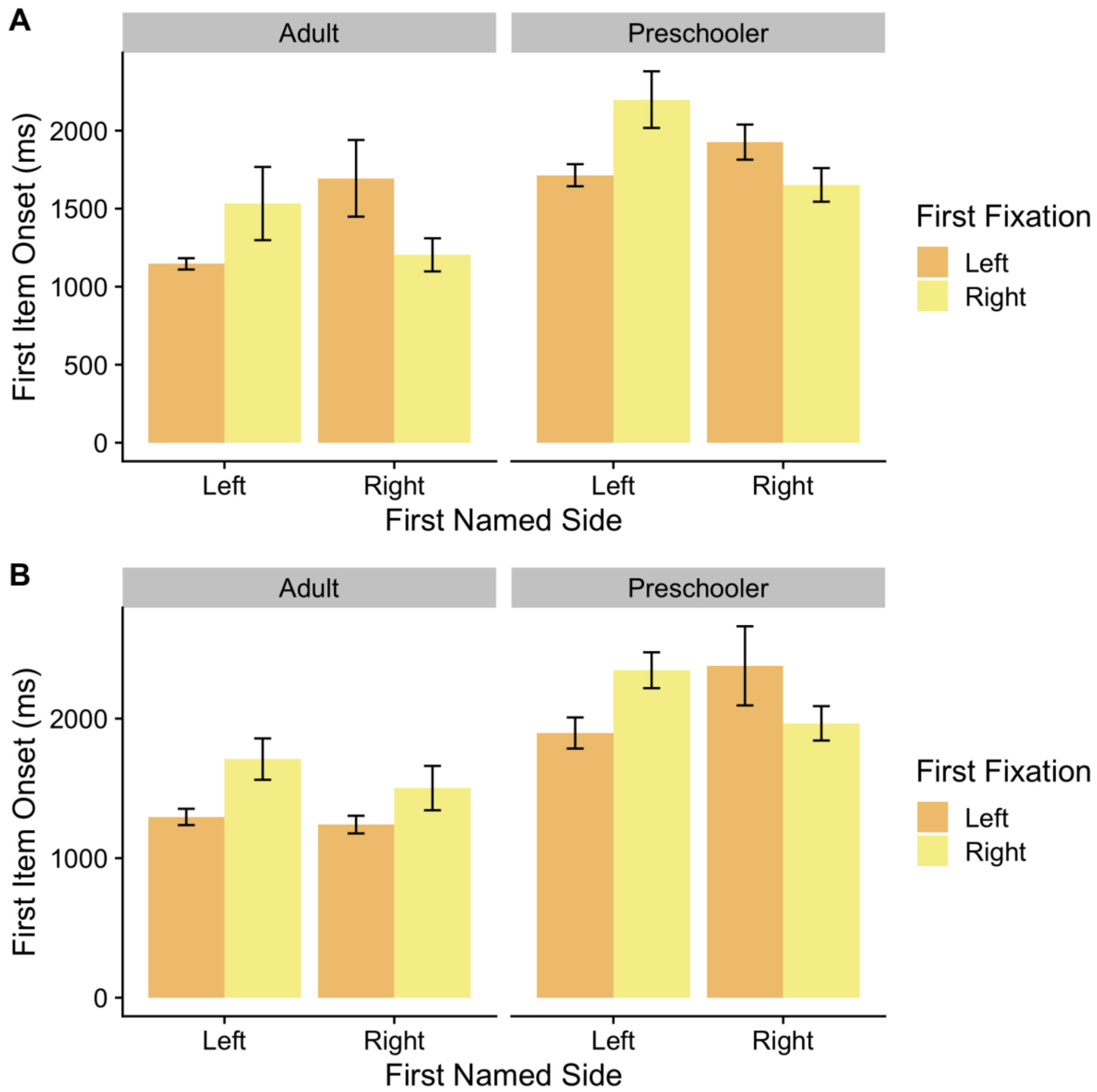

Figure 6. Mean speech onset times (the latency to begin naming the first item) as a function of whether the participants named the left or right image first and whether they fixated on the left (orange bars) or right of the screen first (yellow bars) in Experiment one (A) and Experiment 2 (B). Adults on the left and children on the right. Bars indicate ±1 standard error.

Our combined model indicated that, overall, participants were faster to name the first item when they fixated on the left item first than the right item (1459.45ms [SD=397.81] vs. 1765.01ms [SD=638.82]) and when they named the left item first than the right item (1517.21ms [SD=484.06] vs. 1738.96ms [SD=600.75]). Overall, adults were faster to begin naming the first item than children were (1162.51ms [SD=205.05] vs 1808.71ms [SD=341.22]). The model also indicated two significant interactions. Firstly, there was a significant interaction between fixating on the left item and naming the left item first: when participants named the left item first, they were faster to begin speaking when they fixated on the left item first than if they fixated on the right item (1430.18ms [SD=418.58] vs. 2053ms [SD=888.92]: *B* = -.14, *SE*= .03, *t*=-4.70, *CI*=[-.20, -.08], $\chi^2(1)$=15.32, *p*<.001), and when they named the right item first, they were slower to begin speaking if they fixated on the left item first than if they fixated on the right item (1875.65ms [SD=579.82] vs. 1570.68 [SD=543.88]: *B* = .07, *SE*= .03, *t*=2.72, *CI*=[.02, .12], $\chi^2(1)$=7.26, *p*<.01). We also found a significant interaction between age and naming the left item first: Adults were marginally faster to begin naming when they named the left item first than the right item first (1152.68ms [SD=205.50] vs. 1551.45 [SD=633.35]; *B* = -.10, *SE*= .05, *t*=-1.91, *CI*=[-.21, .00], $\chi^2(1)$=3.46, *p*=.06) whereas there was no difference in children's onset times when they named the left or right item first (1869.97ms [SD=407.85] vs. 1795.21ms [SD=589.97]: *B* = .01, *SE*= .02, *t*=.56, *CI*=[-.03, .05], $\chi^2(1)$=.29, *p*=.59). There was no significant interaction between age and fixating on the left item and there was no three-way interaction.

Our second analysis investigated the effect the cue had on naming times. To If children can use a lexically incremental strategy, meaning they begin lexical retrieval when they fixate on the first named image, then - given that our attention manipulation was successful - they should be faster to begin speaking when they name the cued item first.

However, if they are using a structurally incremental strategy, meaning they do not begin lexical retrieval until after they have developed a sentence plan, the cue should not affect their onset times. Figure 7 shows the mean speech onsets for when the cue appeared on the left or right of the screen and whether participants named the left or right image first. As in our previous analyses, we first ran a model testing whether children's onset times were affected by the cue (with the form First Item Onset ~ Cue * Left Item Named First) followed by a combined analysis testing for age group differences (First Item Onset ~ Cue * Left Item Named First * Age; see Table 11 for all models' output). If participants begin lexical retrieval as soon as they fixate on an object, our model should indicate a significant interaction between cue and naming the left object first: participants should be faster when they name the left object first when the cue appears on the left than the right.

Our cue manipulation did not have a main effect on children's naming times (left cue = 1862.16ms [SD=491.49] vs., right cue = 1879.97ms [SD=407.51]), nor did naming the left item first. However, there was a significant interaction between cue and naming the left item first: when the cue was on the left, children were faster to begin naming the first item when they named the left object first than when they named the right object first (1819.25ms [SD=550.86] vs. 1958.37ms [SD=644.88]; $B = -.05$, $SE= .03$, $t= -2.09$, $CI=[-.10, -.003]$, $\chi^2(1)=4.39$, $p<.05$). When the cue was on the right, children were marginally slower to begin naming the left item first than if they named the right (1979.72ms [SD=462.59] vs. 1807.94ms [SD=523.76]: $B = -.04$, $SE= .02$, $t= -1.79$, $CI=[-.004, .08]$, $\chi^2(1)=3.22$, $p=.07$). This suggests that naming the cued object facilitated pre-schoolers object naming.

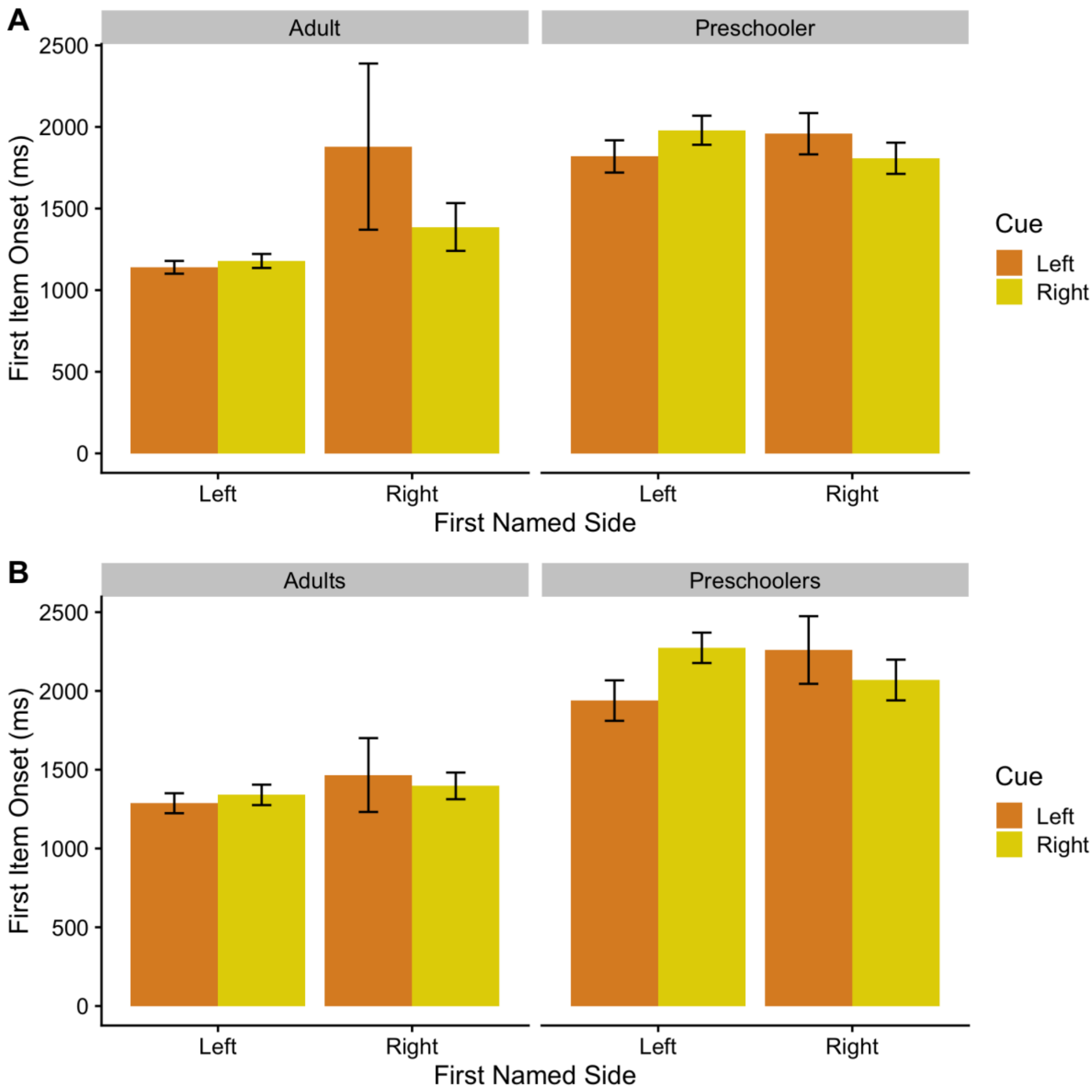

Figure 7. Mean speech onset times (the latency to begin naming the first item) as a function of whether the participants named the left or right image first and whether the cue appeared on the left (orange bars) or right of the screen (yellow bars) in Experiment one (A) and Experiment 2 (B). Adults on the left and children on the right. Bars indicate ±1 standard error.

Finally, our combined model showed that, unsurprisingly, adults were faster to begin speaking than children (replicating our combined model looking at first fixations). Again, we found a main effect of naming the left item first: overall participants were also faster to begin speaking when they named the left object first than when they named the right object first. Furthermore, the cue did not influence participants' onset times (left cue = 1510.37ms [SD=525.10] vs. right cue = 1531.33ms [SD=478.45]). The model also indicated several significant interactions: There was a significant interaction between naming the left item first and cue: When the cue was on the left, participants were faster to begin naming when they named the left item first than the right item first (1485.16ms [SD=539.60] vs. 1950ms [SD=653.80]: *B* = -.12, *SE*= .03, *t*= -3.44, *CI*=[-.18, -.05], $\chi^2(1)$=8.60, *p*<.01). However, when the cue was on the right, participants showed no difference in beginning naming when they named the left item first than the right item first (1564.95ms [SD=539.85] vs. 1719.32ms [SD=526.97]: *B* = -.02, *SE*= .02, *t*= -.93, *CI*=[-.07, .03], $\chi^2(1)$=.74, *p*=.39). These results suggest that naming the left object first when it was cued facilitated naming, regardless of age. We also found a significant interaction between age and naming the left item first (replicating our combined model looking at first fixations results: *B* = -.06, *SE*= .02, *t*= -2.43, *CI*=[-.11, -01], $\chi^2(1)=5.89$, $p<.05$): adults were marginally faster to begin naming when they named the left item first than the right item (*B* = -.10, *SE*= .06, *t*= -1.85, *CI*=[-.21, .006], $\chi^2(1)$=3.28, *p*=.07) whereas children showed no difference in naming times regardless of whether they named the left or right item first or not (*B* = .004, *SE*= .02, *t*= -.19, *CI*=[-.04, .04], $\chi^2(1)$=.04, *p*=.84). We also found an unexpected interaction between cue and age. Further investigations looking at the direction of these interactions showed that adults were faster to begin speaking than children after both a left (1146.85ms [SD=227.13] vs. 1862.16ms [SD=491.49]: *B* = -.23, *SE*= .03, *t*= -7.75, *CI*=[-.29, -.17], $\chi^2(1)$=31.33, *p*<.001) and a right cue (1182.67ms [SD=219.48] vs. 1879.97ms [SD=407.51]: *B* = -.28, *SE*= .03, *t*= -8.05, *CI*=[-.28, -.17], $\chi^2(1)$=31.52, *p*<.001). The three-way interaction between cue, age and left named first was not significant.

Table 10. Linear Mixed Effects Model output for children and adults' log transformed onsets for naming the first item in both experiments as a function of participants' first fixations and what they named first.

		Experiment 1					Experiment 2				
	Predictor	**B**	**SE**	**t**	**CI**	***p***	**B**	**SE**	**t**	**CI**	***p***
Child Model	Intercept	7.45	.04	191.86	[7.37, 7.53]		7.52	.04	180.99	[7.44, 7.60]	
	Left First Fixation	-.08	.03	-3.24	[-.14, -.03]	$\chi^2(1)=9.68$, $p<.01$	-.06	.03	-2.20	[-.11, -.01]	$\chi^2(1)=4.83$, $p<.05$
	Left Item Named First	-.04	.02	-1.64	[-.08, .01]	$\chi^2(1)=2.46$, $p=.12$	-.02	.02	-.93	[-.07, .03]	$\chi^2(1)=.85$, $p=.36$
	Left Item Named First * Left First Fixation	-.09	.02	-5.27	[-.12, .06]	$\chi^2(1)=27.03$, $p<.001$	-.09	.02	-4.23	[-.13, -.05]	$\chi^2(1)=17.45$, $p<.001$
Combined Analysis	Intercept	7.21	.03	267.31	[7.16, 7.27]		7.27	.03	234.53	[7.21, 7.33]	
	Left First Fixation	-.06	.02	-2.93	[-.11, -.02]	$\chi^2(1)=7.47$, $p<.01$	-.08	.02	-3.82	[-.12, .04]	$\chi^2(1)=12.87$, $p<.001$
	Left Item Named First	-.09	.03	-3.03	[-.14, -.03]	$\chi^2(1)=8.68$, $p<.01$	.01	.04	.21	[-.06, .08]	$\chi^2(1)=.05$, $p=.82$
	Age	-.18	.03	-7.29	[-.23, -.14]	$\chi^2(1)=41.80$, $p<.001$	-.19	.03	-7.02	[-.25, -.14]	$\chi^2(1)=43.51$, $p<.001$
	Left First Fixation * Age	.01	.02	.32	[-.03, .04]	$\chi^2(1)=.11$, $p=.74$	-.01	.02	-.78	[-.05, .02]	$\chi^2(1)=.53$, $p=.47$
	Left First Fixation * Left Named First	-.09	.02	-4.48	[-.13, -.05]	$\chi^2(1)=19.31$, $p<.001$	-.05	.03	-1.69	[-.11, .01]	$\chi^2(1)=2.76$, $p=.10$
	Age * Left Named First	-.04	.02	-1.94	[-.09, .00]	$\chi^2(1)=3.70$, $p=.05$	.02	.03	.83	[-.03, .08]	$\chi^2(1)=.71$, $p=.40$
	Left First Fixation * Age * Left Named First	-.01	.02	-.27	[-.04, .03]	$\chi^2(1)=.09$, $p=.76$	.02	.02	1.00	[-.02, .07]	$\chi^2(1)=.93$, $p=.33$

Table 11. Linear Mixed Effects Model output for children and adults' log transformed onsets for naming the first item in both experiments as a function of our attention manipulation and what they named first.

		Experiment 1					Experiment 2				
	Predictor	**B**	**SE**	**t**	**CI**	***p***	**B**	**SE**	**t**	**CI**	***p***
Child Model	Intercept										
	Cue	-.03	.03	1.11	[-.10, ,03]	χ^2(1)=1.27, *p*=.26	-.08	.02	-3.00	[-.13, -.03]	χ^2*(1)*=8.51, *p*<.01
	Left Item Named First	-.01	.02	-.49	[-.04, .02]	χ^2(1)=.25, *p*=.62	.02	.02	.83	[-.03, .06]	χ^2*(1)*=.72, *p*=.39
	Left Item Named First * Cue	-.04	.02	2.82	[-.08, -.01]	χ^2*(1)*=7.98, *p*<.01	.05	.02	-2.74	[-.09, -.01]	χ^2*(1)*=7.48, *p*<.01
Combined Analysis	Intercept	7.22	.03	266.95	[7.16, 7.27]		7.28	.03	239.46	[7.22, 7.34]	
	Cue	-.01	.02	-.26	[-.04, .03]	χ^2(1)=.06, *p*=.81	-.04	.02	-2.26	[-.08, -.01]	χ^2*(1)*=5.10, *p*<.05
	Left Item Named First	-.21	.02	-9.25	[-.25, -.16]	χ^2(1)=57.11, *p*<.001	-.22	.03	-7.95	[-.27, -.16]	χ^2*(1)*=46.28, *p*<.001
	Age	-.07	.03	-2.67	[-.13, -.02]	χ^2(1)=6.98, *p*<.01	-.08	.04	-.17	[-.08, .07]	χ^2*(1)*=.02, *p*=.88
	Cue * Age	.03	.01	2.45	[.001, .05]	χ^2(1)=6.08, *p*<.05	.03	.01	2.16	[.003, .05]	χ^2*(1)*=4.67, *p<.05*
	Cue * Left Named First	-.06	.02	-2.71	[-.11, -.02]	χ^2(1)=7.27, *p*<.01	-.02	.03	-.63	[-.08, .04]	χ^2*(1)*=.41, *p*=.52
	Age * Left Named First	-.06	.02	-2.43	[-.11, -.01]	χ^2(1)=5.89, *p<.05*	-.02	.03	-.56	[-.08, .04]	χ^2*(1)*=.30, *p*=.58
	Cue * Age * Left Named First	-.01	.02	-.67	[-.05, .03]	χ^2(1)=.50, *p*=.48	-.02	.02	-.89	[-.03, .07]	χ^2*(1)*=.78, *p*=.38

2.2.3. Discussion

Overall, our results from Experiment 1 suggest that children's production can be affected by bottom-up factors (in our case, by a subliminal cue). Children tended to begin their utterance with the object they fixated on first, and this occurred even when those fixations were manipulated by an external cue. Children's onset times were also affected by the object they fixated on first, even when those fixations were affected by the cue: they were faster to begin speaking when they named the object they fixated on first, which was often the cued object. These findings provide strong evidence that pre-schoolers can plan in a lexically incremental way when they produce an utterance with minimal semantic content (i.e., conjoined noun phrases).

Adults' production was also affected by our cue manipulation, although this was harder to estimate due to their strong left-right bias. This finding is unsurprising given the findings from previous studies showing that adults' starting points are only weakly affected by perceptual cues.

2.3. Experiment 2a

Our results from Experiment 1 show that children's production can be affected by bottom-up factors for conjoined noun phrases, thus providing support that children can use a lexically incremental planning strategy. However, these utterances did not involve semantic roles. It is possible that for more complex utterances, whereby a speaker must determine both semantic roles and grammatical constituents, that bottom-up factors will have less of an influence on a speaker's starting point. Indeed, previous studies have failed to find effects of attention manipulations on pre-schoolers starting points when they produce more complex utterances, such as transitive sentences.

In the next experiment, we ask if we will continue to see this relationship between visual attention and starting points for more complex utterances. Will children's starting points continue to reflect their first fixation, which is also affected by the visual cue, for utterances that require them to determine both semantic roles and grammatical relations? To test this, we asked participants to produce complete sentences expressing the relative location of the objects (e.g. *the car is next to the doll*), thus requiring participants to determine semantic roles and grammatical relations, and generate a more complex constituent structure than in Experiment 1.

2.3.1. Method

Participants. Participants were 30 adult native English speakers (M=24.04 years, range = 19-34, 15 female) & 42 native English-speaking children aged between 3 and 4 (M=45.79months range = 38-59, 18 female).[8] We recruited most children from private nursery schools around Edinburgh, and the rest from a database of interested families. Ethnicity and SES were not recorded, but were representative of the area (almost entirely white, predominantly from middle-class Scottish families).

Materials. We used the exact same materials as in Experiment 1.

Procedure. The procedure was the exact same as Experiment 1, but this time, we instructed participants to produce locative phrases (e.g., "the car is next to the doll").

Coding. We coded participants' data in the same way as in Experiment 1. We transcribed participants' utterances and scored as to whether they named the left or right item first. Trials

[8] Note that this is the same sample as used in the study reported in Chapter 3; Experiment 2b.

where participants produced a conjoined noun phrase or did not respond (Children: 6%, Adults: 0%) were discarded.

We also coded the time at which participants initiated their utterance (defined as when the participant began speaking about the pictures, including hesitations such *uhm*), the onset of the first determiner and onset of the first named object. We removed observations where participants took more than 5000ms to begin speaking or were 1.5 standard deviations below the age-appropriate mean (Adults: <1%; Children: 1.3%).

We analysed participants' eye movements for the time period in which the images were presented on screen. As in Experiment 1, we established two areas of interest that covered the entire left- and right-hand sides of the screen, respectively, and coded whether fixations were directed to one of those AOIs or not. We removed observations on which no gaze was recorded (due to looking away or tracking ratio < 60%: Adults: <1%; Children: 21.9%).

Analysis. We analysed the data in the exact same way as Experiment 1.

2.3.2. Results

Manipulation Check. As in Experiment 1, we first analysed whether the subliminal cue influenced which picture participants first looked at. We first ran our pre-schooler only model (model form: Left First Fixations ~ Cue), followed by a model looking at the effect of age (model form: Left First Fixations ~ Cue * Age; see Table 4 & 5 for logistic and Bayesian model output for both models).

We replicated our results from Experiment 1 (see Fig 2B). Overall, children were more likely to fixate on the left object first after the cue appeared on the left of the screen than the right (77% [SD = .27] vs. 36% [SD = .30]), suggesting that their first fixations were influenced by our manipulation cue. Furthermore, our combined analysis found a main effect of cue, suggesting that participants were more likely to first fixate on the left item after a left cue than a right cue (88% [SD=.22] vs. 63% [SD=.35]), and a main effect of age, suggesting that adults were more likely to fixate on the left item first than children were (92% [SD=.10] vs. 57% [SD=.24]). Finally, the interaction between cue and age group was not significant. Thus, our cue manipulation influenced participants' visual attention for both age groups and in both experiments.

Choice of Word Order. Our next set of analyses was concerned with the relationship between participants' gaze and their subsequent starting point. We first tested whether children's starting points were predicted by their first fixations (i.e., those sampled within the first 500ms; model form: Left Named First ~ Left First Fixation) followed by a combined analysis looking at the effects of age (model form: Left Named First ~ Left First Fixation * Age Group; See Table 6 & 7 for Logistic and Bayesian model outputs).

Children were more likely to name the left item first when the fixated on the left item first than when they fixated on the right item (57% [SD=.34] vs. 35% [SD=]), thus replicating Experiment 1's results (Fig 3B). Our combined model showed that overall, participants were more likely to name the left item first if they fixated on the left item first than if they fixated on the right item (79% [SD=.32] vs. 60% [SD=.38]); Adults were more likely to name the left item first than children (98% [SD=.04] vs. 50% [SD=.28]); and finally, there was a marginal interaction between left first fixations and age (B = .42, SE= .24, z= 1.79, CI= [-.04, .88], p=.07): Children were more likely to name the left object first if they fixated on the left image first than the right (B = .51, SE= .19, z= 2.68, CI= [.14, .88], p<.01) but we failed to find a significant effect for adults (although note there was a small numerical difference; B = 2.44, SE= 1.75, z= 1.40, CI= [-.99, 5.86], p=.16).

We next looked at the effect our attention manipulation had on participants' word order choice (see Fig 4B) by first testing whether pre-schoolers were more likely to name the left item first after a left cue than a right cue (Left Named First ~ Cue), and then running a combined model to look at the effect of age (Left Named First ~ Cue * Age Group; see Table 8 & 9 for all model outputs).

The cue predicted children's starting points: children were more likely to name the left item first after a left cue than a right cue (57% [SD=.32] vs. 43% [.30]), again replicating Experiment 1's results. The combined analysis replicated Experiment 1's combined model: overall, participants were significantly more likely to talk about the left item first when it was cued than when it was not (77% [SD=.31] vs. 70% [.35]); adults were more likely than children to name the left item first, regardless of cue; and the interaction between cue and age group was not significant, indicating that both pre-schoolers and adults' starting points were affected by our cue manipulation, as in Experiment 1.

Speech Onsets. Our final set of analyses focused on participants first item onset times. First, we tested whether participants were faster to name items that they fixated on first. As before, we conducted a pre-schooler only model (model form: First Item Onset ~ Left First Fixations * Left Named First) and a model that combined the data from each age group (model form: First Item Onset ~ Left First Fixations * Left Named First * Age; see Table 10 for all model output).

As in Experiment 1, children were faster to name the first item when they fixated on the left item first than the right item (2060.81ms [SD=786.41] vs. 2095.79ms [SD=449.09]). Naming the left item first did not predict first item onset (left named first = 2005.90 [SD=492.37] vs. right named first = 2127.06 [643.40]), but we did find a significant interaction between naming the left item first and first fixations: when children named the left item first, they were faster to begin naming when they fixated on the left item first than when they fixated on the right item (1896.82ms [SD=546.43] vs. 2346.58ms [SD=574.14]: *B* = -.09, *SE*= .03, *t*= -3.29, *CI*= [-.15, -.04], $\chi^2(1)$=10.00, p<.01); and when they named the right item first, they were slower to begin naming when they fixated on the left item first than when the fixated on the right item first (2377.91ms [SD=1359.61] vs. 1965.99 [SD=603.27]: *B* = .08, *SE*= .03, *t*= -2.34, *CI*= [.01, .14], $\chi^2(1)$=5.12, p<.05). Our combined analyses revealed a main effect of first fixations on first item onsets: overall, participants were faster to begin naming when they fixated on the left item first than the right item (1664.37ms [SD= 702.20] vs. 1918.65ms [SD= 569.35]); and a main effect of age: adults were faster to begin naming than children were (1310.56ms [SD=333.10] vs. 2060.86ms [SD=424.76]). Neither naming the left item first nor any of the interaction terms were significant.

Our final analyses tested whether naming the cued item affected naming times. We first ran a pre-schooler only models (model form: First Item Onset ~ Cue * Left Named First) and a final model combining both age groups to look at the effect of age (model form: First Item Onset ~ Cue * Left Named First * Age; see Table 11 for all model's output).

Children were faster to begin speaking after a left cue than a right cue (2005.77ms [SD=556.67] vs. 2165.75ms [SD=443.97]). Naming the left item first had no effect on children's naming times (left named first = 2058.70ms [SD=501.16] vs. right named first = 2157.11ms [SD=600]), as in Experiment one, but there was an interaction between cue and naming the left item first: When the cue was on the left, there was no difference in first item onsets when children named the left item first compared to when they named the right item first (1938.80ms [SD=642.39] vs. 2259.87ms [SD=1029.60]: *B*=-.05*, SE*=.04, *t*=-1.40 *, CI*=[-.10, .01]*,* $\chi^2(1)$=1.95, p=.16); However, when the cue was on the right, children were slower to begin naming when they named the left item first than the right item first

(2273.73ms [SD=472.20] vs. 2069.35ms [SD=633.29]: *B*=.08, *SE*=.02, *t*=2.93 , *CI*=[.03, .13], $\chi^2(1)$=8.35, *p*<.01).

Finally, our combined model indicated that, as in experiment one, adults were faster to begin naming the first item than children (1311.66ms [SD=336.16] vs. 2097.61ms [SD=424.69]). It also showed that, overall, participants were faster to begin naming after a left cue than a right cue (1627.52ms [SD=580.57] vs. 1728.95ms [SD=571.68]). Unlike in experiment one, there was no difference in participants naming times regardless of whether they named the left or right item first (1656.61ms [SD=562.46] vs. 1969.38ms [SD=630.38]). We also found a significant interaction between cue and age: Children were marginally faster to begin naming when the cue appeared on the left than the right (2005.77ms [SD=443.97] vs. 2165.75ms [SD=443.97]: *B*=-.05, *SE*=.02, *t*=-1.91 , *CI*=[-.09, .00], $\chi^2(1)$=3.61, *p*=.06) whereas the cue did not affect adults' naming times (left cue = 1287.09ms [SD=346.36] vs. right cue = 1335.82ms [SD=341.84]: *B*= -.02, *SE*=.13, *t*=-1.64 , *CI*=[-.04, .01], $\chi^2(1)$=2.65, *p*=.10).

2.3.3. Discussion

Experiment 2 replicated most of Experiment 1's findings. Children's starting points were predicted by where they fixated first, which in turn was influenced by our cue manipulation. Furthermore, children were faster to name the item they fixated on first, and this occurred even when their first fixations were manipulated by the cue. This again provides strong evidence that children continued to plan in a lexically incremental way, even for more complex utterances that required speakers to determine both semantic roles and grammatical relations.

Overall, participants' first fixations influenced their subsequent response, but children were more likely to name the left item first if they fixated on the left item first than the right, but adults were not (likely because adults showed an overwhelming bias to begin their utterance with the left item). This result differs from Experiment 1, whereby both children and adults' were more likely to name the left item first when they were fixated on the left item rather than the right item, and hints to children and adults relying on different factors when planning their utterances: children's starting points were more influenced by our attention manipulation, whereas adults' starting points are more affected by a perceptual left bias in Experiment 2 when they have to produce a more complex utterance.

2.4. General Discussion

Adults' starting points can be influenced by what they fixate on first, and this can indicate the planning strategy underlying word order choice. Here, we tested whether this was true of pre-schoolers by measuring 3-4-year-olds eye movements as they described objects on a screen. We further tested whether children could use a lexically incremental planning strategy (i.e., one where bottom-up factors affect starting points) by subliminally cueing one of the objects; and whether their starting points continued to be affected by bottom-up factors for utterances whereby they had to determine semantic roles (Experiment 2) or not (Experiment 1).

In both experiments, children's starting points were predicted by their first fixations: children were more likely to name the left item if they fixated on the left item first. Furthermore, our cue manipulation influenced children's first fixations as well as their starting points in both experiments: children were more likely to name the left item first when it was cued than when it was not, even when they had to determine semantic roles and grammatical relations as well as generate a more complex constituent structure. In addition, across both experiments, children were faster to begin naming the first item when they named the object they fixated on first, and this occurred even when they named the cued item first. We suggest that these findings provide preliminary evidence that children begin word retrieval as soon as they begin to fixate on an object.

Adults, on the other hand, showed an overwhelming bias to talk about the left item first, although we did find that their starting points were affected by both their first fixations and the cue (although this was harder to estimate in both experiments). Adults only showed a difference in naming times in Experiment 1 when they named the left item first.

2.4.1. Implications for Sentence Planning Processes in Children

Our results are one of the first to demonstrate that children's word order choice can be influenced by attention manipulations. These results add to the wider research investigating the link between children's attention and language development: not only do infants make use of attentional cues in language learning (Baldwin, 1995; Carpenter et al, 1998) but they also attend to objects in the order they have been described during comprehension. Our results show that children's visual attention corresponds with their own productions: Not only did children focus on each object as they described them, just like adults did, but they also used external attention cues to aid the planning of their utterance. Previous research has provided mixed evidence about whether pre-schoolers can use external influences to aid subsequent production with some studies suggesting they prefer to use implicit linguistic biases when deciding how to begin their utterance (Brough et al., 2018) and others suggest that three-year-olds will only use attentional manipulations when choosing a starting point in cases where the external factor (e.g., a cue) matches their linguistic bias (Ibbotson et al., 2013). Our results provide evidence to the contrary of this and show that three-to-four-year-olds can and do use attention-grabbing cues when planning their own utterances, at least in cases where they do not have any other linguistic biases, such as an animacy bias, to compete with. Moreover, children made use of these for both phrases used in our study, although it should be noted that the utterances participants produced in our study were constrained with minimal semantics.

Overall, our results suggest that pre-schoolers can use a lexically incremental planning strategy when planning a sentence. Not only were children in our study more likely to name the cued than uncued object first, but their first fixations also predicted their subsequent starting point in both experiments. This suggests that children began accessing lexical information as soon as they fixated on the object. Our analysis of children's onset times supports this idea further: children were faster to begin naming in both experiments when they named the cued object. This is in contrast to previous findings suggesting that pre-schoolers tend to choose their starting point after an initial apprehension phase whereby they extract the overall gist of a scene (Brough et al., in prep). Crucially, participants in our study produced sentences involving minimal semantic information, and as a result, they did not have need to extract an overall gist prior to choosing a starting point. As participants in our study produced the exact same sentence frame, it encouraged them to simply inspect the objects they had to name and begin lexical retrieval as quickly as possible and thus, removed the need to inspect the whole scene prior to speaking.

Previous studies indicate that pre-schoolers use a structurally incremental planning strategy (e.g., Brough et al., in prep), whereas our results suggest that pre-schoolers can use a lexically incremental planning strategy, at least for simple sentences with minimal semantic content. Our results alongside previous findings, suggest that children, like adults, are able to flexibly switch between planning strategies depending on the task demands (Van de Velde et al., 2014; Konopka & Meyer, 2014; Konopka, Meyer & Forest, 2019; Swets, Jacovina & Gerrig, 2013). In cases like our task, where children produced highly constrained sentences with minimal semantics, they could begin planning their starting point as soon as they fixated on an object. Furthermore, as children did not have an implicit bias for naming one objects first over the other, our attention manipulation brought one of the objects into focus, and aided lexical retrieval of that object. However, in cases where they have to produce a more complex utterances, such as describing an event, children's linguistic biases may make the agent or most animate characters more accessible, resulting in any attention manipulations having a weaker effect on children's starting points (e.g., as in Brough et al's study). Taken together, this suggests that children, like adults, can use a multitude of sources to aid utterance planning and will place differing weights on these sources depending on the task at hand.

Adults in our study also switched between planning strategies. They were more likely to name the object they first fixated on in Experiment 1, but we did not find this effect in Experiment 2, suggesting they switched their starting point strategies from one where they began lexical retrieval of the first fixated object in Experiment 1 to one that relied on their implicit left bias in Experiment 2. Moreover, adults' naming times were influenced by the cue when they named the left item first in Experiment 1, suggesting they used a lexically incremental planning strategy, whereas in Experiment 2, their naming times were not affected by any of our predictors, suggesting they were no longer beginning lexical retrieval as soon as they fixated on an object (note that the cue still influenced first fixations in Experiment 2). We suggest this null result in naming times in Experiment 2 is due to adults' overwhelming bias to begin their utterance with the left object first. These results are particularly interesting as the only difference between our experiments was the utterance produced: in Experiment 1, they produced a simple conjoined noun phrase and in Experiment 2 they produced a slightly more complex locative phrase – although this particular phrase should have still been very simple for adults to produce. It is possible that for phrases that require speakers to generate relations between objects or characters, such as transitives or even locative phrases, speakers will prefer to use biases when planning their utterance. Biases are less likely to change, whereas attention cues will. As a result, a speakers' attention may not be as reliable a cue for planning more complex utterances. Alternatively, the connotations of the "next to" phrase encourages a left-right relationship, resulting in a stronger left bias effect in Experiment 2. It is interesting that we do not see this difference in children, especially considering the locative phrase produced in Experiment 2 should have been more difficult for them to produce. However, as we will discuss in the next section, children in our study did not seem to have the same left bias as adults. Thus, there was no bias for them to use when producing the more complex utterance, meaning they relied more on the attention manipulation when choosing their starting point. Future studies should investigate this idea further, for example by contrasting the effect a cueing manipulation has on children's starting points for simple utterances when there are other biases competing with it (e.g., animacy bias).

2.4.2. Implications for Left-Bias Models

Secondly, our results have a further implication in that they can tell us about the nature of the non-linguistic left-bias. English-speaking adults tend to view pictures left-to-right, but also process events in a left-to-right manner (e.g., Christman & Pinger, 1997). Different accounts have been proposed to explain the origin of the left-to-right bias: The neuropsychological account assumes this bias derives from functional properties of left hemisphere processing (Chaterjee, Southwood & Basilico, 1999; Chaterjee, 2001), whereas the cultural account assumes the bias is a manifestation of a participant's scanning habit, which is affected by the direction their native language is written (e.g., Chokron & De Agostini, 2000; Nachson, Argaman, & Luria, 1999; Spalek & Hammad, 2005; Tversky, Kugelmass, & Winter, 1991). Most work investigating the left bias has asked participants who use different writing systems (e.g., contrasting languages who use a left-to-right direction, such as English, with those using a right-to-left direction, such as Arabic) to draw pictures of events and measures whether they put the subject on the left or right of the drawing. In tasks such as this, adults tend to draw the subject in the position that corresponds to their writing system (on the left in English, but on the right in Arabic), whereas children draw the subject in both the left and right position, regardless of their native language (Dobel, Diesendruck & Bölte, 2007). This suggests the left-to-right bias can be explained by a cultural account. However, in tasks asking adults to match pictures to sentences have found that even adults whose writing system uses a right-to-left direction show a facilitation effect when the subject of an action appears in the left position (Maass & Russo, 2003), suggesting that the left bias may be influenced by a neuropsychological predisposition to prefer attending to the left of a scene first, but there is also a cultural preference to put the subject in a position canonical to your writing system.

Children in our study did not show a left bias in their starting points when they had to produce a simple conjoined noun phrase nor when they had to produce a more complex phrase. Adults, on the other hand, showed an overwhelming preference to begin their utterance with the left object, and to a greater extent in Experiment 2. These findings on their own suggest that the spatial bias seen in adults derives from exposure to a speaker's written language rather than an innate neuropsychological preference to view and talk about scenes left-to-right as the children in our study had not yet been through intensive exposure to reading and writing. This is consistent with previous findings showing that 3- to 6-year-old children do not draw events in a left-to-right manner, and instead randomly assign the agent and patient to either the left or right position (Dobel et al., 2007). Our results add to these findings as we failed to find any evidence of a left bias in children's language production. This is particularly striking as if the left bias originates from exposure to a writing system, you might expect the left bias to appear in children's production first and broadens out to other modalities, such as those involved in determining spatial relations. Overall, these results suggest that children are born with the flexibility to represent spatial relations and any bias to one side derives from exposure to a writing system (Göbel. McCrink, Fischer & Shaki, 2018).

Whilst we failed to find evidence of the left bias in our pre-schoolers productions, previous research suggests there is a directionality bias in children the corresponds to their writing system (Chokron & De Agostoni, 1995; Tversky et al., 1991). For example, in a line bisection task, Hebrew-speaking pre-schoolers showed a rightward bias in determining the centre of a line (i.e., a bias corresponding to their writing system), whereas French-speaking pre-schoolers did not (Chokron & De Agostoni, 1995). It is important to remember that whilst pre-schoolers in our study have not yet gone through formal instruction of their writing system, they are still typically exposed to it during reading story books with their caregiver. It is possible that children have some form of a left bias, but this fails to manifest in their own productions. In line with this hypothesis, we found that children were faster to begin speaking when they fixated on the left object in both experiments and after the left object was cued in Experiment 2. Moreover, children seemed to show a left bias in their gaze behaviour: whilst the cue did influence their first fixations, children showed a slight preference to fixate on the left object first more than the right object in both experiments (Exp 1: 56% vs. 44%, Exp 2: 57% vs. 43%; see also Figure 3). These findings may provide partial support for the hypothesis that there is an implicit bias to direct attention from the left to the right when generating relations between objects and characters (e.g., Chaterjee et al., 1999). Thus, when children were asked to express grammatical relations between the two objects, naming times were facilitated when the cue matched their innate preference to direct their attention to the left. This is consistent with previous research suggesting that the left bias in spatial attention is a function of hemispheric specialization which can be overruled by other factors, such as cultural pressures (Maass & Russo, 2003). However, if our naming times finding are a manifestation of the left bias, it would suggest that the implicit left bias is not strong enough to affect production, as it failed to have as big an effect on children's subsequent word order choices compared to adults.

Thus far, we have argued that children in our study used a lexical incrementality strategy, as their starting points were affected by our cue manipulation. However, we have also suggested that pre-schoolers demonstrated some evidence for a left bias which may have affected their gaze and their naming times. According to a lexical incrementality account, pre-schoolers should begin their utterance talking about whatever they attend to first (and so we should see a slight left bias in their productions as well). Yet, children in our study named the left object about half of the time (Exp 1: 53 %, Exp 2: 50%). It is possible that for some children, the left object was too difficult a name for them to retrieve (and therefore use as their starting point), resulting in them naming the right object first instead. Indeed, there were some objects that children found difficult to name (e.g., shoulder) despite being matched with an object with a similar AOA. It is possible that pre-schoolers do have a left bias that affects production, but we just failed to find evidence of it. However, it should be noted that pre-schoolers in our study only showed a very slight tendency to fixate on the left object first, and so this potential left bias is nowhere near as strong as adults' (who first fixated on the left object to a much greater extent than the right object: Exp 1: 94% vs. 6%, Exp 2: 93% vs 7%).

To conclude, we have shown that three-to-four-year-olds starting points are affected by visual attention for both very simple and more complex utterances. Furthermore, they are susceptible to low-level bottom-up factors for simple stimuli (i.e., two objects next to each other). Overall, these results suggest that children can plan in a lexically incremental way, although we cannot generalise this to production of events. Furthermore, our results provide evidence that the left-bias seen in adults is from exposure to the direction their native language is written.

3. Study 2. Mechanisms Underlying Pre-Schoolers Morpho-Phonological Encoding: Can Pre-schoolers retrieve word forms incrementally?[9]

3.1. Introduction

Despite considerable progress in explaining how children learn language, we know much less about how they develop the ability to formulate their utterances. Children can understand much more than they can produce (e.g., Golinkoff, Ma, Song & Hirsh-Pasek, 2013), and this production gap likely exists because speaking involves the coordination and overlap of a series of complex cognitive processes. To produce a multi-word utterance (e.g., *the dog is next to the cats*), children must learn to coordinate the speech planning processes of each successive word (Levelt & Meyer, 2000). How do they do this?

In order to produce a multi-word utterance, speakers must engage in several levels of speech planning: They first must create a prelinguistic representation of what they are going to say, select the lexical items associated with that prelinguistic representation, and then generate the corresponding morphological and phonological plans in order to develop a phonetic plan that can be subsequently articulated (e.g., Dell, 1986; Garrett, 1975; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999; Levelt & Meyer, 2000). For instance, to name an object, a speaker must first identify the object (via various visual-conceptual processes), select the correct lexical unit (or *lemma*), retrieve the morpho-phonological representation

[9] Note that the method and sample used in this study are identical to that presented in chapter 2.

associated with the selected lemma (i.e., the word form), and generate a phonetic representation ready for articulation (e.g., Levelt & Meyer, 2000).

Alongside this, when producing a sentence, speakers must also plan the order of words that they will speak. How is this achieved? The *structural incrementality account* assumes that speakers use top-down information to generate an utterance plan (e.g., Griffin & Bock, 2000). Here, word order is determined by a higher-order conceptualisation of an event's semantic relations (i.e., determining who is doing what to whom; Griffin & Bock, 2000). This kind of strategy is consistent with language production models assuming that syntactic structure is generated based on the thematic structure of a scene without reference to lexical representations (e.g., Chang, Dell & Bock, 2006). A 'gist extraction' mechanism quickly assigns thematic roles to each character, such as the agent and patient, and creates an overall structure for the sentence plan, which is imposed before speaking. Thus, word order can be affected by linguistic biases, such as a bias to have animate objects earlier in an utterance than inanimate objects. This means that speakers do not begin lexical retrieval until *after* a sentence plan has been generated (proposed to be after the first 400ms; e.g., Griffin & Bock, 2000).

Alternatively, speakers can also use a bottom-up *lexically incremental* strategy. Under this strategy, speakers begin talking about the most salient or accessible character first and build a sentence plan that is consistent with this starting point (e.g., Gleitman, January, Nappa & Truesewell, 2007; Tomlin, 1997). This kind of strategy is consistent with language production models assuming that visual information immediately activates lexical information prior to the generation of a syntactic structure (Bock, 1982; Bock & Levelt, 1994). This strategy may be useful under conditions in cases where an event is hard to interpret (van de Velde et al, 2014), or after a character has been lexically primed (Konopka

& Meyer, 2015). Recent work suggests that pre-schoolers can make use of both these planning strategies to guide word order (e.g., Brough, Branigan, Gambi & Rabagliati, 2018, Lindsay, Rabagliati & Branigan, 2018), but how exactly do they retrieve the word forms for their sentence?

A key finding in the adult psycholinguistic literature is that adult's language production is incremental (e.g., Griffin & Bock, 2000). That is, they plan small units of their utterance, begin speaking, and then plan later parts of their utterance whilst simultaneously producing earlier parts. Indeed, eye-tracking studies have shown there is a tight link between a speaker's eye movements and their speech output (Gleitman et al., 2007; Griffin & Bock, 2000; Griffin, 2001; Meyer, Sleiderink & Levelt, 1998; van der Meulen, Meyer & Levelt, 2001; see Meyer, 2004 for review). For example, when describing two objects as "*a cat is next to a dog,*" speakers typically fixate on each object just before mentioning them. That is, they will fixate on the picture of a cat first and then switch their gaze to the next object (in this case, a dog) a few hundred milliseconds before they begin naming, suggesting that speakers' eye movements are tightly coordinated with speech planning processes.

Indeed, eye-tracking studies have shown that the time speakers spend inspecting an object depends not only on how long it takes for them to identify the object (i.e., visual-conceptual processing; Griffin & Oppenheimer, 2006), but also on the time it takes for them to select the appropriate name (or lemma) and retrieve and generate the phonological word form (Belke & Meyer, 2007; Griffin, 2001; Meyer & van der Meulen, 2001; see also Meyer, Belke, Telling & Humphreys, 2007). For example, Meyer et al (1998) asked participants to name object pairs (e.g., *a scooter and a hat*) but manipulated participants' ability to not only recognise the objects (by presenting them with intact or degraded line drawings), but also their ability to retrieve the phonological form of the object name by varying the frequency of object names (see Jescheniak & Levelt, 1994). They found that both manipulations affected the time speakers gazed at an object: gaze durations were longer for degraded objects than intact objects, and they were also longer for low frequency names than high frequency names. Similarly, speakers tend to look longer at objects with long names than at objects with short names (Meyer, Roelofs & Levelt, 2003; but see Griffin, 2003).

Overall, these findings suggest that speakers' gaze remains on an object until its name has been planned to the level of phonological form and they will only shift their gaze to the next object *after* that phonological form has been retrieved (see also Roelofs, 2007, 2008). But why do speakers continue fixating on an object until they have retrieved the phonological form of an object name? It is possible that long gaze durations facilitates lexical access processes: By attending to (and therefore fixating on) an object, both object identification and lexical retrieval processes are facilitated (for example, through a process of spreading activation; e.g., Roelofs, 1992; Wühr & Waszak, 2003; Wühr & Frings, 2008). Furthermore, attending to an object may also reduce the interference from the names of other objects, especially in cases of multiple object naming (e.g., Meyer et al., 1998; see Griffin, 2004 for review). Thus, when naming multiple objects, the conceptual and lexical representations of the fixated-on-object will receive greater activation (thus inhibiting alternative names), and thus will facilitate subsequent lemma selection and word form retrieval. But by fixating on the to-be-named object, a speaker will primarily attend to lexical processes for *that* object, and therefore limits the amount of interference of other object names. This proposal fits well with findings demonstrating that lexical access requires processing capacity (e.g., Cook & Meyer, 2008; Ferreira & Pashler, 2002; Roelofs, 2008). Thus, when planning to name two objects, speakers will direct their visual attention to the first and then second object and late shifts in gaze to the second object (i.e., just before naming of the first object), allows for objects to appear in the correct word order.

Chapter 3

The studies reviewed so far have mainly focused on what eye movements reflect during speech planning. They suggest that visual attention and speech planning processes are tightly coordinated. Furthermore, they suggest that speakers retrieve the phonological form of each object name in a highly sequential (and incremental) fashion (Griffin, 2001; Griffin, 2003). But when speakers name multiple objects, processes of lexical access must be running partially in parallel (i.e., retrieving one name while speaking another) so that strings of words can be produced fluently (Morgan & Meyer, 2005; Meyer, Ouellet & Hacker, 2008; Malpass & Meyer, 2010). For example, when producing "*a cat and a dog,*" speakers must access two different content words *cat* and *dog*. Thus, later stages of lexical access (e.g., phonetic encoding) would be occurring for the first word (*cat*) while earlier stages of lexical access (e.g., conceptual processes) will be occurring for the second word (*dog*). Indeed, adult speaker's shift their gaze from the first to the second object about 150ms before they begin naming the first object, suggesting that they are still planning some aspect of the first word as they begin naming the second (e.g., Griffin & Bock, 2000; van der Meulen, Hartl & Tappe, 2003). Such parallel planning serves fluency as if a speaker encodes each next word from scratch after completing the current word's articulation, then speech would appear disconnected and disfluent. Indeed, in typical multiple-object naming studies, adult speakers can still identify and process objects extrafoveally (meaning that they don't necessarily have to directly focus their attention on an object to process it; Morgan & Meyer, 2005; Meyer, Ouellet & Hacker, 2008), suggesting there is some temporal overlap in the processing of successive objects names. However, the ability to retrieve names of the fixated on and extrafoveal object in parallel is dependent on the difficulty of processing the fixated-on object (Malpass & Meyer, 2010), which implies that processing load may be relevant to parallel planning.

Adult studies suggest that speakers can flexibly engage in parallel planning depending on the length of time between the first and second word. For example, Griffin (2003) asked speakers to name object pairs where the first object name was either short (e.g., *wig carrot*) or long (e.g., *windmill carrot*). When the first word was short, speakers took longer to begin speaking but analyses of their eye movements showed that they also gazed at the second-named-object for longer before they began naming but spent less time gazing at the second object after they began speaking, suggest that speakers began planning of the second-named-object whilst still planning the first-named-object. However, when speakers produced a 'next to' phrase between objects (e.g., *the windmill is next to the carrot*), this effect disappeared, suggesting that speakers postponed planning of the second object until after speech began. These findings indicate that speakers only began preparing as much of the second-named-object as they needed to prior to speaking, thus minimizing buffering of the first-named-object.

In this paper, we are concerned with how children plan their utterances. Do children's eye movements reflect speech planning processes? And specifically, do they, like adults, fixate on objects in a sequential fashion? If they can, this would indicate that they retrieve word forms in an incremental fashion. In addition, we ask when do children shift their gaze to the second named object before they begin naming. If they begin fixating on the second named object *after* they begin naming, this would indicate that children begin planning of the second object in a serial way (i.e., they only begin lexical retrieval of the second object once they have completed planning processes of the first object). However, if they fixate on the second object *before* they begin naming, it could indicate that they begin lexical access processes of the second object whilst also planning later aspects of the first object (i.e., parallel planning). In what follows, we review the evidence about how children plan their utterances.

Chapter 3

Studies investigating the mechanisms underlying sentence planning in children are limited and, where available, have typically relied upon speech error data. Findings from such studies suggests that the overall nature and organisation of the language production system in children is similar to that of adults (e.g., Jaeger, 1992; McDaniel, Mckee & Garrett, 2010; see McKee, McDaniel & Garrett, 2018 for review). Based on this, we might expect children to formulate their sentences in the same way as adults. Indeed, we can draw some inferences about whether children incrementally retrieve word forms by looking at the errors they produce. Children and adults differ in their proportions of perseverations (e.g., Daddy, me watching *Daddy* cooking ... no ... *Mommy's* cooking) and anticipatory errors (e.g., Yeah, *it* likes it ... *I* like it): Adults produce significantly more anticipatory errors than children do, who instead produce substantially more perseverations (Jaeger, 2005; Wijnen, 1990). For a speaker to produce an anticipatory error, they must have already encoded later elements of the upcoming speech which then trigger the error. The fact that children do produce anticipatory errors implies that they are incrementally retrieving words whilst speaking, though the fact that they do so to a lesser extent than adults implies that their planning window may be smaller. Evidence of parallel planning comes from word exchange errors (e.g., *her run is nosing*): children, like adults, produce word exchange errors, suggesting that lexical retrieval and phrasal planning occur concurrently for both age groups (Jaeger, 2005). Taken together, these findings suggest that not only are children able to plan incrementally, but the planning process is parallelized, just like in adults.

Further evidence for incremental planning in children comes from studies that elicit speech errors. For example, McDaniel, Mckee and Garrett (2010) asked younger (three- to five-year-olds) and older children (six- to eight-year-olds) and adult controls to complete an elicited production task which encouraged the production of different relative clauses and measured the kinds of errors participants produced. They found that children produced more filled pauses and restarts than adults did. Furthermore, children typically produced a relativizer (e.g., *who*) or a complementizer (e.g., *that*) more frequently than adults, and also tended to pause after producing one, suggesting that children were using these components of their utterance as a stalling device so they could begin planning the next part of their utterance. Based on these findings, McDaniel et al (2010) proposed that adults are able to plan further ahead than children (e.g., they have planned a whole clause prior to speaking), whereas children need to plan more frequently and more locally than adults, resulting in more pauses. That is, children need to stop to plan each clause as it comes, resulting in more disfluencies and pauses, whereas adults can continue planning the next clause as they are articulating the current one, resulting in fluent speech (see also McKee, McDaniel, Garrett, Lozoraitis & Mutterperl, 2013). Findings such as these could suggest that children are able to plan their utterances in small chunks (i.e., incrementally), but struggle with parallel planning.

So far, the evidence suggests that children can plan incrementally, although their planning scope seems to be much smaller than adults'. However, when it comes to parallel planning, the evidence is mixed: error data implies that children are retrieving lexical items in parallel whereas the disfluency data suggests that children struggle to begin planning the next clause of their utterance whilst they are still speaking. Whilst speech error and disfluency data can give us some insight into the developing language production system, they fail to tell us about the moment-by-moment mechanisms involved in children's language production. Eye-tracking methods in adult production studies have revealed how the production system operates during *successful* language use (e.g., Griffin & Bock, 2000; Meyer, Sleiderink & Levelt, 1998) and whilst many studies have used eye-tracking paradigms to investigate how children comprehend language online (e.g., Fernald, Pinto, Swingley, Weinberg & McRoberts, 1998; Snedeker & Trueswell, 2004), little work has used these paradigms to investigate how children plan and structure their own utterances (but see Bunger, Trueswell

& Papfragou, 2012; Norbury, 2014; Rabagliati & Robertson, 2017). Indeed, work in older children has found a link between children's language production abilities and their visual attention (Jongman, Roelofs, Scheper & Meyer, 2017): children who were faster at naming pictures tended to have greater sustained visual attention than children who were slower at picture-naming. These findings could suggest that children visually attend to objects to aid retrieval for forthcoming words like adults do. Furthermore, a recent case study investigating typically and atypically developing children's eye movements during picture naming found that their typically developing participant (age 10.5 years) tended to look to the first named item (i.e., the agent of an event) before speaking, and then looked to the second named item (i.e., the patient of an event) after they had named the subject (Norbury, 2014), suggesting that the child's eye movements were tightly synchronized with their speech planning processes, indicating that they were planning incrementally. However, this data only looked at one typical child's eye movements (to compare against atypically developing children's eye movements) who was much older than the population we are interested in. In fact, to the authors' knowledge, no study to date has investigated how pre-schoolers inspect multiple objects as they name them. Indeed, studies that have measured pre-schoolers' eye movements during production have typically looked at their eye movements as they inspect a scene before speaking (Bunger et al., 2012) or to measure where they look during referential communication tasks (where they name one object; Davies & Kreysa, 2018; Rabagliati & Robertson, 2017).

3.1.1. The Current Study

In this study, we examine how pre-schoolers formulate their utterances and ask whether they do this in the same way as adults. Specifically, we examine the relationship between pre-schoolers' eye movements and their speech, and ask if consider what their eye movements can tell us about how they are retrieving word forms (i.e., do they retrieve word forms in an incremental fashion or not)? To do this, we tracked participants' eye movements as they described stimuli involving two objects. In Experiment One, participants named the objects using simple noun phrase conjunctions (e.g., *a car and a doll*). In Experiment Two, participants expressed the relative location of the objects using complete sentences (e.g., *a car is next to a doll*), thus requiring participants to determine semantic roles and grammatical relations, whilst also generating a more complex constituent structure. We also tested adult controls to compare children's eye movements against.

Our main goal was to examine children's eye movements as they named objects, and consider whether their eye movements can provide insight into their lexical retrieval processes. We considered this in a number of ways. We first measured whether pre-schoolers were more likely to look at the first named object than the second named object *before* speaking, and whether they were more likely to look at the second named object than the first named object *after* they began speaking. If children retrieve word forms in an incremental fashion, we expect them to look at the first named object more than the second named object before speaking, but they should then look to the second named object after they began speaking.

In addition, we also determined the time children began looking to the second-named-object after they had finished fixating on the first-named-object (defined as the offset of the final fixation to the first object before they began naming the first object. Note that the onset of the fixation may have occurred before they began naming, but the offset may have occurred *after* they began naming). If pre-schoolers begin fixating on the second-named-object after they begin speaking, it would imply that they are planning in a more serial way and are engaging in less parallel planning (because they are only beginning retrieval processes of the second object *after* they have planned the first object all the way to articulation).[10] However, if they begin fixating on the second object before they begin naming the first, it could suggest that they begin initial planning processes of the second word (e.g., visuo-conceptual processes, morpho-phonological encoding) whilst they are still completing later processes (e.g., phonetic encoding) of the first word. That is, they are engaging in parallel planning (note that we are aware that simply fixating on the second named object does not necessarily mean they are engaging in any linguistic planning of that word, but these are initial findings for future work).

To further investigate the language production process in pre-schoolers, we also measured gaze durations (the time participants fixated on each object before finishing naming them), eye-voice spans (the time between participants first fixating an item and naming it) and gaze-to-speech lags (the time between the offset of gaze to the named object and the onset of its spoken name; Meyer, 2004). Gaze durations reflect the time it takes for speakers to retrieve and generate the phonological word. Therefore, longer gaze durations would indicate that participants took longer to retrieve a word. Of interest to us is whether there are differences in gaze durations between the first and second named object: if gaze durations are longer for the second named object than the first named object, it would indicate that it took longer to plan a word whilst participants were completing another task (i.e., speaking). However, if a speaker has a shorter gaze duration for the second named word compared to the first named word, it would indicate that speakers may have been processing the second named picture extrafoveally, and therefore were planning the second word in parallel with the

[10] Note that this is different from our first analysis described above. In this analysis, we are looking at the time participants begin fixating on the second object whereas the first analysis is looking at the average proportion of looks to each object (i.e., do they look at the first object more *before* speaking and do they look at the second object more *after* speaking).

first. Eye-voice spans can tell us how long it took for speakers to begin planning the to-be-named word to then naming it. As with gaze durations, a longer eye-voice span for the second named picture compared to the first named picture would indicate that planning for the second named picture took longer to plan. Finally, gaze-to-speech lags can tell us about how speakers coordinate eye gaze and speech. If speakers' gaze reflects planning to the phonological code, we would expect them to have a longer gaze-to-speech lags (which would mean that they stop fixating on the to-be-named object before they begin speaking). However, it is also possible that speakers will continue fixating on to the to-be-named picture until shortly before naming – or even after – they begin naming. If speakers continue fixating on the first named picture, it could indicate failures to begin planning of the second named picture until after they began speaking (perhaps to remove any interference that could occur when planning of the next-named-object), and would suggest that they were planning in a more serial fashion.

3.2. Experiment 1b

3.2.1 Method

Participants: Participants were 30 adult native English speakers (M=23.13 years, range = 19-32, 17 female) & 30 native English-speaking children aged between 3 and 4 (M=46.33 months, range = 36-58, 13 female).[11] We recruited most children from private nursery schools around Edinburgh, and the rest from a database of interested families. Ethnicity and SES were not recorded, but were representative of the area (almost entirely white, predominantly from middle-class Scottish families).

[11] Note that this is the same sample as used in the study reported in Chapter 2; Experiment 1a.

Materials On each trial, participants saw two pictures side by side (Fig.1).[12] We created two lists which were essentially mirror-reversed versions of each pair to counterbalance the effect of item name difficulty (i.e. in list one, the car appeared on the left of the screen and the doll on the right, and vice versa for list two).

To create our picture pairs, we chose 60 words with an average age of acquisition (AOA) of less than 3 years (Johnston, Dent, Humphreys & Barry, 2010; mean AOA = 2.31) and paired each word with another that had a similar AOA, creating 30 pairs of words (e.g., car and doll had the same AOA of 1.83 and were paired together). Our stimuli were these 30 pairs of words and their corresponding pictures (see Figure 1). Each pair was only used once throughout the experiment (total of 30 trials). We then chose child-friendly pictures of each word. Each picture ranged in size from 155 x 480 pixels (*e.g. carrot*) to 480 x 480 pixels (*e.g. clock*), depending on what the item was (narrow words like carrot had a smaller width than other words – although most pictures were 480 x 480 pixels). These pictures were then spliced together, with one item appearing on the left and the other on the right.

Procedure Prior to the study, participants were told that they would see pictures on the screen, and were instructed to name each picture as a conjoined noun phrase as quickly as possible e.g., *a car and a doll*). The study began with 4 practice trials. If participants didn't know what to say or they made a mistake, the experimenter would produce a correct version

[12] We included a cue manipulation whereby one of the pictures had been preceded by a subliminal cue (a red star 445 x 445 pixels) that flashed on either the left or the right of the screen for 75ms before picture onset. The cue manipulation formed part of a further study (reported in Lindsay, Rabagliati & Branigan, in prep). We created two lists which were essentially mirror-reversed versions of each pair to counterbalance the effect of cue for each item (i.e. in list one, the car appeared on the left of the screen and the doll on the right, and vice versa for list two). Thus, the cue appeared before each item an equal number of times. **We do not include cue in any of our analyses in this paper, but because children did not name their pictures left-to-right, we coded which object was their first named object and which was their second named object.**

of the utterance. Participants then completed 30 trials, with a short break after 15 trials. During this time, the experimenter would not name the pictures for them but, in cases where children became distracted, they would prompt them back to looking at the screen by saying *What do you see here? Can you tell me what these pictures are?*

Each trial began with a central fixation point, and then a cue (a red star) was presented on either the left or right side of the screen for 75ms. Participants then saw two pictures on the left and right side of the screen, and named them as instructed. The experimenter ended each trial once the participant had named the picture. If participants did not know the name for the items, the experimenter moved on to the next trial. Order of trials was individually randomized.

Stimuli were presented using SMI's Experiment Centre on a laptop fitted with a REDn Scientific eye-tracker (SensoMotoric Instruments GmbH, www.smivision.com). The tracker was calibrated once at the start of the session and once after 15 trials using a 2-point grid. It recorded fixations binocularly at 30Hz, but we only analysed right-eye fixations. Sessions lasted approximately 10-15 minutes.

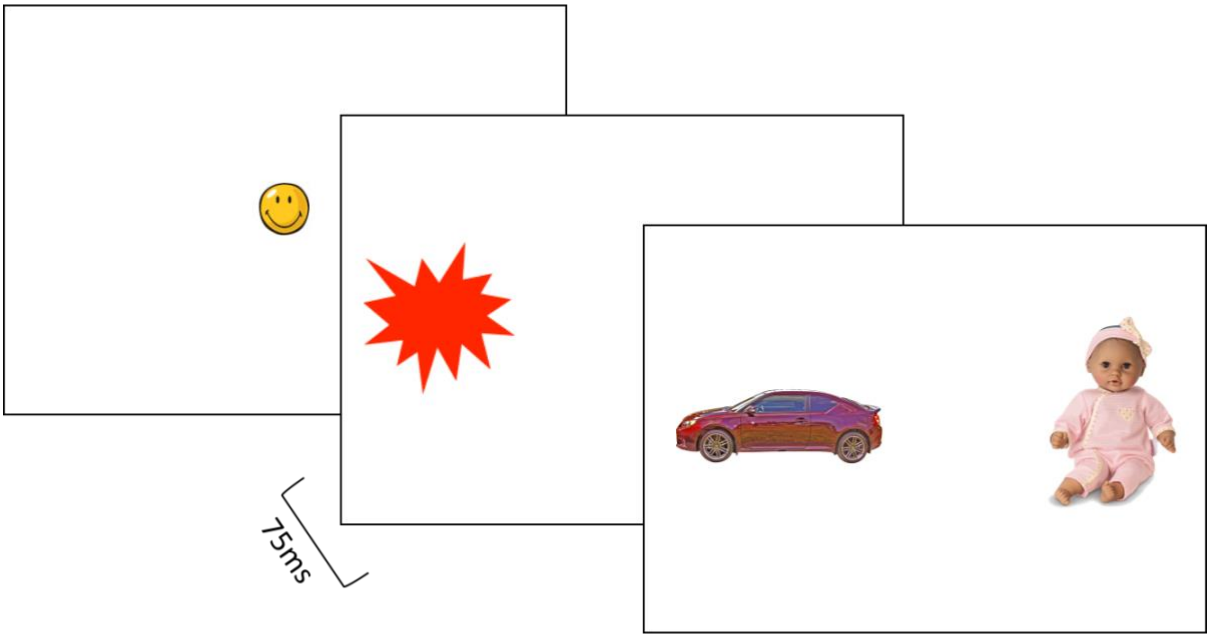

Figure 8. Example trial for experiments 1 and 2. The cue appeared on either the left or right of the screen an equal number of times throughout the experiment. Participants were asked to describe the pictures as quickly as they could. In Experiment 1, participants described the images as a conjoined noun phrase (a car and a doll); in Experiment 2, they produced a locative phrase (the car is next to the doll). A subliminal cue appeared before each trial (the effects of which are reported in a further study) but we do not discuss this in this study.

Coding. Participants' utterances were transcribed, and scored as to whether they named the left or right item first or second. Trials were discarded if participants did not produce a response (Adults: 0.003%, Children: 5%).

We also coded the time during the trial at which participants initiated their utterance (defined as when the participant began speaking about the pictures[13], including hesitations

[13] If participants asked a question about the pictures or told a story about one of objects, this did not count as speech onset and that trial would be discarded. They had to name the objects (e.g., *a car and a doll*) in order for us to include it in our analyses.

such *uhm*), the onset of the first determiner, onset of the first named object, offset of the first name, onset of second determiner, onset of second named object and offset of second name. We removed observations where participants took more than 5000ms to begin speaking or were 1.5 standard deviations below the age appropriate mean (<1% of data points for each age group).

We analysed participants' eye movements for the time period in which the images were presented on screen. Using BeGaze software (Version 3.6), we established two areas of interest that covered the entire left- and right-hand sides of the screen, respectively. We coded whether fixations were directed to one of those AOIs, offscreen, or could not be measured due to track loss. We removed observations on which no gaze was recorded (due to looking away or where tracking ratio was < 60%: Adults: 1.0%; Children: 20.7%).

In this paper, we are concerned with where participants looked over time (specifically, the time they looked to the first named object before naming it, and then shifted their gaze to the second named object). We first coded where participants were looking (first or second named object) for the first 2000ms before speaking and 2000ms after speaking within 100ms time bins, where time 0 was fixed as the time each participant began naming the first picture. We therefore created two variables: one which focused on looks to the first named picture and another that focused on looks to the second named picture. If participants fixated to the first named picture, it was coded as 1 for our fixations to first named picture variable and 0 for all other fixations (and vice versa for our second named picture variable: fixations to the second named picture were coded as 1 and all other fixations were coded as 0). This meant that some participants could have looked at the same picture within the same time bin (i.e., fixations to first named variable and fixations to second named variable are both 1 in the same 100ms time bin).

Data Analysis. We analysed participants' utterances and eye movements over the trial. We used linear mixed-effects models to analyse our data. Specific details of each model are given in the relevant section, using lme4 syntax. All models included participants and items as random effects (Baayen, Davidson & Bates, 2008), and incorporated maximal random effects structure (Barr, Levy, Scheepers & Tily, 2013), except when otherwise specified. In cases where the model would not converge, correlations among random effects were fixed to zero (Bates, Kliegl, Vasishth, & Baayen, 2015), and in cases of singularity, we used a partially Bayesian approach using the *blme* package – an extension for *lme4* (Chung, Rabe-Hesketh, Dorie, Gelman & Liu, 2013). Unless otherwise specified, all predictors were contrast coded (-1, 1). For the LMM models, we report coefficient estimates (*b*), standard errors (*SE*), *t* and *p* values for each predictor; 95% confidence intervals (CI) are from the *confint* function (method=Wald). For all models, *p* values are from log- likelihood ratio tests.

3.2.2. Results

In this paper, our focus is on children's eye gaze, and whether they show the same pattern as adults (for work looking at how pre-schoolers choose their starting point, see Lindsay et al., in prep). To do this, we analysed the following: Firstly, we focused on whether pre-schoolers were more likely to shift their gaze to the second named picture after they begin naming, which would indicate that children are retrieving each word form in an incremental fashion (gaze paths section). We also analysed participants' gaze durations and eye-voice spans, which can tell us how long it took for participants to retrieve a word to its phonological word form (longer times here would indicate retrieval difficulties), and their gaze-to-speech lags, which can tell us about when participants stopped fixating on a named object in relation to speech (i.e., it can tell us about how participants coordinated eye gaze and speech – a longer gaze-to-speech lag could suggest that participants could begin planning of upcoming words whilst completing later production processes, such as phonetic encoding). Finally, we look at the time pre-schoolers shift their gaze to the second named picture: if they begin looking to the second named picture after they begin naming, it would indicate that they are planning in a serial fashion (that is, they are retrieving word forms one at a time) whereas if they begin looking at the second named picture before they begin speaking, it could indicate parallel planning.

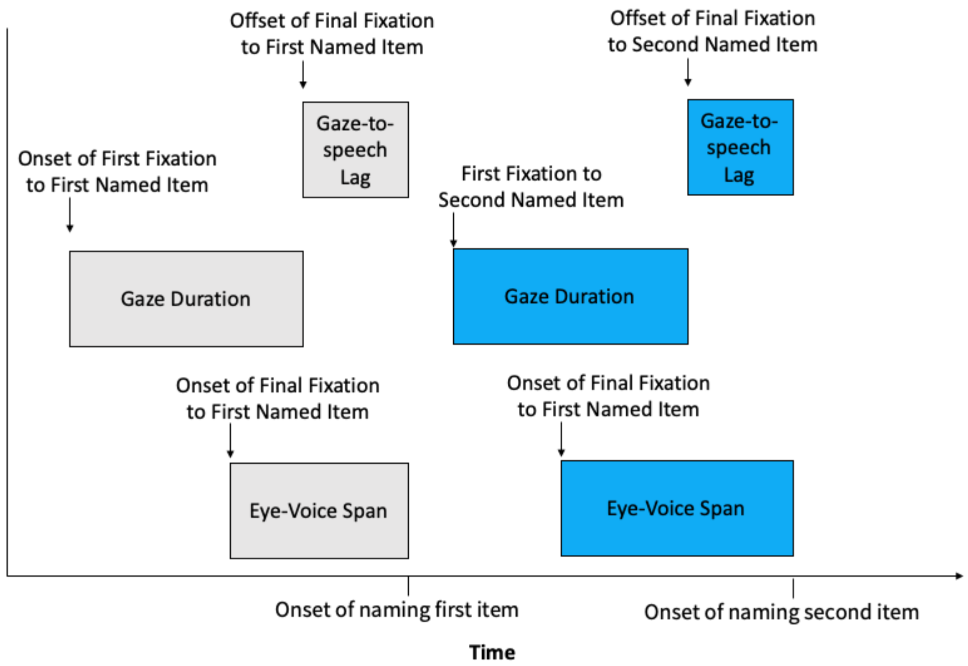

Figure 9: Schematic timeline of each eye-tracking measurement over a trial. The gaze duration is the total time a participant fixates on an object before the participant finishes naming it (note that this can include fixations to the item after the participant begins speaking about that item); the eye-voice span is the time between the onset of the final fixation to the named item and the onset of naming that item – both these measures can tell us about any difficulties the participant had retrieving the word form before ultimately producing it; and the gaze-to-speech lag is the time between the offset of the final fixation to a named item (i.e., when they stop fixating on the object) and naming it, which can tell us about how the participant coordinated eye-gaze and speech and perhaps indicate when they began planning of words to be produced later in the utterance.

Gaze Paths. Here, our focus is on how participants' eyes move over time (before and after they begin naming). Figure 9 shows the mean proportion of gaze to the first named and second named picture for each age group 2000ms before and 2000ms after they began naming the first item. The figure shows that both adults and children preferentially fixated the objects in the order they mentioned them, suggesting their utterance planning was incremental; they looked to the first named picture before they began naming it, and then shifted their gaze to the second named picture just before they named the first item.

To analyse whether participants were fixated on the first named object more than the second named object before speaking, and whether they fixated on the second named object more than the first named object after they began speaking, we calculated the average proportion of fixations to each area of interest (first named picture vs. second named picture) in two different time windows: one 500ms before they began naming the first picture and one 500ms after naming the first picture(shaded areas in Figure 9). To do this, we used our time bin dataset (i.e., the one in which we coded where participants were looking within 100ms time bins) and calculated the average proportion of fixations to each picture 500ms *before* and 500ms *after* they began naming the first item (so if a participant looked at the first named picture in all 5 100ms time bins before naming, their average score for fixations to the first named picture would be 1; and if the same participant also fixated on the second named picture in 3 out of 5 100ms time bins, their average fixations for the second named picture would be 0.6 before naming; we did the exact same thing for our 500ms time window after speaking). We therefore had an average fixation score to each AOI for each time window for each trial for each participant. We tested whether participants had a greater fixation score as a function of AOI in our pre-schoolers' analysis (fixation score ~ AOI) and in our combined model, we also looked at the effect of age (fixation score ~ AOI * Age Group). If participants continue looking at the second picture whilst they are naming the first, they should have a greater fixation score for the first named picture than the second named picture before they began naming the first item (i.e., in the first time window), but should have a greater fixation score for the second named picture than the first named picture after they began speaking (i.e., in the second time window). However, it is possible that participants will continue looking to the first named picture whilst they begin naming it (so therefore, they will have a greater fixation score for the first named object after they begin naming it), which would suggest that they need to continue fixating on the first named picture to aid later lexical retrieval processes (such as articulation or monitoring processes).

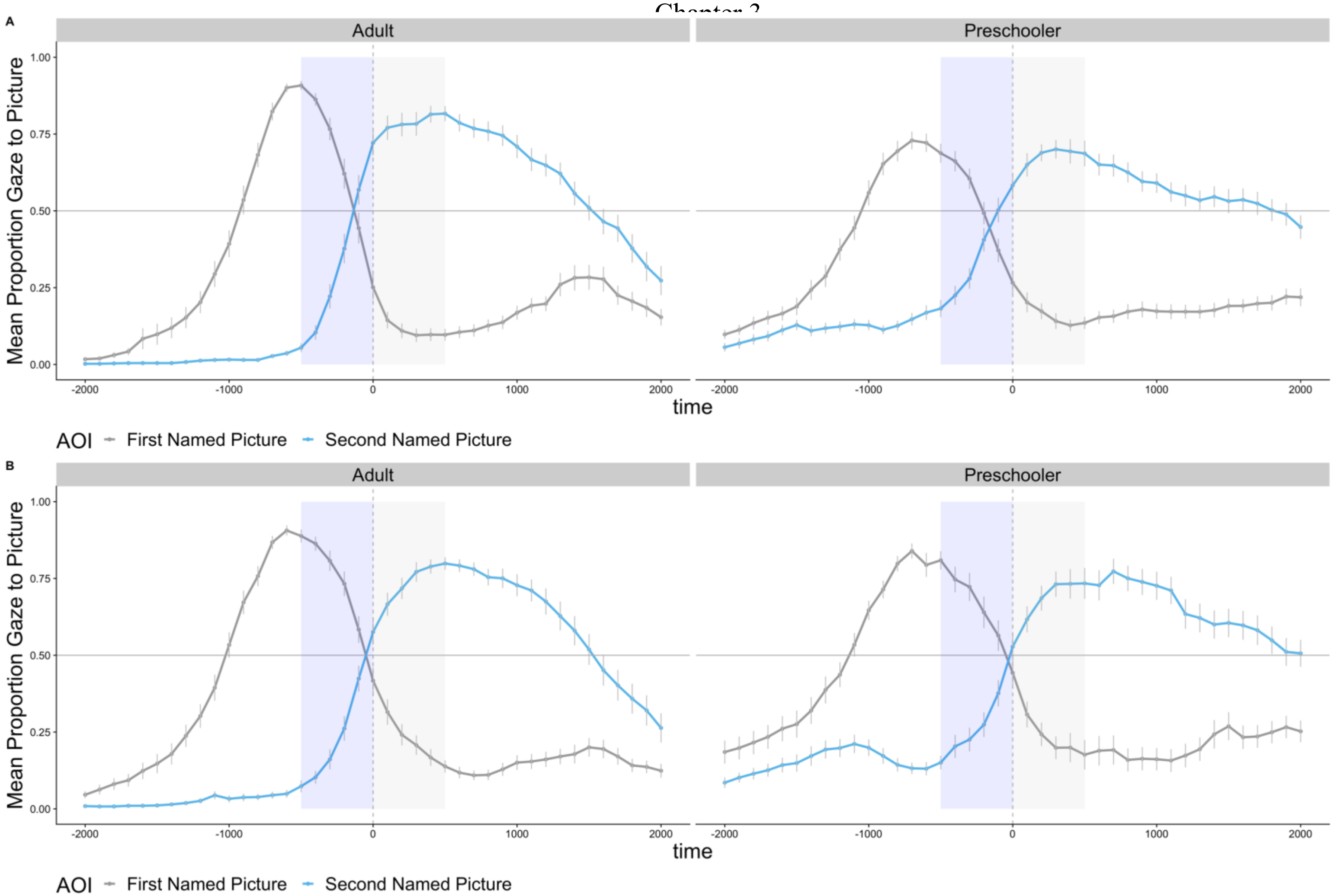

Figure 10. Mean proportion of looks to the first and second named picture for children and adults across time (in ms) in Experiment 1 (A) and Experiment 2 (B). Time 0 corresponds to the time participants began naming the first item. Bars indicate ±1 standard error. Only plotted trials where the second item was named. The shaded areas are the relevant time windows used for our analysis (500ms before participants named the first item and 500ms after they named the first item).

Figure 10 shows the mean proportion of looks to the first and second named picture within each of our 500ms time windows as a function of age. Table 1 reports our models for our first time window looking only at preschoolers' average fixations (fixations ~ AOI) followed by a combined model which looks at the effect of age (fixations ~ AOI * Age Group); Table 2 reports our models for the second time window.

Within our first time window (500ms before naming the first picture), pre-schoolers fixated on the first named object more than the second named object (51% [SD=.16] vs. 36% [SD=.17]). In our second time window (500ms after naming), pre-schoolers fixated on the second named object more than the first named object (67% [SD=.18] vs 17% [SD=.12]). Overall, this suggests that pre-schoolers' gaze was coordinated with their speech output: they looked significantly more at the picture they were just about to name (i.e., the first named picture) than to the other picture, and shifted their gaze to the second named picture (and away from the first picture) as they began naming the first picture.

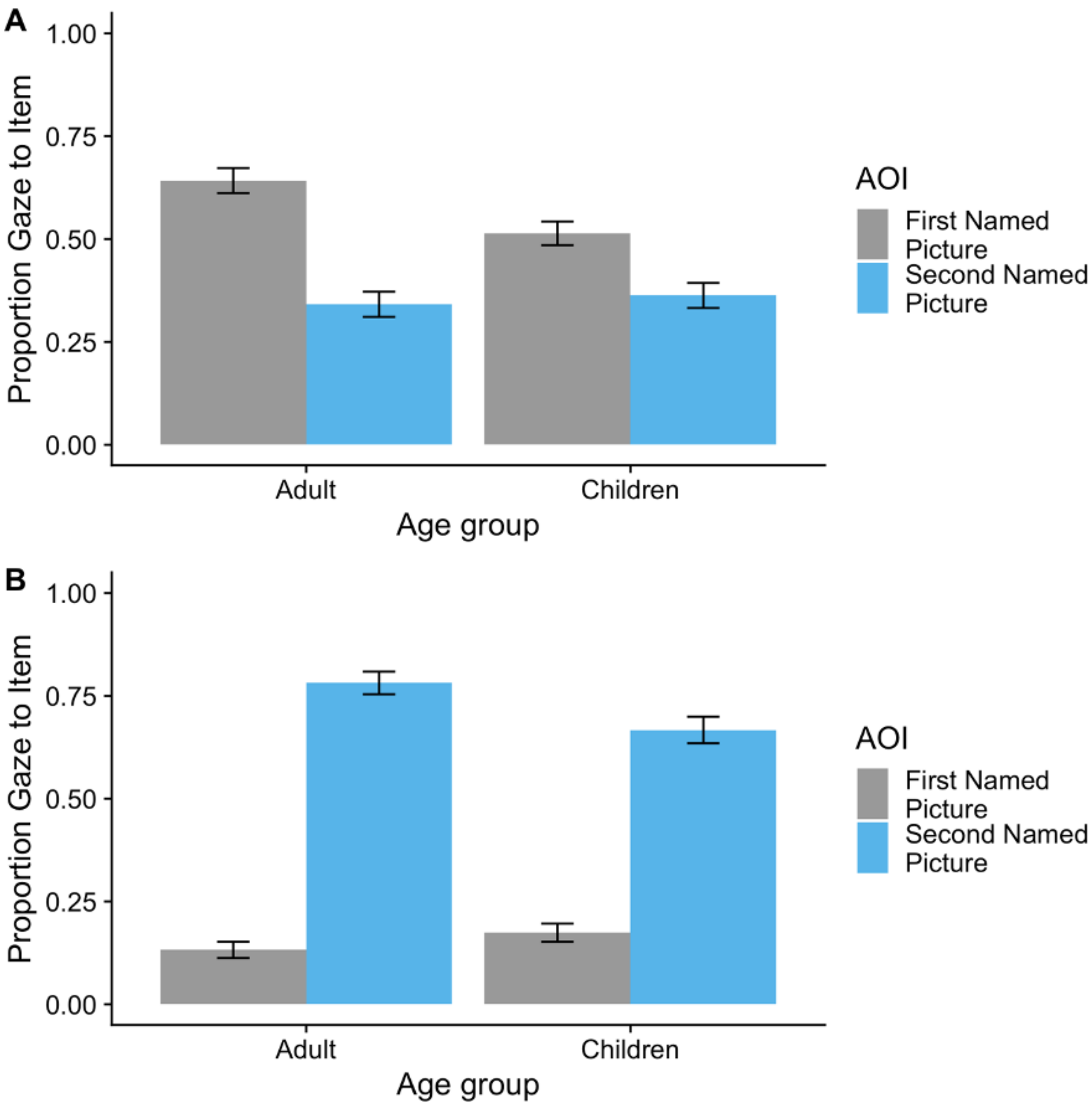

Figure 11. Mean proportion of looks to the first and second named picture for children and adults in the 500ms time window before they begin naming the first picture (A) and the 500ms time window after they began naming the first picture (B) in Experiment One. Bars indicate ±1 standard error. Only plotted trials where the second item was named.

Table 12. Linear Mixed Effects Model output for children and adults' average fixation score within our first (500ms before naming) and second time window (500ms after naming) in both experiments as a function of AOI (first or second named picture) and Age Group.

			Experiment 1					Experiment 2				
		Predictor	**B**	**SE**	**t**	**CI**	***p***	**B**	**SE**	**t**	**CI**	***p***
Time Window 1 (500ms before naming)	**Child Only Model**	Intercept	.44	.01	30.07	[.41, .47]		.47	.01	38.42	[.45, .50]	
		AOI	.07	.01	6.41	[.05, .09]	$\chi^2(1)$=40.03, p<.001	.21	.01	20.37	[.19, .23]	$\chi^2(1)$=348.06, p<.001
	Combined Analysis	Intercept	.47	.01	67.77	[.46, .49]		.48	.01	80.04	[.40, .57]	
		AOI	.12	.02	5.43	[.07, .16]	$\chi^2(1)$=25.68, p<.001	.22	.01	38.99	[.14, .28]	$\chi^2(1)$=1215.10, p<.001
		Age Group	.02	.01	3.72	[.01, .04]	$\chi^2(1)$=12.89, p<.001	.01	.01	1.51	[-.001, .02]	$\chi^2(1)$=2.43, p=.12
		AOI * Age Group	.04	.02	1.91	[-.00, .07]	$\chi^2(1)$=3.69, p=.05	.01	.01	1.54	[-.05, .09]	$\chi^2(1)$=2.36, p=.12
Time Window 2 (500ms after naming)	**Child Only Model**	Intercept	.43	.01	31.64	[.40, .46]		.46	.02	30.81	[.43, .49]	
		AOI	-.27	.01	-27.18	[-.29, -.26]	$\chi^2(1)$=549.65, p<.001	-.21	.01	-20.07	[-.23, -.19]	$\chi^2(1)$=338.39, p<.001
	Combined Analysis	Intercept	.45	.01	60.39	[.44, .46]		.47	.01	68.68	[.46, .49]	
		AOI	-.31	.01	-56.57	[-.32, -.30]	$\chi^2(1)$=2091.30, p<.001	-.23	.01	-37.06	[-.23, -.21]	$\chi^2(1)$=1116.20, p<.001
		Age Group	.01	.01	2.30	[.002, .03]	$\chi^2(1)$=5.67, p<.05	.01	.01	1.48	[-.003, .02]	$\chi^2(1)$=2.50, p=.11
		AOI * Age Group	-.03	.01	-5.59	[-.04, -.02]	$\chi^2(1)$=31.04, p<.001	-.01	.01	-1.79	[-.02, .001]	$\chi^2(1)$=3.21, p=.07

Our combined model looked at whether pre-schoolers' looking behaviour was different to our adult controls. Within our first time window, our participants fixated on the first named picture more than the second named picture (58% [SD=.17] vs. 35% [SD=.17]). However, we also found a main effect of age: adults had a greater fixation score than pre-schoolers did (49% [SD=.03] vs. 44% [SD=.08]), which could reflect the fact pre-schoolers had noisier data than adults (and therefore fixated on the pictures less because they would look away). There was also a marginal interaction: adults fixated on the first named item more than pre-schoolers (64% [SD=.16] vs. 51% [SD=.15]; *B* =.06, *SE*=.02, *t*=3.06, *CI*=[.02, .10], $\chi^2(1)$=8.95, *p*<.01) but there was no difference in adults and pre-schoolers fixations to the second named picture (34% [SD=.16] vs. 36% [SD=.17]; *B* = -.01, *SE*=.02, *t*=-.48, *CI*=[-.05, .03], $\chi^2(1)$=.23, *p*=.63).

Our combined model for our second time window showed similar results: overall, participants fixated more on the second picture than the first picture (72% [SD=.17] vs. 15% [SD=.12]). Again, we found a main effect of age: adults had a greater fixation score than pre-schoolers did (45% [SD=.06] vs. 42% [SD=/-7]), and we found a significant interaction: there were no age differences in fixations to the first named picture (adults: 13% [SD=.11] vs. 17% [SD=.12]; *B* =-.02, *SE*=.01, *t*=-1.23, *CI*=[-.04, .01], $\chi^2(1)$=1.64, *p*=.20), but adults fixated on the second named picture more than pre-schoolers did (78% [SD=.15] VS 67% [SD=.18]; *B* =.05, *SE*=.02, *t*=3.06 *CI*=[.02, .09], $\chi^2(1)$=9.44, *p*<.01). Overall, these results suggest that pre-schoolers' and adults' eye movements were similar before and after speech: they tended to look more to the first named picture than the second named picture before speaking and shifted their gaze to the second named picture once they began naming the first.

Gaze Durations. Our next analysis focuses on the time during which participants fixated on a picture before they finished naming that picture. Gaze durations have often been used as a measurement of processing time for a named object. We determined participants' gaze durations by calculating the total time participants fixated on the named picture until they finished naming it (the offset of named picture).[14] That is, we calculated the duration of each fixation to the named picture from the start of the trial until the moment they finished naming it (unless their first fixation was affected by our cue manipulation and they did not name the cued object), and if there were multiple fixations, summed the total of these, to calculate total gaze durations. We tested whether there were differences in the total time preschoolers fixated on each object (model form: Gaze Duration ~ Utterance Term) followed by a model looking at age differences (model form: Gaze Duration ~ Utterance Term * Age Group; see Table 13 for all model output).

[14] Because of our cue manipulation (reported in a further study), some participants' first fixations were to a picture that was not the named picture (e.g., from Fig 1; if the car was cued and so their first fixation was to the car but they named the doll first in their utterance). Thus, we excluded these fixations from our gaze duration analysis (i.e., in our example, we excluded the first fixation to car, but included later ones)

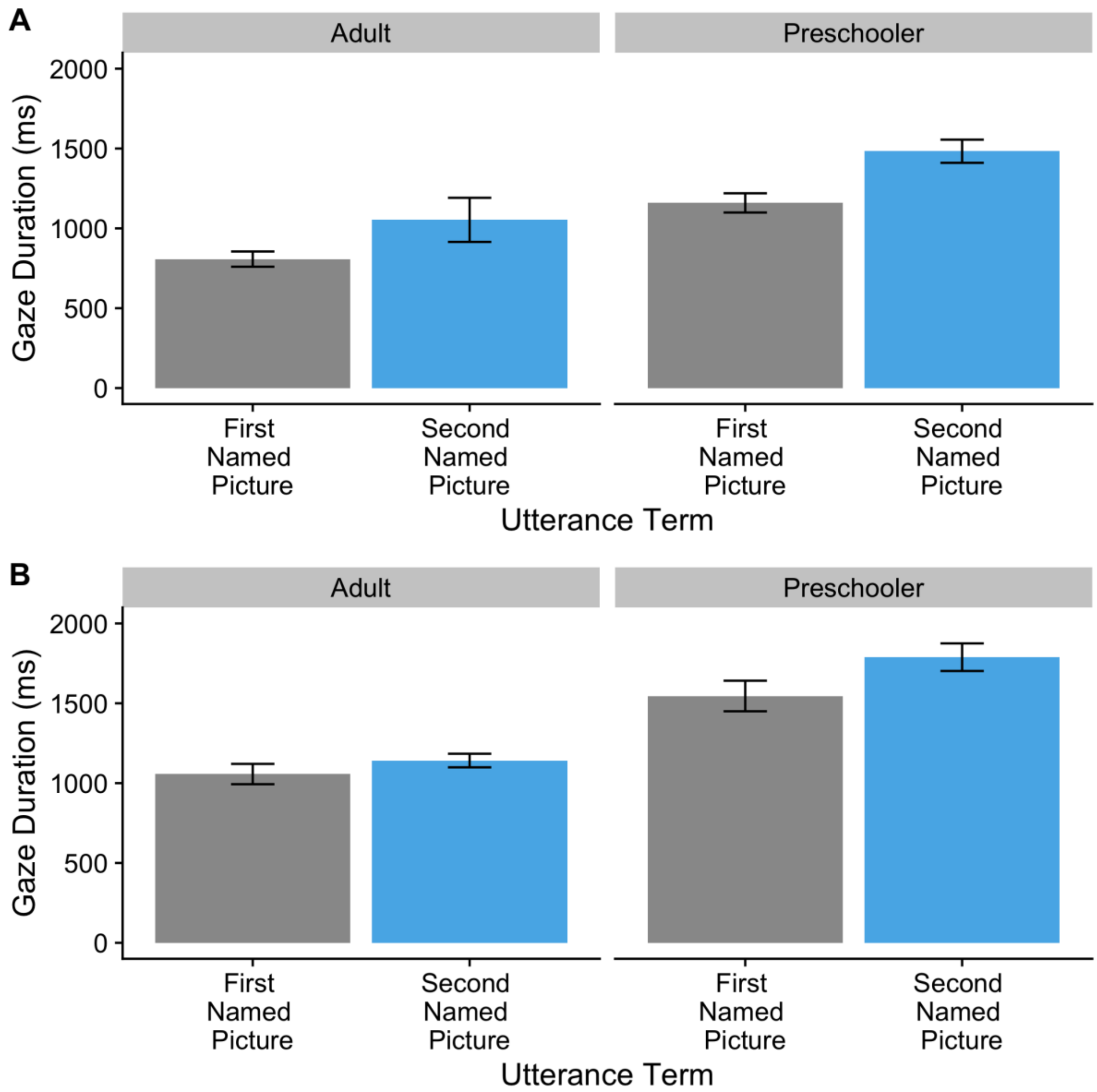

Figure 12. Mean gaze durations spans for each picture (first named or second named) for children and adults in Experiment 1 (A) and Experiment 2 (B). Bars indicate ±1 standard error. Only plotted trials where the second item was named.

Table 13. Linear Mixed Effects Model output for children and adults' log transformed gaze duration in both experiments as a function of the character they were fixating on.

		Experiment 1					Experiment 2				
	Predictor	**B**	**SE**	**t**	**CI**	***p***	**B**	**SE**	**t**	**CI**	***p***
Child Only Model	Intercept	6.97	.06	120.99	[6.86, 7.09]		7.21	.05	160.21	[7.12, 7.30]	
	Utterance Term	-.17	.02	-6.89	[-.22, -.12]	$\chi^2(1)$=25.92, *p<.001*	-.14	-.03	-4.78	[-.20, -.09]	$\chi^2(1)$=16.83, *p<.001*
Combined Analysis	Intercept	6.76	.04	170.18	[6.68, 6.84]		6.99	-.03	212.06	[6.93, 7.06]	
	Utterance Term	-.11	.02	-6.23	[-.15, -.08]	$\chi^2(1)$=29.68, *p<.001*	-.09	.02	-4.08	[-.13, -.05]	$\chi^2(1)$=15.11, *p<.001*
	Age Group	-.17	.02	-5.86	[-.22, -.11]	$\chi^2(1)$=25.91, *p<.001*	-.17	.03	-6.43	[-.22, -.12]	$\chi^2(1)$=31.85, *p<.001*
	Utterance Term * Age Group	.04	.02	2.38	[.01, .08]	$\chi^2(1)$=5.48, *p<.05*	.05	.02	2.37	[.01, .09]	$\chi^2(1)$=5.53, *p<.05*

Figure 11 shows the average gaze durations for adults and pre-schoolers. Overall, pre-schooler's gaze durations were shorter for the first named picture than the second named picture (1159ms [SD=335.73] vs. 1483ms [SD=403.88]). Our combined model showed the same pattern: overall, gaze durations were shorter for the first named picture than the second named picture (986ms [SD= 347.55ms] vs. 1272ms [SD=636.33]); adults' gaze durations were shorter than pre-schoolers' gaze durations (928ms [SD=470.05] vs. 1323ms [SD=317.94]). We also found a significant interaction: adults' gaze durations were shorter than children's gaze durations for both the first named picture (807.54ms [SD=756.29] vs. 1159ms [SD=335.73]; B = -.14, SE= .04 , t= -3.86, CI=[-.21, -.07], $\chi^2(1)$= 13.47, p<.001) and the second named picture (1053ms [SD=756.29] vs. 1483ms [SD=403.89ms]; B = -.23, SE= .03, t= -6.28, CI=[-.28, -.15], $\chi^2(1)$= 29.28, p<.001). Overall, this suggests that participants required more processing time for the second named picture, but pre-schoolers required more time overall.

Eye-Voice Spans. Eye-voice spans can tell us about how long participants took to plan the named word before they began speaking. To calculate eye-voice spans, we focused only on fixations that ended before the participant began naming the appropriate picture (i.e., only looked at fixations to the first named picture that ended before first named picture onset; and only looked at fixations to the second named picture that ended before second item onset). Eye-voice spans were calculated as the time between the onset of speakers' final fixation to the named picture and the onset of naming that picture.

Figure 12 shows the mean eye-voice span for each picture. As before, we first tested whether there were differences in pre-schoolers' eye-voice spans[15] as a function of the named picture (model form: Eye Voice Span ~ Utterance Term), and then whether there were any age differences (model form: log transformed Eye Voice Span ~ Utterance Term * Age Group; see Table 14 for all model output).

Overall, there were no differences in pre-schoolers' eye-voice spans for the first or second named picture (1133ms [SD=194.63] vs. 1176ms [SD=275.61]). Our combined model showed that, overall, participants' eye-voice spans were greater for the first named picture than the second named picture (1054ms [SD=210.51] vs. 991ms [SD=298.73]); adults had a shorter eye-voice span than pre-schoolers did (902ms [SD=158.39]VS 1156ms [SD=191.56]); and there was a significant interaction – adults had a longer eye-voice span for the first named picture than the second named picture (973ms [SD=197.75] vs. 799ms [SD=177.58]; B = .12, SE= .02, t= 5.65, CI=[.08, .17], $\chi^2(1)$= 22.66, p<.001), but pre-schoolers' did not show any difference. These findings suggest that pre-schoolers took the same amount of time to plan the first and second word, whereas adults took less time to plan their second word.

[15] For our analyses, we log-transformed eye-voice spans.

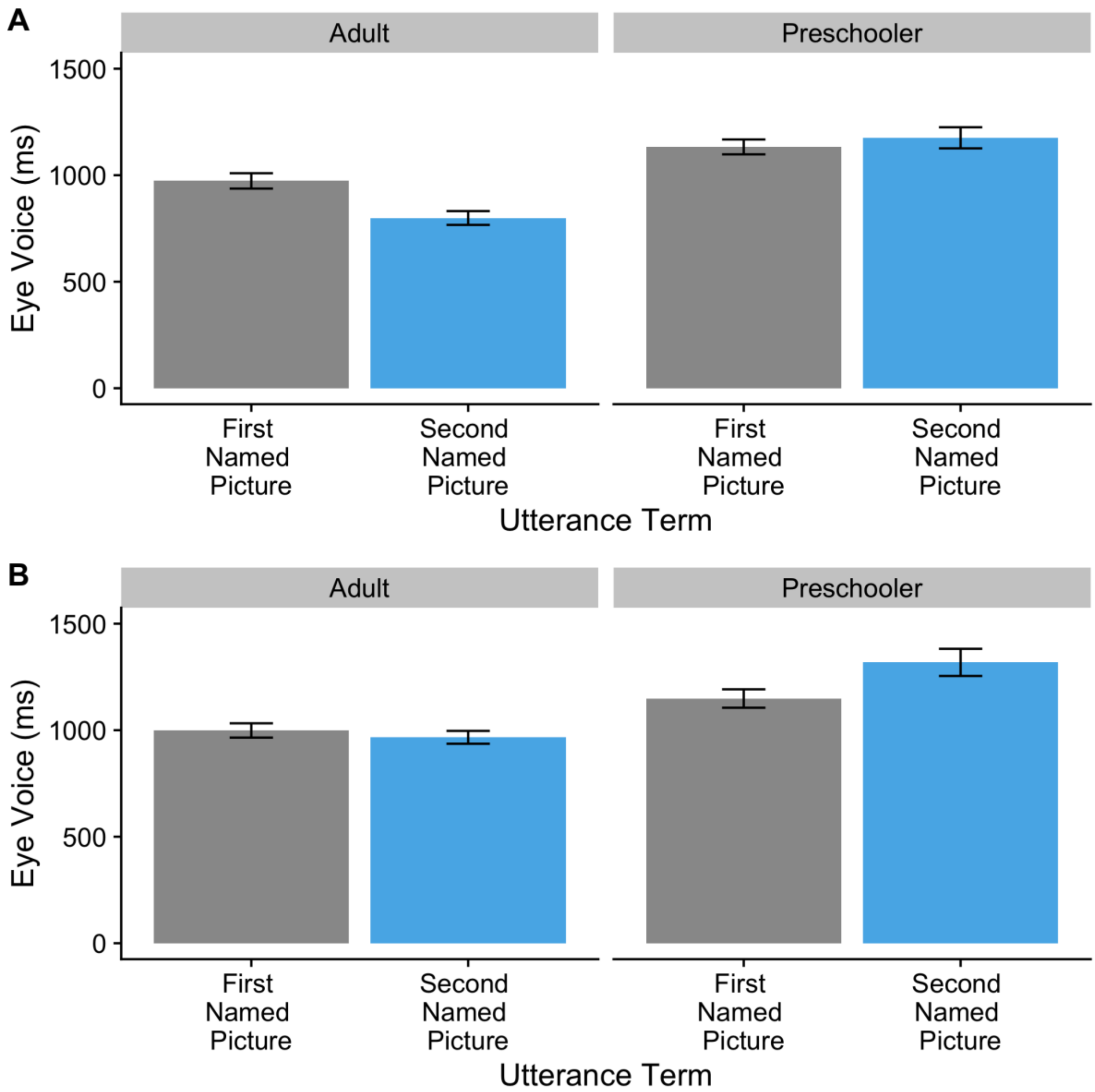

Figure 13. Mean eye-voice spans for each picture (first or second) for children and adults in Experiment 1 (A) and Experiment 2 (B). Bars indicate ±1 standard error. Only plotted trials where the second item was named.

Table 14. Linear Mixed Effects Model output for children and adults' log transformed eye-voice spans in both experiments as a function of the character they were fixating on.

		Experiment 1					Experiment 2				
	Predictor	**B**	**SE**	**t**	**CI**	***p***	**B**	**SE**	**t**	**CI**	***p***
Child Only Model	Intercept	6.92	.04	175.48	[6.84, 6.99]		6.96	.04	193.19	[6.89, 7.03]	
	Utterance Term	-.01	.02	-.62	[-.06, .03]	$\chi^2(1)$=.38, *p*=.54	-.08	.02	-3.74	[-.12, -.04]	$\chi^2(1)$=13.86, *p*<.001
Combined Analysis	Intercept	6.75	.03	232.96	[6.69, 6.81]		6.81	.03	241.19	[6.75, 6.86]	
	Utterance Term	.07	.02	4.56	[.04, .11]	$\chi^2(1)$=18.67, *p*<.001	-.01	.01	-1.01	[-.04, .01]	$\chi^2(1)$=1.05, *p*=.31
	Age Group	-.13	.02	-5.55	[-.18, -.09]	$\chi^2(1)$=25.03, *p*<.001	-.12	.02	-4.73	[-.17, -.07]	$\chi^2(1)$=18.89, *p*<.001
	Utterance Term * Age Group	.07	.02	3.98	[.03, .10]	$\chi^2(1)$=14.42, *p*<.001	.05	.02	3.29	[.02, .08]	$\chi^2(1)$=10.18, *p*<.01

Gaze-to-Speech Lag. To further examine how pre-schoolers coordinate gaze with speech, we calculated their gaze-to-speech lag – the time between the offset of participants' final gaze to a named picture and the onset of its spoken name. Positive gaze-to-speech lags indicate that speakers stopped fixating on the named picture before naming it, whereas negative numbers indicate that speakers stopped fixating on the picture after they began speaking.

As before, we first tested whether there were differences in pre-schoolers' gaze-to-speech lag[16] as a function of the named picture (model form: Gaze-To-Speech Lag ~ Utterance Term), and then whether there were any age differences (model form: Gaze-To-Speech Lag ~ Utterance Term * Age Group; see Table 15 for all model output).

Figure 13 shows the mean gaze-to-speech lag for each picture for each age group. Overall, pre-schoolers had a longer gaze-to-speech lag for the first named picture than the second named picture (298.23ms [SD=163.73] vs. 61.67ms [SD=213.06]).

[16] For our analyses, we log-transformed gaze-to-speech lags. Because we had negative numbers, we added a constant (maximum gaze-to-speech lag) to each value.

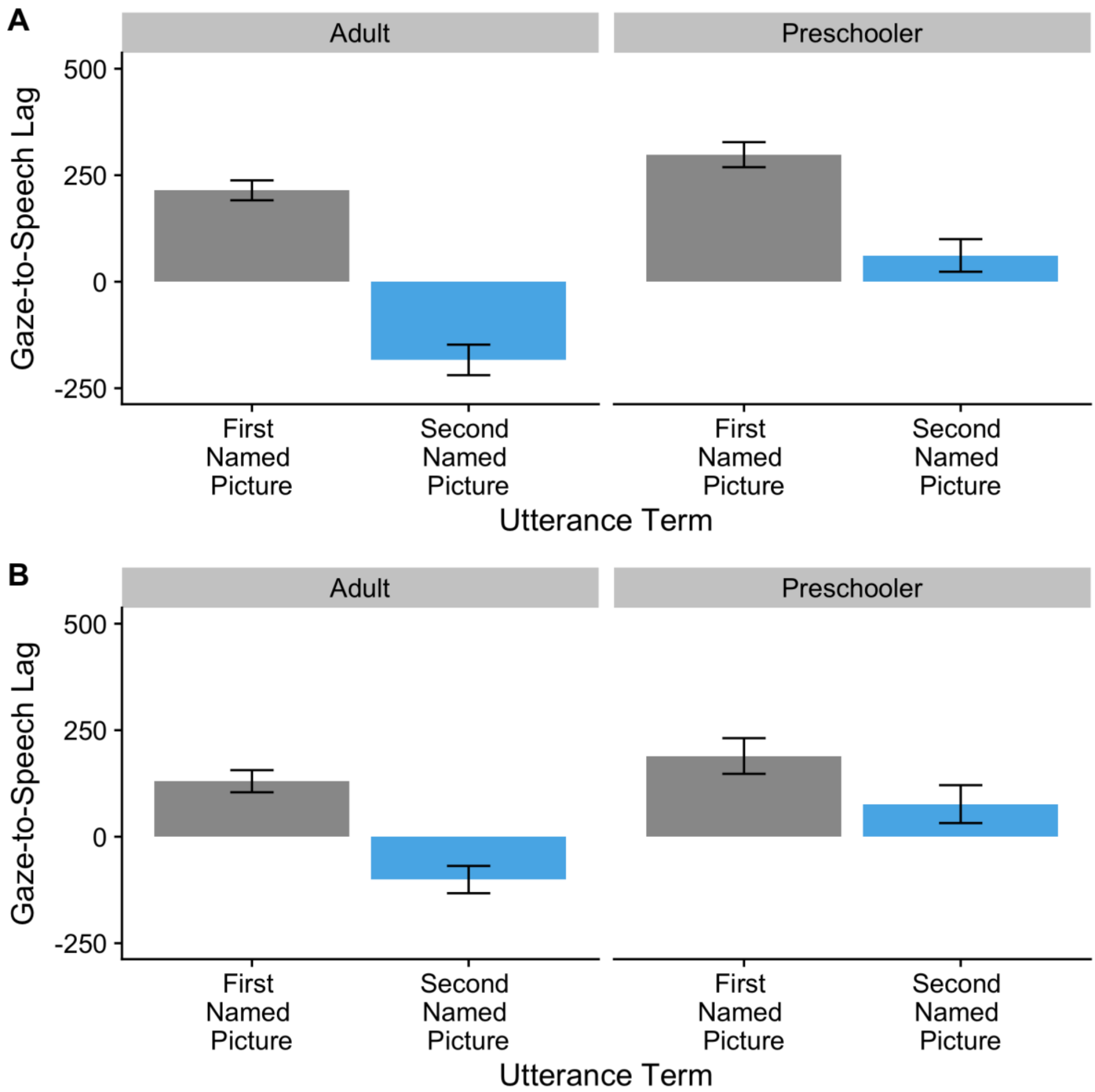

Figure 14. Mean gaze-to-speech lag for each picture (first or second) for children and adults in Experiment 1 (A) and Experiment 2 (B). Bars indicate ±1 standard error. Only plotted trials where the second item was named.

Table 15. Linear Mixed Effects Model output for children and adults' log transformed gaze-to-speech lags in both experiments as a function of the character they were fixating on.

		Experiment 1					Experiment 2				
	Predictor	**B**	**SE**	**t**	**CI**	***p***	**B**	**SE**	**t**	**CI**	***p***
Child Only Model	Intercept	7.98	.01	947.95	[7.96, 7.99]		8.72	.004	1830.31	[8.72, 8.73]	
	Utterance Term	.05	.01	6.08	[.05, .06]	$\chi^2(1)$=23.38, p<.001	.01	.004	2.40	[.001, .02]	$\chi^2(1)$=5.45, p<.05
Combined Analysis	Intercept	7.94	.01	1093.54	[7.93, 7.95]		8.72	.003	2785.38	[8.71, 8.72]	
	Utterance Term	.07	.01	13.16	[-.06, .08]	$\chi^2(1)$=80.06, p<.001	.02	.001	13.54	[.01, .02]	$\chi^2(1)$=177.73, p<.001
	Age Group	-.03	.01	-4.23	[-.04, -.02]	$\chi^2(1)$=16.28, p<.001	-.01	.003	-2.82	[-.01, -.003]	$\chi^2(1)$=7.70, p<.01
	Utterance Term * Age Group	.02	.01	3.31	[.01, .03]	$\chi^2(1)$=10.50, p<.01	.01	.001	4.29	[.002, .01]	$\chi^2(1)$=18.33, p<.001

Our combined model showed that overall participants' gaze-to-speech lag was longer for the first named picture than the second named picture (257ms [SD=152.02] vs. -59ms [SD=237.89]); adults' gaze-to-speech lag was shorter than children's (meaning that they stopped looking to the fixated on picture closer to speech onset; 14ms [SD=140.58] VS. 179ms [SD=136.15]); and finally, there was a significant interaction: adults had a smaller gaze-to-speech lag for the second named picture than children did (-184ms [SD=196.23] vs 62ms [SD=213.06]; *B* = -.04, *SE*= .01 , *t*= -4.16, *CI*=[-.07, -.02], $\chi^2(1)$=15.54, *p*<.05), likely because they continued looking to the second named picture after naming it. Adults still had a smaller gaze-to speech lag (i.e., fixated on the object closer to the time they name it) than children for the first named picture, but the difference was less pronounced than for the second named picture (215ms [SD=128.07] vs 298ms [SD=163.73]; *B* = -.02, *SE*= .01, *t*= -1.95, *CI*=[-.03, .00], $\chi^2(1)$= 3.88, *p*<.05). Overall, these findings suggest that children finished fixating on a to-be-named picture *before* adults do (see also Fig 14 for a timeline of when participants finished fixating on the first named picture relative to first item onset). This means that pre-schoolers stopped gazing at the named picture earlier than adults did. We discuss this surprising finding in our discussion below.

Gaze Shifts. To determine when participants began fixating on the second named picture after fixating on the first named picture (i.e., shifted their gaze to the second object), we also calculated the time participants began fixating on the second named picture after their final fixation to the first named picture had ended prior to naming the first picture.

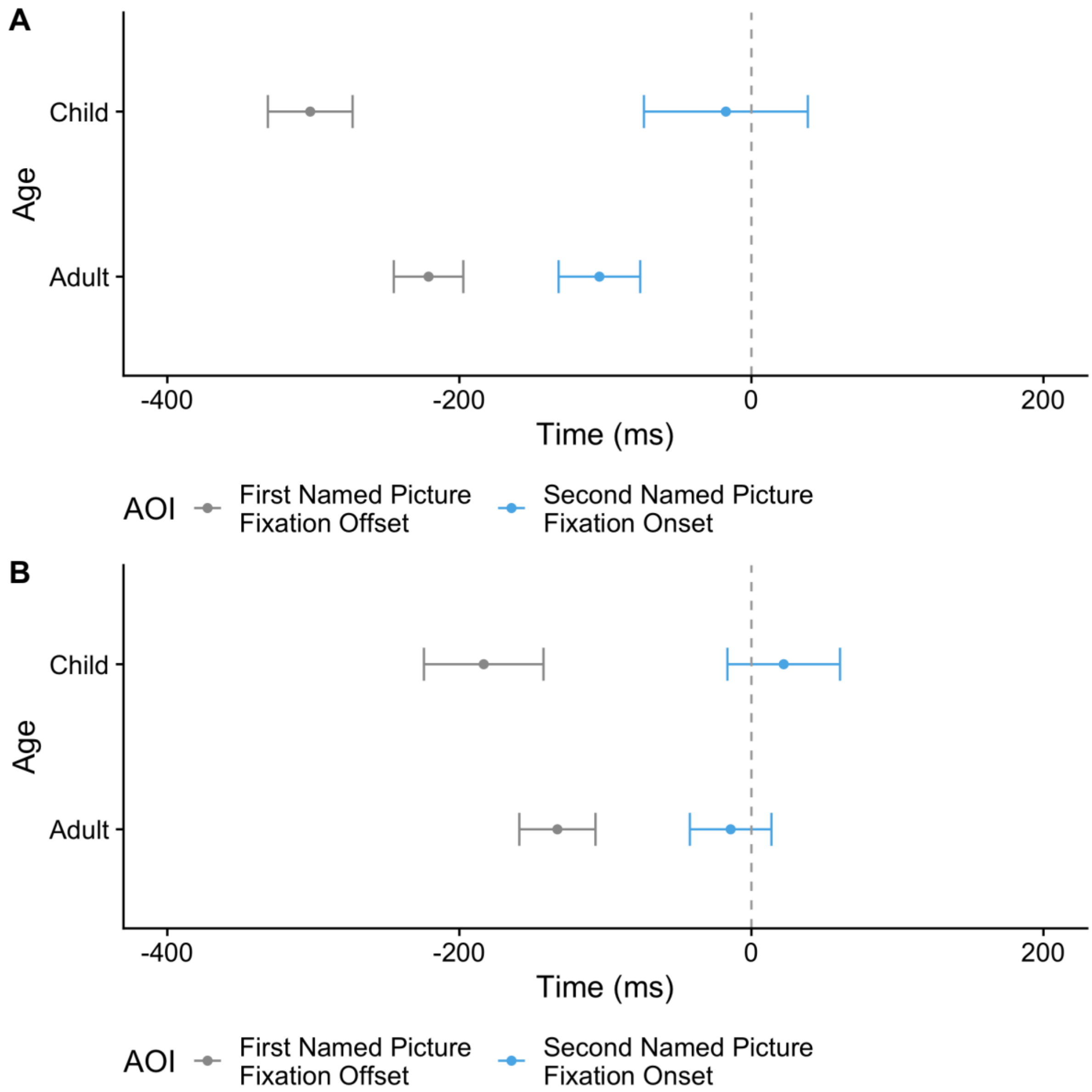

Figure 15. Mean offset of first named picture fixation (in grey) and mean onset of second named picture fixation (in blue) relative to first object onset (time 0 = first item onset) for children and adults in Experiment 1 (A) and Experiment 2 (B). Bars indicate ±1 standard error. Only plotted trials where the second object was named.

Table 16 Linear Mixed Effects Model output for children and adults' gaze shifts: log-transformed onset times in both experiments as a function of whether onset type was the onset of their fixation to the second named picture or onset of naming the first item

		Experiment 1					Experiment 2				
	Predictor	**B**	**SE**	**t**	**CI**	***p***	**B**	**SE**	**t**	**CI**	***p***
Child Only Model	Intercept	7.40	.04	188.93	[7.32, 7.48]		7.48	.04	208.56	[7.41, 7.55]	
	Onset Type	.05	.01	4.04	[.02, .07]	$\chi^2(1)=16.15$, $p<.001$	.01	.01	.99	[-.01, .04]	$\chi^2(1)=.99$, $p=.32$
Combined Analysis	Intercept	7.11	.03	243.42	[7.05, 7.17]		7.22	.03	246.86	[7.16, 7.28]	
	Onset Type	.06	.01	6.77	[.04, .08]	$\chi^2(1)=31.99$, $p<.001$	.02	.01	3.18	[.01, .03]	$\chi^2(1)=10.08$, $p<.01$
	Age Group	-.22	.02	-9.46	[-.27, -.18]	$\chi^2(1)=53.78$, $p<.001$	-.19	.03	-7.72	[-.25, -.15]	$\chi^2(1)=41.75$, $p<.001$
	Onset Type * Age Group	.01	.01	1.52	[.003, .03]	$\chi^2(1)=2.37$, $p=.12$	.01	.01	.94	[-.01, .02]	$\chi^2(1)=.89$, $p=.35$

Fig 14 shows the mean time participants stopped fixating on the first named object and the mean time they began fixating on the second named object (relative to the time they began naming the first object; see also Table 17).[17] On average, children began looking to the second named object 17ms [SD=312.47] before they began naming the first object (adults began fixating on second named picture 104ms [SD=153.06] before naming the first object). We tested whether there were age differences in the onset of fixations to the second named picture relative to the onset[18] of naming the first picture (model form: Onset of Fixation to Second Picture (relative to onset) ~ Age Group). Numerically, children looked to the second named picture (relative to onset) somewhat slower than adults, but this difference was not statistically significant (B = -.01, SE= .01, t= -1.60, CI=[-.01, .001], $\chi^2(1)$=2.61, p=.11).

To test whether participants began fixating on the second named picture before or after they began naming the first picture, we first ran a model testing whether there were differences in the time pre-schoolers began fixating on the second named picture and the time they began naming the first named picture (model form: Time[19] ~ Onset Type, where Onset Type is naming onset time vs., second picture fixation onset time), and a second model to investigate age differences (model form: Time ~ Onset Type*Age Group; see Table 16 for all model output).

Our child only model showed that overall, the onset of naming was greater than the onset of fixating to the second object, indicating that children began fixating to the second named object before they began naming the first item (fixation onset: 1751ms [SD = 461.07]

[17] We calculated this by finding the time participants finished fixating on the first named picture and the onset of their first fixation to the second named picture subtracted by first item onset.
[18] We log-transformed fixation onset times. Because we had negative numbers, we added a constant value (max value of fixation onset time) to each value prior to log-transforming.
[19] We log-transformed time variable.

vs. naming onset: 1780ms [SD=310.38]). Our combined model showed the same pattern: overall, participants began fixating to the second named picture before they named the first picture (fixation onset: 1419ms [SD = 529.25] vs. naming onset: 1482 [SD=414.06]); overall, adults were faster than children (for both onset of fixations to second picture and naming the first picture; 1124ms [SD=290.58] vs. 1767ms [SD=359.01]); and there was no interaction, suggesting that overall, children and adults showed the same pattern (see Table 16).

Table 17. Mean onset times for second fixations and naming in adults and pre-schoolers

	Onset Type	**Mean**	**SD**	**SE**
Adults	Second Fixation	1074ms	344.00	62.81
	Naming	1175ms	249.52	45.55
Pre-Schoolers	Second Fixation	1752ms	461.07	82.81
	Naming	1780ms	310.38	55.75

3.2.3. Discussion

Overall, our results from Experiment One suggest that pre-schoolers were planning each word in an incremental fashion, just like adults. They tended to fixate on the picture they were just about to name and then shifted their gaze to the next-named object just before they began speaking. Furthermore, pre-schoolers began fixating on the second-named object before they began naming the first, just like adults. This could indicate that they began planning the second named object whilst they were still planning later stages of the first word (i.e., parallel planning).

We also measured children and adults' gaze durations, eye-voice spans and gaze-to-speech lags. Gaze durations (defined as the total time participants fixated on a picture from trial onset until they finished naming the picture) can tell us about the processing time for a named object; eye-voice spans (the time between first fixating on an object and the onset of naming it) can tell us how long participants took to plan the named word before they began speaking; and gaze-to-speech lags (defined as the time between the offset of participants' final gaze to a named picture and the onset of its spoken name) can tell us about the coordination between eye gaze and speech.

Both children and adults had longer gaze durations for the second named picture than the first named picture, but we propose that this is for different reasons. Adults' gaze-to-speech lags indicated that they continued looking to the second named picture after they began naming (which would result in longer gaze durations for the second named picture, as these were measured as the sum of all fixations from the time participants stopped fixating on the first item until the offset of naming the second picture). This pattern is consistent with previous studies showing that adults tend to continue looking at the final picture they name (e.g., Meyer, Wheeldon & Konopka, 2012). However, adults' eye-voice spans were shorter for the second named picture than for the first named picture which could indicate they began extrafoveal processing of the second picture whilst planning their first word. Pre-schoolers, on the other hand, did stop fixating on the second named picture before naming it, and their eye-voice spans for the first and second named picture were equivalent. Therefore, the increase in gaze durations for the second named picture may indicate that it took longer for them plan that particular word, which could suggest that speaking has a detrimental effect on pre-schoolers' abilities to plan later aspects of an utterance.

3.3. Experiment 2b

Experiment One showed that pre-schoolers, like adults, not only coordinated their eye gaze with speech but also planned their utterance incrementally. Our results also suggest that both adults and pre-schoolers initiated planning of the second word whilst still planning later aspects of the first word (such as phonetic encoding and articulation). However, we asked participants to produce very simple sentences in Experiment One. It is unclear if this pattern would still hold for more complex sentences requiring participants to determine semantic roles and grammatical relations and generate a more complex constituent structure. Indeed, previous studies have shown that adult speakers will always plan in an incremental fashion regardless of utterance complexity, but the extent to which they plan in a parallel fashion is affected by the difficulty of the to-be-produced utterance (Malpass & Meyer, 2010).

In Experiment Two, we therefore showed participants the same array of pictures as in Experiment One, but instead asked them to produce more complex and complete sentences expressing the relative location of the objects (e.g., *the car is next to the doll*), thus requiring participants to determine semantic roles and grammatical relations, and generate a more complex constituent structure than in Experiment One.

3.3.1. Method

Participants. Participants were 30 adult native English speakers (M=20.04 years, range = 19-34, 15 female) & 42 native English-speaking children aged between 3 and 4 (M=45.79 months, range = 36-59, 18 female)[20] that had not participated in Experiment One. We recruited most children from private nursery schools around Edinburgh, and the rest from a database of interested families. Ethnicity and SES were not recorded, but were representative of the area (almost entirely white, predominantly from middle-class Scottish families).

Materials & Procedure. We used the exact same materials as in Experiment 1. The procedure was the exact same as Experiment 1, but this time, we instructed participants to produce locative phrases (e.g. “the car is next to the doll”).

[20] Note that this is the same sample as used in the study reported in Chapter 2; Experiment 2a.

Coding & Analysis. We coded and analysed the data in the same way as Experiment 1. We removed data from trials where participants took more than 5000ms to begin speaking or were 1.5 standard deviations below the age-appropriate mean (Adults: <1%; Children: 1.3%); where participants produced a conjoined noun phrase (children: 6%, adults: 0%); and where no gaze was measured (Adults: <1%; Children: 21.9%).

3.3.2. Results

Gaze Paths. As in Experiment One, our focus in this section is how our participants' eyes moved over time (before and after they began naming; see Fig 2B for mean proportion of looks to each picture) and whether they were more likely to fixate on the first or second named picture before or after naming. Figure 2B shows that both adults and children preferentially fixated the objects in the order they mentioned them, as in Experiment One, which suggests that they planned their utterance incrementally even for these more complex sentences; they looked to the first named picture before they began naming it, and then shifted their gaze to the second named picture just before they named the first item. We first ran separate models looking at whether pre-schoolers had a greater fixation score as a function of AOI (fixation score ~ AOI) in each time window, followed by two other models looking at the effect of age (fixation score ~ AOI * Age Group; see Table 12 for all model output).

Figure 15 shows the mean fixation scores to the first and second named picture within each of our 500ms time windows as a function of age. Overall, pre-schoolers had a greater fixation score for the first named picture than the second named picture within our first time window (500ms before naming; 65% [SD=.29] vs. 29% [SD=.17]); and had a greater fixation score for the second named picture than the first named picture within our second time window (500ms after naming; 67% [SD=.18] vs. 26% [SD=.20]), thus replicating our results from Experiment one. This suggests that pre-schoolers' gaze was still coordinated with their speech output, even when they were producing a more complex utterance; they looked significantly more at the picture they were just about to name and shifted their gaze to the next-named picture as they began naming.

Our combined model also found a main effect of AOI in both time windows; overall participants looked more to the first named picture than the second named picture in the 500ms before first name onset (68% [SD=.18] vs. 28% [SD=.16]), and looked to the second named picture more than the first named picture after first name onset (70% [SD=.17] vs. 25% [SD=.17]). Unlike in Experiment 1, there was no effect of age in either of our time windows: there was no difference in adults and children's fixation scores before (adults: 49% [SD=.03] vs. children: 47% [SD=.06]), or after naming (adults: 48% [SD=.03] vs. children: 46% [SD=.07]) nor was there a significant interaction in either time window. Overall, this suggests that both children and adults were planning the word of the first named picture before naming, and then began planning the word of the second named picture after they began naming the first.

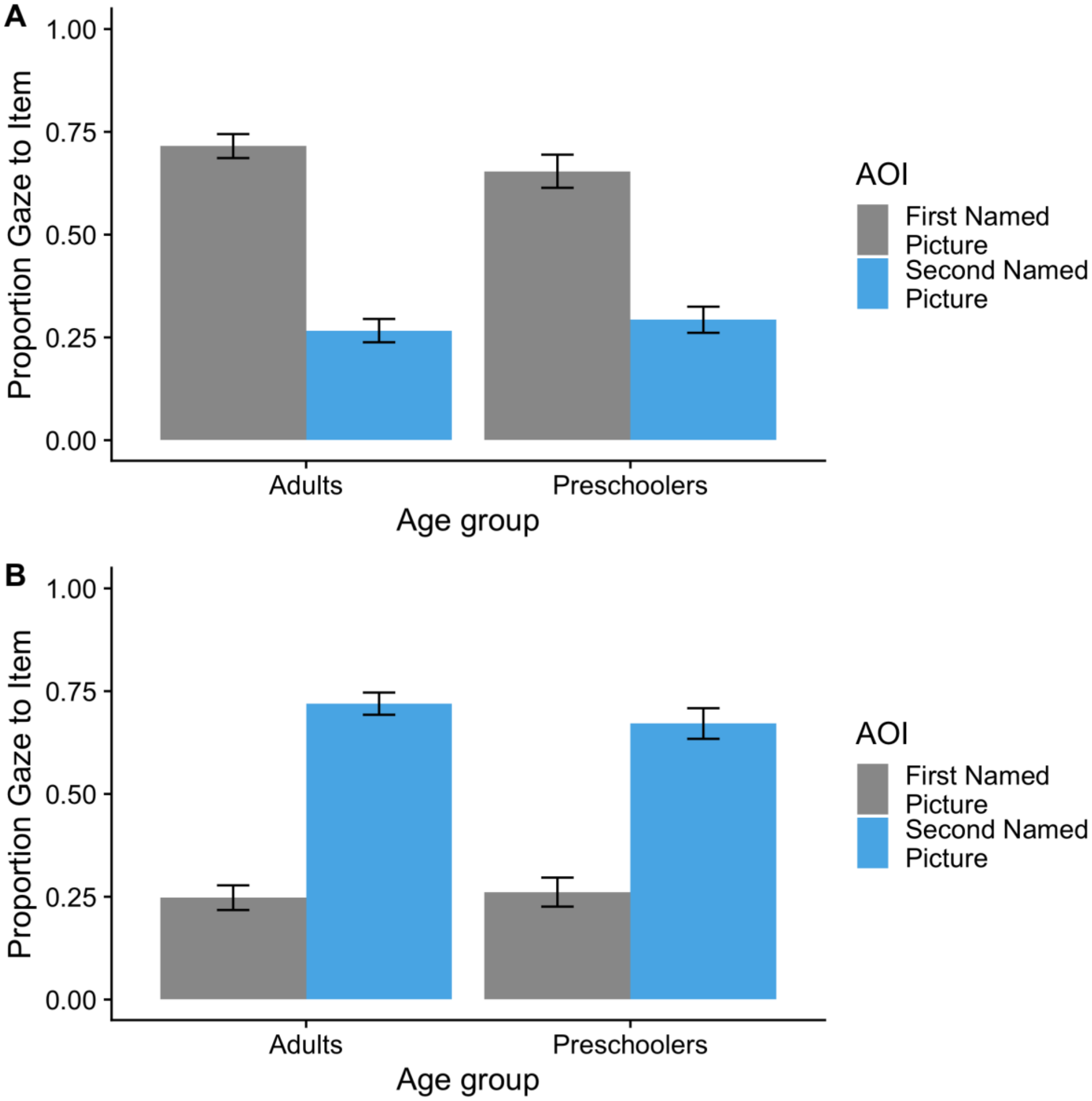

Figure 16. Mean proportion of looks to the first and second named picture for children and adults in the 500ms time window before they begin naming (A) and the 500ms time window after they began naming (B) in Experiment Two. Bars indicate ±1 standard error. Only plotted trials where the second item was named.

Gaze Durations. Our next analysis focused on the total time speakers fixated on each picture before they finished naming that picture (Fig 4B). We first tested whether pre-schoolers' gaze durations were affected by the picture they were naming (Gaze Duration ~ Utterance term) followed by a model looking at age differences (Gaze Duration ~ Utterance Term * Age Group; see Table 12 for all model output).

Pre-schoolers looked longer at the second named picture than the first named picture (1789ms [SD=457.38] vs. 1546ms [SD=506.34]), replicating our results from Experiment One. Our combined model showed that overall, participants had longer gaze durations for the second named picture than the first named picture (1454ms [SD=482.83] vs. 1293ms [SD=493.79]); adults had shorter gaze durations than pre-schoolers (1094ms [SD=262.16] vs. 1672 [SD=367.09]); and there was a significant interaction: children had longer gaze durations for the second named picture than the first named picture, but there was no difference in adults' gaze durations for either picture (1057ms [SD=348.15] vs. 1141ms [SD=232.92]; B = -.05, SE=.03 , t=-1.78 , CI=[-.05, .005], $\chi^2(1)$=3.13 , p=.08). Overall, this suggests that pre-schoolers required more processing time for the second named picture than adults did.

Eye-Voice Spans. As in Experiment One, we tested whether there were differences in pre-schoolers eye-voice spans (the time between first fixating on a picture to the onset of naming) as a function of the utterance term (Eye-Voice Span ~ Utterance Term), followed by a combined model looking at the effect of age (Eye-Voice Span ~ Utterance * Age Group; see Table 13 for model output).

Our combined model indicated that, overall, there was no difference in participants' eye-voice spans for each utterance term (first named picture: 1071ms [SD=218.58] vs. second named picture: 1136ms [SD= 314.98]); adults had a shorter eye-voice span than pre-schoolers (983ms [SD= 156.44] vs. 1232ms [SD=252.22]). However, there was a significant interaction: children had a longer eye-voice span for the second named picture than the first (1148ms [SD= 228.81] vs. 1318ms [SD=337.32] ; see Fig 5B)), but there was no difference in adults' eye voice spans for either picture (first named picture: 999ms [SD= 184.25] vs. second named picture: 967ms [SD= 165.11]; *B* = .02 , *SE*= .02, *t*= 1.19, *CI*=[-.01, .06], $\chi^2(1)$= 1.45, *p*=.23). This is different from Experiment 1's eye-voice span results, where children showed no differences in eye voice space for either picture, and adults had a shorter eye-voice span for the second named picture than the first named picture. We discuss the implications for this in our discussion.

Gaze-to-Speech Lag. Our next analyses focused on participants' gaze-to-speech lags (the time between the offset of participants' final fixation to the named picture and the onset of naming it). We first tested whether pre-schooler's gaze-to-speech lags were affected by utterance term or not (Gaze-to-Speech Lag ~ Utterance Term), followed by a combined model looking at whether there were differences as a function of age (Gaze-to-Speech Lag ~ Utterance Term * Age Group; See Table X for all model output).

We replicated our results from Experiment One. Our first model indicated that pre-schoolers had a longer gaze-to-speech lags for the first named picture than the second named picture (189ms [SD=221.55] vs 77ms [SD=235.29]), indicating that pre-schoolers finished fixating on the second named picture close to when they began naming it. Our combined model replicated our results from Experiment One: overall, participants had a longer gaze-to-speech lag for the first named picture than the second named picture (159ms [SD= 185.57] vs. -15ms [SD=223,31]); adults had shorter gaze-to-speech lags than children (15ms [SD=134.68] vs. 132ms [SD=184.91]); and there was a significant interaction: adults had a smaller gaze-to-speech lag for the second named picture than children did (-101ms [SD=175.50] vs. 77 [SD=235.30]; *B* = -.02, *SE*=.01 , *t*=-2.98, *CI*=[-.03, -.01], $\chi^2(1)$=8.15, *p*<.01), likely because adults continued looking to the second named picture after naming it. However, there were no age differences in participants' gaze-to-speech lags for the first named picture (130ms [SD=142.25] vs. 189 [SD=221.55]; B = -.004, SE= .003, t= -1.19, CI=[-.-1, .002], $\chi^2(1)$=1.49, p=.22).

Gaze Shifts. Our final set of analyses focused on when participants began looking to the second named picture. On average, children began looking to the second named object 17ms [SD=312.47] before they began naming it (adult second named picture fixation onset: 104ms [SD=153.06] before naming; see Fig 7B). We first tested whether there were age differences in the onset of fixations to the second named picture relative to the onset of naming the first picture (model form: Onset Fixation to Second Picture (relative to onset) ~ Age Group). Overall, there were no age differences in the time participants began looking to the second named picture (relative to naming the first picture; B = -.01, SE= .01, t= -.93, CI=[-.02, .01], $\chi^2(1)$=.89, p=.35), suggesting that children began looking to the second named picture (relative to naming the first item onset) at the same time as adults.

Our final set of models tested whether the onset of pre-schoolers' fixations to the second named picture was before or after the onset of naming the first picture (Onset Time ~ Onset Type) followed by a model testing for any age differences between adults and pre-schoolers (Onset Time ~ Onset Type * Age Group; see Table 6 for both model outputs).

In our pre-schoolers only model, there was no difference in onset times for either onset type (second picture fixation onset: 1974ms [SD=485.47] vs. naming first picture onset: 1947ms [SD=340.71]), suggesting that they began looking to the second named picture at the same time they began speaking (which contrasts with our results from Experiment One). However, our combined model found a main effect of onset type: overall, participants began fixating on the second named picture before they began naming (1608ms [SD=545.50] vs. 1610ms [SD=460.92]). There was also a main effect of age group: adults had faster onset times than children (1282ms [SD=318.66] vs. 1960ms [SD=407.32]). However, we did not find a significant interaction (despite the fact that the pre-schooler only model failed to find an effect of onset type). To investigate this further, we also ran a separate adult model, which found a main effect of onset type: adults onset times were faster for their second fixation onset than naming onset, indicating they began looking to the second picture before they began naming (1267ms [SD=341.21] vs. 1296ms [SD=313.70]; $B = -.02$, $SE= .01$, $t= 2.06$, $CI=[.001, .05]$, $\chi^2(1)=4.11$, $p<.05$). This failure to find a main effect of type in our pre-schooler only model could be due to the fact that the distribution of our pre-schooler onset data had an ex-gaussian shape (even though we removed values that were greater than 5000ms from our model; see Fig 9)[21], meaning that the pre-schooler only model was unable to accurately estimate the effect of onset type. Overall, these results suggest that both adults and pre-schoolers look to the second named picture before naming, but pre-schooler's ability to do this is less precise.

[21] When we run a model excluding values greater than 3500ms, we do find a main effect of onset type ($B = .03$, $SE= .01$, $t= 2,54$, $CI=[.01, .05]$, $\chi^2(1)=6.46$, $p<.05$)

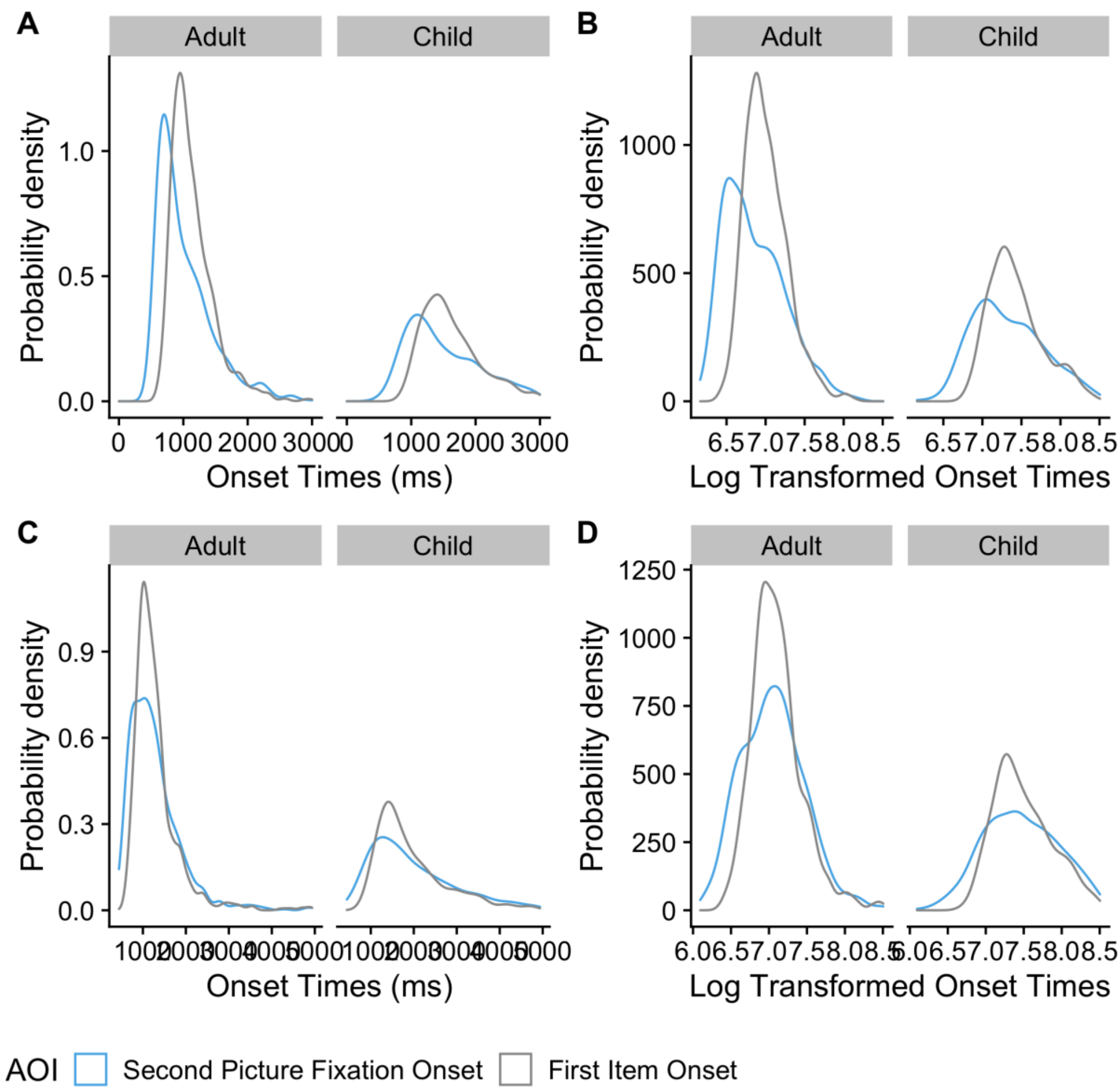

Figure 17. Fig 9: Distribution of onset times for children and adults in Exp1 (A & B) and Experiment Two (C & D). Plots B & D show the distribution of log-transformed onset times.

3.3.3. Discussion

As in Experiment One, both pre-schoolers and adults tended to fixate on each object in the order they mentioned them, suggesting they coordinated their eye gaze with speech. Furthermore, they fixated on the first named object before naming it and shifted their gaze to the second named object, indicating they were planning each word incrementally. Furthermore, we found some evidence that children and adults began looking to the second named picture before they began naming the first (in our combined model), but there was much more variation in our child data, suggesting that children were much more variable in the time they would begin fixating on the second named picture. However, pre-schoolers began fixating to the second named picture closer to naming onset in Experiment Two – where participants produced more words between object names – than in Experiment One (see Fig 7). As there were more function words between object names, participants had more time to begin preparing the second named object and as a result, could begin planning the second object later in their utterance (i.e., just in time planning). As a result, they fixated on the second item later. This suggests that children, as well as adults, may coordinate planning of the next noun so it becomes available as they need it. As a result, they don't need to buffer the first word for too long as they begin planning of the second noun (see also Griffin, 2003).

We also found that children's gaze durations were longer for the second named picture than the first named picture, replicating our results from Experiment One. These results, together with the fact that pre-schoolers had longer eye-voice spans and gaze-to-speech lags, suggest that planning of the second word took longer than planning for the first word, which could indicate that speaking interfered with planning of the second word.

3.4. General Discussion

The present experiments investigated the relationship between pre-schoolers' eye movements and their speech and consider what their eye movements can tell us about how they are retrieving word forms (i.e., do they retrieve word forms in an incremental fashion or not). In two eye-tracking experiments, we measured pre-schoolers eye-movements as they described pictures and compared them to adult controls' eye movements. We varied the complexity of the utterance participants had to produce across experiments: in Experiment One, participants produced simple conjoined noun phrases, whereas in Experiment Two, they produced more complex locative noun phrases. We showed that pre-schoolers, like adults, coordinated their eye movements with their speech output speech in both experiments: they tended to preferentially fixate on each object in the order that they mentioned them in. Furthermore, our results from both experiments showed that pre-schoolers fixated on the first named object before naming it and shifted their gaze to the next named object before they began naming the first. We also found that pre-schoolers had longer gaze durations than adults did, indicating that it took longer for children to retrieve each word than adults, but their gaze durations were longer for the second named picture than the first named picture in both experiments[22]. Children also had longer gaze-to-speech lags than adults did, suggesting they stopped fixating on each picture earlier than adults did. We discuss each of these findings in turn.

Firstly, our results are the first to suggest that pre-schoolers' gaze is coordinated with their speech in a similar way to adults. Previous studies have shown that adults tend to retrieve word forms in a highly sequential way, and this is reflected in their eye movements (Meyer et al., 1998). We found this same effect in pre-schoolers: children in our study first fixated on the object they were going to name and shifted their gaze to the next-named-object as they began naming. Overall, these results suggest that pre-schoolers planned their utterances in an incremental fashion, just like adults – regardless of whether they were planning a simple sentence (Experiment One) or a more complex sentence (Experiment Two). These findings are consistent with previous error-based and disfluency studies

[22] Note that adults also had longer gaze durations for the second named picture, but this is likely due to the fact they continued fixating on the second object whilst naming. Thus, adults' eye voice spans are more informative for understanding planning processes for the second named object

suggesting that pre-schoolers have smaller planning scopes than adults, and as a result plan small chunks of their utterance at a time (e.g., McDaniel et al., 2010).

Whilst we found clear similarities in adults and pre-schoolers eye movements, we also found some differences in the way children and adults coordinated their eye movements. In typical eye-tracking studies, a speaker will finish fixating on an object once they have completed phonological encoding of that object (e.g., Meyer et al, 1998), but they can deviate from this pattern when a task is easy (Meyer et al., 2012). Children had longer gaze-to-speech lags than adults in both experiments, indicating that they tended to finish fixating on each object before adults did. Adults' short gaze-to-speech lags in both experiments are likely due to the fact the task was very easy: they were repeating the same syntactic structure whilst naming words that have very low AoAs. Therefore, they didn't need to spend as much time planning upcoming words (which is reflected also in their shorter gaze durations). This is consistent with previous findings demonstrating that adults have shorter gaze-to-speech lags for easy tasks (Meyer et al., 2012). Furthermore, shorter looks to the second named picture (i.e., the final word) are due to the fact that adults continued fixating on the second picture (resulting in longer gaze durations for the second named picture), which is consistent with previous studies demonstrating that adults continue fixating on final-named picture (e.g., Meyer et al., 2012).

Children's long gaze-to-speech lags are quite a surprising finding. Previous studies have found that slower speakers have shorter gaze-to-speech lags than faster speakers (e.g., Griffin, 2003; Mortensen, Meyer & Humphreys, 2008; but see Belke & Meyer, 2007; Spieler & Griffin, 2006). Children typically have a much slower speech rate than adults (see McDaniel et al., 2018 for review) and so you might expect children to finish fixating on an object later than adults (and thus have a shorter gaze-to-speech lag than adults), perhaps to remove any interference of the next named picture (Meyer, 1998). Why might children have a longer gaze-to-speech lag? One potential explanation for this result centres on the efficiency of the developing production system. If we assume that children's fixations reflect lemma selection and retrieval of lexical word forms (like adults; Meyer et al., 1998), then the offset of a fixation would reflect the beginning of later planning processes of a word (e.g., phonetic and articulatory planning, etc.). Thus, longer gaze-to-speech lags may reflect the fact that children take a longer time to complete these later planning processes. In fact, Meyer et al (2012) showed that the less practice adults have in producing an object name, the longer their gaze-to-speech lag is, suggesting that later planning processes are affected by the availability of a given name. The names of objects in our study may have been less accessible to pre-schoolers than adults (even though we chose objects that had low AoAs), resulting in the longer gaze-to-speech lags. Overall, this would suggest that pre-schoolers' production processes are similar to adults', but pre-schoolers are just slower (and less efficient) to complete planning.

Our results also indicate that pre-schoolers do not plan in a wholly serial fashion (and thus, they could be planning in a parallel fashion): pre-schoolers, like adults, began fixating on the second named picture before they began naming the first (as seen from our gaze shift findings). If pre-schoolers planned in a wholly serial fashion (i.e., only begin planning the second word after completing processing of the first), we would expect them to begin fixating on the second picture *after* they began naming. However, it is unclear to what extent pre-schoolers did engage in parallel planning. Simply fixating on a picture does not necessarily imply that our pre-schoolers began linguistic encoding of that picture.

However, our additional eye gaze measurements indicate some ways in which parallel planning is something that children struggle with. Adults' eye-voice spans for Experiment

One were shorter for the second named picture than the first, which could be indicative of extra-foveal (and secondary) processing of the second word whilst their primary focus was planning the first word. Indeed, previous studies have shown that adult speakers' gaze durations are shorter for words they have processed extrafoveally, which is indicative of parallel planning (e.g., Morgan & Meyer, 2005; Malpass & Meyer, 2010). Furthermore, there was no difference in adults' eye-voice span in Experiment Two, suggesting that adults did not engage in extra-foveal processing of the second word here. This is consistent with previous findings suggesting that adults are less likely to engage in parallel processing for more complex utterances (Malpass & Meyer, 2010). However, pre-schoolers did not show this same pattern: their eye-voice spans were the same for both pictures in Experiment One, and in fact were longer for the second named picture than the first in Experiment Two. We can explain this finding in two ways. On the one hand, it is possible that children did not engage in any secondary processing of the second object (i.e., they were not engaging in parallel planning). Indeed, the fact that pre-schoolers showed no difference in their eye-voice spans for both pictures could indicate that it took the same amount of time to complete planning processes for the first and second word (and thus, did not engage in parallel planning). However, if this lack of difference in eye-voice spans reflects pre-schoolers not engaging in parallel planning, then you would expect to still see this pattern in Experiment Two (i.e., no difference in eye-voice spans between first and second picture). Instead, children had a longer eye-voice span for the second named picture than the first. It is possible that children in Experiment One were in fact engaging in parallel planning, just like adults, but they were less efficient in their ability to do so, resulting in there being no difference in their eye voice spans for either picture. However, in Experiment Two, which may have required more processing demands because they were producing a more complex sentence, they did not engage in parallel planning, and as a result, their eye-voice spans for the second picture were longer, perhaps because it took longer to retrieve the second word whilst they were speaking (see section below discussing pre-schoolers long gaze durations for the second named picture). However, this interpretation should be taken lightly. Further research is needed to test whether pre-schoolers are in fact engaging in parallel processing of upcoming words.

So far, we have claimed that children and adults' longer eye voice-spans for the second named picture indicate that they were not engaging in parallel planning in Experiment Two. But why might this be the case? It is possible that the complexity of the to-be-produced utterance in Experiment Two (compared to Experiment One) affected participants' ability to engage in parallel processing of the second word. However, both children and adults' gaze shifts to the second named picture were much closer to naming onset in Experiment Two than in Experiment One (indeed, the time pre-schoolers began fixating on the second named picture was much more variable in Experiment Two, with the distribution of gaze shift onsets being skewed to the left), which may indicate that they were engaging in 'just-in-time' planning, to minimise buffering of the first-named-object (and thus reducing processing load prior to naming; Griffin, 2003). That is, both pre-schoolers and adults began planning of the second word later in Experiment Two than in Experiment One because there were more words (and therefore more time) between naming the first and second picture. Therefore, they had more time to retrieve the second word, and as a result, they could shift their gaze to the second picture later. Indeed, this finding is consistent with previous findings showing that adults begin planning final words later when they have more time between the first and second word (Griffin, 2003). Thus, participants in our study did not have to engage in parallel planning in Experiment Two, and so they did not.

We now discuss the time it took for pre-schoolers' to retrieve a word whilst speaking. Pre-schoolers in our study had longer gaze durations for the second named picture than the first in both experiments, suggesting that they took longer to retrieve the second word. Both to-be-named pictures were matched for complexity and we counter-balanced whether they appeared on the left or right. This means that children's long gaze durations for the second named picture are unlikely to be due to differences in retrieving easier or harder words (that is, children are unlikely to be naming pictures that are easier to retrieve first, and harder pictures second), but is rather due to a difficulty in planning whilst also speaking at the same time resulting in a delay retrieving the second word. There are several reasons why this would be the case. On the one hand, it is possible that naming the first item resulted in interference retrieving the second name. Indeed, previous studies have shown that much younger children (two-year-olds) are affected by interference of previously retrieved words resulting in more naming errors (Gershkoff-Stowe, 2001; Gershkoff-Stowe & Smith, 1997; Gershkoff-Stowe, Connell & Smith, 2006). It is possible that the interference from the previously named word affected the efficiency of retrieving the next word, resulting in longer gaze durations.

Alternatively, pre-schoolers' longer gaze durations for the second named picture could reflect inefficient retrieval processes whilst engaging in a secondary task (i.e., speaking). That is, speaking interferes with children's abilities to retrieve subsequent word forms. Indeed, previous disfluency studies have shown that pre-schoolers tend to use more stalling devices in their own speech. This has been thought to show that children have smaller planning scopes than adults do, resulting in them having to stop and plan more often (and thus, producing more disfluencies, e.g., McDaniel et al., 2010). However, these disfluency findings could also reflect a slow-down in pre-schoolers' retrieval abilities whilst speaking. Even adults' retrieval processes are affected by interference from linguistic stimuli (Schriefers, Meyer & Levelt, 1990; Fairs, Bögels & Meyer, 2018; Fargier & Laganaro, 2016). For example, adults' naming times are slower when they complete a linguistic task, such as distinguishing between syllables, alongside picture-naming than when they complete a non-linguistic task (Fairs et al., 2018, Fargier & Laganaro, 2016), suggesting that the linguistic task interfered with word retrieval. Thus, pre-schoolers are behaving in a similar way to dual-tasking adults, indicating that, overall, their production processes are less efficient than adults' and so are affected more heavily by interference (be it from previous words or from dual-tasking).

To conclude, our results suggest that the planning processes involved in naming multiple objects is similar in pre-schoolers and adults. Pre-schoolers, like adults, tended to look at each object as they were about to name it, suggesting that they were retrieving each word form incrementally. However, our results also suggest that children are less efficient at retrieving later word forms, and instead show similar patterns as adults under cognitive load. Furthermore, our results are the first to investigate how pre-schoolers inspect objects as they name them. We have shown that it is possible to use eye-tracking paradigms in developmental production research, and hope that this study will pave the way for future studies investigating the online mechanisms involved in the developing production system.

4. Study 3: A horse by any other name: Referential alignment as evidence for flexible perspective-taking in pre-schoolers' language use [23]

4.1. Introduction

People can conceptualize things in more than one way, so that in different situations they might think of the same entity as AN ANIMAL, A HORSE, THAT PONY, MY PET, or SCARY THING, and correspondingly refer to it in different ways that reflect their current conceptualization, or *perspective* (e.g., *an animal, a horse, that pony, my pet*, *monster*). It is well-established that adults are highly flexible in the perspectives that they can adopt, and that the ability to adopt others' favored perspectives is crucial for successful communication (H. H. Clark & Wilkes-Gibbs, 1986). But how early does this ability develop, and what mechanisms subserve it? Previous research suggested that children's abilities to adapt to conversational partners' perspectives develop relatively late (Krauss & Glucksberg, 1969),

---

[23] Experiment 1 in this study was designed by Holly Branigan and carried out by undergraduate students who submitted this work as part of their undergraduate dissertation for a degree in Psychology at the University of Edinburgh. Experiments 2-3 were designed and carried out by the author who submitted this work as part of her Masters dissertation for a Masters degree in the Psychology of Language at the University of Edinburgh. However, since the dissertation was submitted, the data has been reanalysed using Bayesian methods and the addition of Bayes Factors, and we have also included combined analyses. Experiment 5 was designed by Holly Branigan and carried out by undergraduate students (who submitted this work as part of their undergraduate dissertation for a degree in Psychology at the University of Edinburgh) and by the author. The author coded, analysed and wrote up the manuscript.
This chapter is based on a submitted manuscript to in Developmental Psychology (Lindsay, L., Hopkins, Z. L., Branigan, H.B. (submitted). A horse by any other name: Referential alignment as evidence for flexible perspective-taking in pre-schoolers' language use. *Developmental Psychology.*).

but more recent evidence has suggested that even pre-schoolers can accommodate others' perspectives under some circumstances (Matthews, Lieven, & Tomasello, 2010). Here we report five experiments that examine whether pre-schoolers spontaneously adopt in their own language production the same perspective as a conversational partner previously used, and the mechanisms that might support this behavior.

Adult speakers flexibly adopt different perspectives when conversing, as reflected in their referential choices. For example, they refer to the same referent using a basic-level term (*dog*) or a more specific subordinate (*labrador*) depending on whether other category exemplars are present (Brennan & Clark, 1996), and using a proper name (*UN building)* or a description (*the building with all the flags)* depending on their partner's presumed knowledge (Brennan & Clark, 1996; Isaacs & Clark, 1987). Critically for our interest, they are also responsive to their conversational partners' perspectives, so that they show a robust tendency to spontaneously adopt the perspective, and hence referring expression, that a partner has previously used (e.g., using *pony* after their partner has previously used *pony*). That is, adult speakers tend to *referentially align* (or entrain) with a partner's perspective, and so maintain the *referential precedent* that a partner has set (and adult comprehenders correspondingly expect speakers to maintain precedents once established; Metzing & Brennan, 2003; Shintel & Keysar, 2007). This tendency occurs both for novel objects without conventional names (e.g., *floor* vs. *row* for a configuration of points*;* Garrod & Anderson, 1987) and familiar objects with conventional names (e.g., *basket* vs. *hamper;* Branigan, Pickering, Pearson, McLean, & Brown, 2011). Importantly, speakers align with a partner's perspective even when this entails adopting a perspective that they would not usually favor (e.g., using *pony* where they would normally strongly favor using *horse;* Branigan et al., 2011; Tobar, Rabagliati, & Branigan, 2020). Such alignment is associated with successful and satisfying communication (Fusaroli et al., 2012).

Adults' referential alignment has been attributed to a range of underlying mechanisms (which importantly may not be mutually exclusive; Branigan et al., 2011a). First, it may be driven by speakers' motivation to be understood, so that speakers consider what their addressee is likely to understand when deciding how to conceptualize and so refer to an object (i.e., *audience design*; Clark, 1992, 1996). These judgments are shaped by assumptions about *common ground* (information assumed to be shared by the speaker and addressee), based on the speaker's prior interactions with the addressee (e.g., Clark & Marshall, 1981; Galati & Brennan, 2010; Wilkes-Gibbs & Clark, 1992), as well as beliefs about their likely knowledge (Fussell & Krauss, 1992; Isaacs & Clark, 1987). This collaborative account implicates theory of mind (*ToM*; Wimmer & Perner, 1983): Speakers must model their partners' mental states, including their favored perspective on an object, and accommodate these in their referential production. When someone refers to an animal as a *pony*, it signals that she understands and favors this perspective, causing her partner to adopt the same perspective during subsequent reference and so creating a partner-specific referential pact that commits both interlocutors to using that expression in future, even when overinformative in context (Brennan & Clark, 1996).

In contrast, an alternative account suggests that alignment can arise implicitly, without modelling a partner's mental states, via domain-general episodic memory representations (Horton & Gerrig, 2005; Horton, 2007; Horton & Gerrig, 2016). Under this account, normal memory processes lead people to store and associate referential expressions with different objects, partners and contexts. When a speaker subsequently talks about the same object with the same partner, both the object and partner act as memory cues that activate previously used referential expressions (and associated perspectives), pre-empting the use of other terms.

Alignment might also occur without partner modelling via automatic priming mechanisms. According to the interactive alignment account, when people converse they prime each other's linguistic representations in a way that facilitates subsequent re-use (Pickering & Garrod, 2004); importantly, such facilitation is not partner-specific (Branigan, Pickering, McLean, & Cleland, 2007). Alignment at one level of representation (e.g., lexis) promotes aligned representations at other levels (e.g., syntax), leading ultimately to aligned semantic representations, so that alignment of linguistic representations is causally implicated in the development of shared perspectives and mutual understanding between interlocutors. That is, using the same language can lead conversational partners to implicitly adopt the same perspective without explicitly modelling the other's perspective.

Finally, speakers might align in order to promote positive social relations. Social psychological accounts of behavioral mimicry suggest that people automatically imitate aspects of a partner's behavior for social-affiliative reasons. Studies in this area have focused on motor imitation (e.g., posture or body movements; Chartrand & Bargh, 1999; Lakin & Chartrand, 2003), but some evidence suggests that language imitation in adults also supports affiliative goals (Abrahams, Hartsuiker, De Fruyt, & Bajo, 2019; van Baaren, Holland, Steenaert, & van Knippenberg, 2003; see also Giles, Coupland, & Coupland, 1991).

In contrast to the overwhelming evidence that adults are able to spontaneously, flexibly and responsively adopt different perspectives on objects during dialogue, young children's ability to recognize and accommodate multiple perspectives in their language use remains contested. Early claims for egocentricity and inflexibility in young children's referential communication (e.g., Deutsch & Pechmann, 1982; Dickson, 1982; Krauss & Glucksberg, 1969; Whitehurst & Sonnenschein, 1981) are supported by findings that young children reject alternative perspectives - and hence alternative names - for objects during comprehension. For instance, three-to-five year olds show a mutual exclusivity bias during word-learning, so that they assume that new words refer to new objects (Markman & Wachtel, 1988; Markman, 1989). Furthermore, three-year-olds deem a partner's use of an object name to be unacceptable if this contravenes their own preferences (Doherty & Perner, 1998; Doherty, 2000; Perner, Stummer, Sprung & Doherty, 2002).

But more recent evidence suggests that young children are in fact able to recognize and take different perspectives on objects during comprehension. Matthews, Lieven, and Tomasello (2010) showed that three- and five-year-olds are flexibly responsive to different conversational partners' use of different perspectives within the same pragmatic context. In their study, children moved objects with two possible names (e.g., *horse* vs. *pony*) in response to instructions. Some children heard the same experimenter give instructions in two blocks, using the same name in both blocks (*horse-horse;* i.e., maintaining the same perspective) or switching names between blocks (*horse-pony;* i.e., changing to a new perspective); other children heard different experimenters in each block. Both three- and five-year-old children were able to switch perspectives across blocks. However, they were slower to do so in the different name/same experimenter condition than in the different name/different experimenter condition, suggesting that they associated particular perspectives with particular partners. Four-year-olds demonstrate a similar ability to switch perspectives (and a similar association between particular perspectives and particular partners) when comprehending modified noun phrases (e.g., *striped ball* vs. *purple ball;* Graham, Sedivy, & Khu, 2014). Moreover, they can rapidly shift between multiple speakers' perspectives during comprehension (Khu, Chambers, & Graham, 2020).

In language production, preschoolers' capacity to flexibly take alternative perspectives on objects appears more limited. Three-year-olds can recognize and use different perspectives when given pragmatically appropriate contexts and/or cues that explicitly elicit multiple perspectives (Clark, 1997; Deák & Maratsos, 1998; Mervis, Golinkoff, & Bertrand, 1994); for example, calling a dinosaur-shaped crayon both *a dinosaur* and *a crayon* when probed to use alternative perspectives (Deák, Yen & Pettit, 2001). However, the ability to spontaneously switch between perspectives (i.e., to abandon an established precedent and adopt a new perspective) seems to emerge later in development. In Deák et al.'s study, three-year-olds were less likely to adopt a particular perspective on an object after they already taken another perspective on it than when they had not. Moreover, although both four- and six-year-old speakers spontaneously switch from using a proper noun (e.g., *Peter,* an individual-level perspective) to a common noun (e.g., *boy,* a category-level perspective) in response to changes in the pragmatic context, only six-year-olds similarly switch from a subordinate-level perspective (*woman's shoe*, in the context of more than one shoe) to a category-level perspective (*shoe*, in the context of a single shoe*;* Koymen, Schmerse, Lieven, & Tomasello, 2014; see also Birch & Bloom, 2002; Diesendruck, 2005).

Existing evidence does not discriminate underlying mechanisms. Most findings have been explained in terms of children's ability to explicitly model a partner's perspective, but are compatible with other explanations. For example, preschoolers' rejection of others' names for objects might reflect impaired ability to model others' perspectives (Doherty & Perner, 1998; Doherty, 2000; Perner, Stummer, Sprung & Doherty, 2002), but could equally reflect the cognitive demands of the experimental task (i.e., the need to maintain two perspectives simultaneously; Clark, 1997). Similarly, three- and five-year-olds slowed comprehension when speakers fail to maintain referential precedents (Graham et al., 2014; Matthews et al., 2010) might arise because they explicitly model and accommodate partner-specific common ground, but is also compatible with automatic memory-based associations between particular contexts, partners, objects, and expressions (Horton & Gerrig, 2005).

Indeed, children in Matthews et al.'s study were slowed whenever they encountered an alternative perspective, irrespective of the speaker, suggesting a basic advantage for previously processed words that is compatible with lexical priming (Ostashchenko, Deliens, Geelhand, Bertels & Kissine, 2019).

Importantly, studies to date have focused on how pre-school speakers choose perspectives when they initiate reference to target objects. No studies have investigated how they choose perspectives in response to a partner's previous referential expressions, and specifically whether they spontaneously adopt the same perspective as a conversational partner - in other words, whether they referentially align. This gap in the literature is perhaps surprising, given claims about the pervasiveness of alignment and its role in successful communication, and substantial evidence for robust referential alignment in typically and atypically developing school-aged children as well as in adults. For instance, typically developing seven-to-twelve-year-olds navigating a maze tend to re-use referential expressions that their partners have just used (e.g., *line* vs *row;* Garrod & Clark, 1993), and both typically developing and autistic eight-to-ten-year-olds tend to name pictures in a picture-matching game using the same perspective - even when usually disfavored - as an adult experimenter previously used (Branigan, Tosi, & Gillespie-Smith, 2016; Hopkins & Branigan, 2020; Hopkins, Yuill, & Branigan, 2017).

As with adults, school-aged children's alignment is compatible with more than one underlying mechanism. It could in principle reflect explicit modelling of a partner's perspective, consistent with evidence that children are able to engage in some forms of audience design from a young age (e.g., Hansson, Nettelbladt, & Nilholm, 2000; Hoff, 2010; Sachs & Devin, 1976; Shatz & Gelman, 1973), with more sophisticated forms developing over time (Koymen et al., 2014). However, the fact that autistic children show the same tendency to align as typically developing children, and that this tendency is independent of individual differences in ToM abilities, implicates other mechanisms that are not contingent on partner-modelling, under at least some circumstances (Branigan et al. 2016; Hopkins et al. 2017). One candidate might be automatic memory associations (Schneider, 2002), though this possibility is not supported by Branigan et al's. (2016) finding that alignment was equally strong whether children named the same or a different referent as the experimenter (e.g., the same or a different token of a horse, where the former should act as a stronger memory cue).

An alternative possibility would be priming mechanisms, whereby a partner's use of a name activates the corresponding word form and concept in the mental lexicon. This explanation is consistent with evidence that lexical priming effects emerge early in development (Friedrich & Friederici, 2005; Gershkoff-Stowe, Connell, & Smith, 2006; Mani, Durrant, & Floccia, 2012), and with patterns of alignment effects found in cd typically developing and autistic children (Branigan et al. 2016; Hopkins et al. 2017). It is also in keeping with Garrod and Clark's (2003) finding that seven-to-eight-year-olds tended to re-use their partners' words even with different reference, suggesting that the word form was sometimes facilitated independently of the mapping to a referent in the discourse model.

Finally, referential alignment in older children might have a partial source in social-affiliative mechanisms. Hopkins and Branigan (2020) found that typically developing seven-to-ten-year-olds' showed a heightened tendency to referentially align with the experimenter in a picture-matching game following an ostracism manipulation (Cyberball; Williams & Jarvis, 2006) , compared with controls who were not ostracized. These findings are consistent with evidence that young children's non-linguistic imitation may be modulated by affiliative motivations (Over & Carpenter, 2009b; Watson-Jones, Legare, Whitehouse, & Clegg, 2014; Watson-Jones, Whitehouse, & Legare, 2016).

4.1.1. The Current Study

We suggest that we can gain insight into children's cognitive and communicative development by investigating the age at which children adopt the same perspective on objects as their conversational partners, the circumstances under which such referential alignment occurs, and the mechanisms that underlie this behavior. It is informative about young children's cognitive flexibility in conceptualizing the world, and the developmental origins of a behavior that is pervasive in adult communication. Most importantly, it can cast light on important yet unresolved questions about the extent to which successful communication crucially depends upon the ability to model others' mental states (Nilsen & Graham, 2009, 2012) versus automatic and potentially resource-free mechanisms that children may develop early, and which might support the emergence of more complex communication skills.

Across five experiments, we therefore investigated whether three-to-four-year-olds spontaneously and flexibly adopt a conversational partner's perspective in their own language production in response to their partner's language use, and examined the mechanisms underlying such effects. In Experiments 1-4, we used a 'snap' task in which a child and experimenter alternated turning over cards and naming pictures of familiar objects in a card-matching game (see Branigan, McLean & Jones, 2005). On experimental trials, the experimenter named an object compatible with two perspectives (reflected in two alternative acceptable names; e.g., *horse/pony*) using either the favored or disfavored perspective (as established by a pre-test). Following two filler trials, the child named the same target object. In Experiment 5, we used a novel collaborative 'Moving house' task in which a child and experimenter took turns giving each other a series of instructions about where to place objects in a house. The experimenter gave instructions involving target objects compatible with two perspectives, using the favored or disfavored perspective; in the next block, the child gave instructions involving the same objects. In both tasks, our dependent measure was children's likelihood of adopting the disfavored perspective when referring to an object (i.e., aligning on the same name) after hearing the experimenter use the disfavored vs. favored perspective; alignment manifested as an increased likelihood of using the disfavored perspective after the experimenter had used the disfavored perspective than the favored perspective.

In Experiment 1, we investigated whether children spontaneously adopted the same perspective used by their partner even if this was a normally disfavored perspective. In Experiment 2, we examined whether children aligned on a perspective when referring to a different token from the same category. In Experiment 3, we tested whether children aligned on a perspective even after they had explicitly established their own, different, perspective for the referent earlier in the interaction. In Experiment 4, we investigated the partner-specificity of these effects, by having one experimenter present when the child established their favored perspective and a different experimenter present during the picture-matching game. Finally, in Experiment 5, we examined whether alignment occurred in less structured, more cognitively demanding contexts than the 'snap' task (i.e., the 'Moving house' task). In all experiments, we additionally examined whether children would be more likely to align after exposure to third-party ostracism. Half of the children in each experiment played the picture-matching game after watching videos that depicted third-party ostracism; the other half played after watching control videos (Over & Carpenter, 2009a).

Across the experiments, patterns of children's use of the disfavored perspective would inform our understanding of three-to-four-year-olds' ability to spontaneously and flexibly adopt different perspectives during communication, and the extent to which these effects might arise from partner-modelling (audience design), memory-based, lexical priming, or social-affiliative mechanisms.

4.2. Experiment 1

In Experiment 1, we used the ‘snap’ picture-matching game to investigate whether three-to-four-year-olds adopted the same perspective as a conversational partner had previously used when referring to the same object. On experimental trials, the experimenter named objects that were compatible with two perspectives (one strongly favored and one strongly disfavored), and the child named the same referent two turns later. If children were more likely to use the disfavored perspective after the experimenter used this perspective than after she used the favored perspective, it would suggest that three-to-four-year-olds are able not only to recognize multiple perspectives during communication, but moreover to flexibly and spontaneously adopt a disfavored perspective (and by implication, give up a favored perspective) during language production; and furthermore that they do so in response to their partner’s language use. Any such tendency would be compatible with partner-modelling (audience design), episodic memory, or lexical priming mechanisms.

We additionally examined whether children’s alignment was influenced by factors that might affect social-affiliative motivation (Hopkins & Branigan, 2020). Before playing the ‘snap’ game, participants watched either two videos depicting third-party ostracism or two control videos (Over & Carpenter, 2009). If children were more likely to align with their partner’s use of an otherwise disfavored perspective after exposure to ostracism, it would suggest that the tendency to adopt a partner’s perspective is underpinned in part by social-affiliative mechanism.

4.2.1. Method

Participants. Twenty-four native English-speaking children (female = 12; mean age = 47.6 months; range = 36-56 months) from nurseries in and around Edinburgh participated. None of the children had any known language or developmental disorders. Ethnicity and SES were not recorded but were representative of the area (almost entirely White, predominantly from middle-class Scottish families). We were unable to use traditional power analyses to determine a sample size for our experiment due to the lack of prior studies examining the influence of both a partner's previous language use and third-party ostracism on pre-schoolers' lexical alignment. We therefore aimed to test a similar sample size as Over and Carpenter (2009) for our affiliative manipulation, and a similar sample size to other lexical alignment studies in older children (Branigan et al., 2016a). To mitigate for our lack of a power analysis, we also calculated Bayes Factors (BF, Dienes, 2014; see data analysis section below). Ethical approval for this experiment and all other experiments was obtained from the Psychology Research Ethics Committee at the University of Edinburgh.

Design and Materials. We used a 2 (Partner Perspective: favored vs. disfavored; within-subjects) x 2 (Social-affiliative Condition: ostracism vs. control; between-subjects) mixed design.

For our ostracism manipulation, we used the videos used successfully in Over and Carpenter's (2009) study of five-year-olds. In the ostracism condition videos, a group of coloured shapes interacted with each other, but repeatedly moved away from a single shape that repeatedly approached the group; in the control condition videos, a group of coloured shapes interacted with each other while a differently shaped object made random movements around the screen.

In the ‘snap’ game, we used a subset of the experimental items used by Branigan *et al*. (2016a) that had been pre-tested to establish that children in this population knew and understood both perspectives but had a strong preference for one alternative (i.e., the favored perspective).[24] The twelve experimental items in the snap game comprised pairs of identical objects (a prime and target) that had two common alternative perspectives (*horse* vs. *pony*), and a scripted prime name (favored vs. disfavored partner perspective; henceforth *favored prime* vs *disfavored prime*). The prime and target pictures used the same tokens of the same object (see Figure 17).

We prepared two paired (experimenter/participant) lists, each containing one version of the experimental items in a Latin Square design, as well as 24 filler pictures. Item order was randomized for each participant with the following constraints: two filler items appeared between the experimenter’s prime card and the child’s associated target card; and four snap trials (where the experimenter and participant’s consecutive cards matched) were distributed among trials.

[24] Pre-tests established that 1) a group of 18 typically developing three- and four-year-olds used both perspectives when naming the target (e.g., horse and pony), but the favored perspective was used at least twice as frequently as the disfavored perspective; and 2) more than 70% of a further group of 12 typically developing three- and four-year-olds correctly chose the target (from an array of four including a semantic competitor) when given the name corresponding to the disfavored perspective (see Branigan et al., 2016a). The alternative names were not drawn systematically from any particular dialect or register.

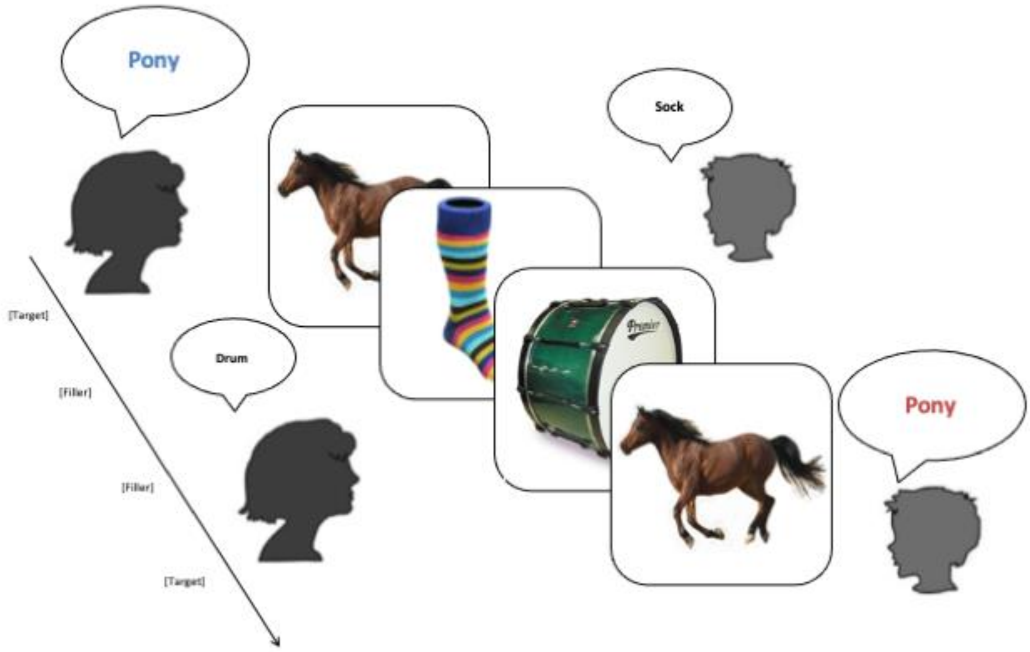

Figure 18. Example trial (Disfavored partner-perspective condition). The experimenter named the prime card using the Disfavored perspective. After two filler trials, the child named their target card.

Procedure. Testing was conducted by an experimenter (E1) who was blind to the social-affiliation condition to which each child had been assigned. E1 invited children individually into a quiet room (see Figure 18 for a timeline of who interacted with the child during each stage). After sitting down, E1 introduced a second experimenter (E2) to the child. E1 told the child that she (E1) needed to read something outside but that the child could watch videos with E2. When the videos finished, E1 returned and explained how to play ‘snap’ to the child.

E1 placed a pile of pre-arranged picture cards face-down in front of each player (the child and E1). E1 told the child that they would each take turns naming the picture and looking for ‘snap’ items to win. The experimenter and child alternated turning over and naming the top card. The experimenter always began the game by turning over the top card and naming it (following a script) – this constituted the prime name. After two intervening fillers (one named by the child, one by the experimenter), the child named his or her target card – this constituted the target response. The game continued with players alternating until all the cards had been described. When both players had the same picture on their top card, the first player to shout *snap!* won the cards in play.

After completing the ‘snap’ game, E1 and the child watched a final video in which all the shapes played together. This video was designed to alleviate any negative feelings induced by the ostracism videos. Children were then verbally asked what they thought the first videos were about.[25] Experimental sessions were audio-recorded for transcription and coding.

[25] The majority of children did not provide any response.

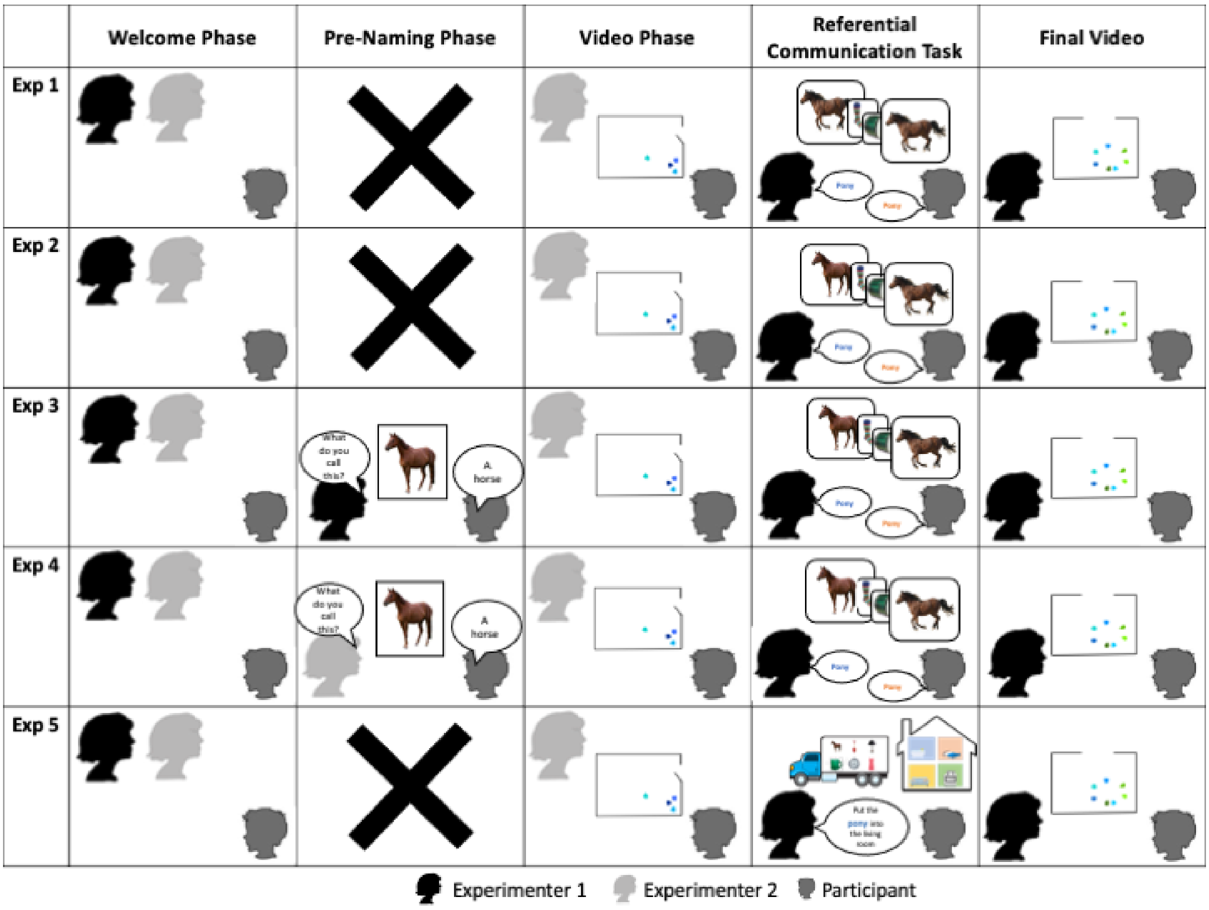

Figure 19. Experiment structure for all experiments, showing which experimenter(s) interacted with the child (participant) during each stage. In the welcome phase, E1 introduced E2 to the child. In the pre-naming phase (Experiments 3 and 4), E1 elicited the child's preferred perspective in Experiment 3; E2 elicited the child's preferred perspective in Experiment 4. In the video phase (all experiments), E1 left while E2 and the child watched videos together. In the referential communication task, E2 left while E1 returned to play with the child. In the final video phase, E1 and the child watched a video together.

Data Analysis. Target responses were coded as favored name (e.g., *horse*), disfavored name (e.g., *pony*), other, or no response. 'No responses' were excluded from our analysis. As our analysis was concerned with establishing whether children were more likely to use a disfavored perspective after hearing a disfavored prime name, we coded responses where a disfavored perspective was used to describe an experimental item as 1. All other responses were coded as 0.

We used Bayesian generalized linear mixed-effects models[26] (GLMMs; Baayen, Davidson & Bates, 2008), using the *bglmer* function of the *blme* package (Chung, Rabe-Hesketh, Dorie, Gelman & Liu J 2013) – an extension of the *lme4* package (Bates, Mæchler, Bolker & Walker, 2015) - in RStudio (version 1.2.5033), with a binomial link function, and participants and items as random effects. We used the maximal random effects structure justified by our design (Barr, Levy, Scheepers & Tily, 2013). In cases where the model would not converge, correlations among random effects were fixed to zero (Bates, Kliegl, Vasishth, & Baayen, 2015). Unless otherwise specified, all predictors were contrast coded (-0.5, 0.5) and centered. The following fixed effects were included in the model: Social-affiliation Condition (reference level: control vs. ostracism), Partner Perspective (reference level: favored vs. disfavored) and a two-way interaction (Social-affiliative Condition * Partner Perspective). For the GLMM models, we report coefficient estimates (*B*), standard errors (*SE*) and *z* and *p* values for each predictor; 95% confidence intervals (CI) are from the *confint* function (method=Wald).

We also calculated Bayes Factors by using the models' Bayesian Information Criteria (BIC) values. The use of Bayes Factors has been argued to offer advantages over traditional power analyses (Dienes, 2014) and also allowed us to quantify the strength of evidence for the alternative hypothesis versus the null hypothesis. Following Dienes (2014), we interpret a Bayes Factor of greater than 3 as strong evidence for the alternative hypothesis over the null,

[26] We were unable to use a normal generalised linear mixed effects model (*glmer*) as even a model with a simple random effects structure produced a singular fit error. Singular fit errors can appropriately be addressed using Bayesian approaches (e.g., using the *blme* package) that adds a weak prior and avoids singularity (Chung et al., 2013).

less than 1/3 as strong evidence for the null hypothesis over the alternative, and anything between these values as support for neither the alternative nor the null. We estimated a Bayes Factor as $e^{(BIC\_null - BIC\_experimental)/2}$ from the BIC values of both the experimental and null models (Wagenmakers, 2007). The raw data and analysis scripts are available on request.

4.2.2. Results

Table 18 reports the frequency of different responses for each condition for each experiment; Table 19 reports the proportion of favored and disfavored responses for each condition across each experiment. The alignment effect for favored responses is the difference between the percentage of favored responses in the favored vs. disfavored partner-perspective conditions; the alignment effect for disfavored responses is the difference between the percentage of disfavored responses in the disfavored vs. favored partner-perspective conditions.

Across conditions, children aligned with the experimenter on 74% of trials (favored partner-perspective condition: 84%; disfavored partner-perspective condition: 64%). Importantly, children were more likely to use a disfavored perspective after the experimenter had used a disfavored perspective than a favored perspective (64% vs. 6%; Figure 19A; *B* = 4.71, *SE*= .82, *z*= 5.67, *CI*=[3.09, 6.32], *p*<.001, BF= 167934). This tendency to align with the experimenter's perspective did not vary as a function of Social-affiliative condition (*B* = 1.50, *SE*=1.44 , *z*= 1.04, *CI*=[-1.33, 4.34], *p*=. 30, BF= .12). Unexpectedly, children used the disfavored perspective more often after watching the control videos than the ostracism videos; note however that the corresponding Bayes Factor supports neither the alternative nor the null hypothesis (control: 42% trials; ostracism: 29% trials; *B* = -1.68, *SE*= .78, *z*= -2.18, *CI*=[-3.18, -.17], *p*=<.05, BF = 1.31).

Table 18. Frequency of Each Response Type by Social Affiliative Condition and Partner Perspective for each Experiment

Experiment	Social-Affiliative Condition	Response	Partner Perspective: Favored	Partner Perspective: Disfavored
1	Control	Favored	57	17
		Disfavored	8	52
		Other	6	3
		No Response	1	0
	Ostracism	Favored	63	20
		Disfavored	1	40
		Other	8	11
		No Response	0	1
2	Control	Favored	56	17
		Disfavored	5	40
		Other	8	13
		No Response	3	2
	Ostracism	Favored	54	16
		Disfavored	8	47
		Other	4	7
		No Response	6	2
3	Control	Favored	64	45
		Disfavored	2	21
		Other	5	5
		No Response	1	0
	Ostracism	Favored	64	41
		Disfavored	4	29
		Other	4	1
		No Response	0	1
4	Control	Favored	61	30
		Disfavored	3	28
		Other	7	9
		No Response	1	1
	Ostracism	Favored	60	45
		Disfavored	1	16
		Other	4	6
		No Response	0	1
5	Control	Favored	159	107
		Disfavored	14	62
		Other	14	19
		No Response	4	5
	Ostracism	Favored	156	96
		Disfavored	11	63
		Other	9	20
		No Response	1	0

Table 19. Proportion of Favored and Disfavored Responses by Social-Affiliative Condition and Partner Perspective for Each Experiment.

Experiment	Social-affiliative condition	Response	Partner Perspective: Favored	Partner Perspective: Disfavored	Alignment Effect
1	Control	Favored	.80 (.05)	.24 (.04)	56%
		Disfavored	.11 (.03)	.72 (.06)	61%
	Ostracism	Favored	.88 (.03)	.28 (.05)	60%
		Disfavored	.01 (.01)	.57 (.07	57%
2	Control	Favored	.80 (.06)	.18 (.05)	62%
		Disfavored	.06 (.04)	.56 (.08)	50%
	Ostracism	Favored	.83 (.05)	.23 (.05)	60%
		Disfavored	.11 (.04)	.68 (.07)	57%
3	Control	Favored	.90 (.03)	.63 (.06)	27%
		Disfavored	.03 (.02)	.30 (.06)	27%
	Ostracism	Favored	.89 (.03)	.58 (.09)	31%
		Disfavored	.06 (.02)	.41 (.09)	35%
4	Control	Favored	.86 (.03)	.45 (.05)	41%
		Disfavored	.04 (.02)	.41 (.06)	37%
	Ostracism	Favored	.92 (04)	.67 (.08)	25%
		Disfavored	.01 (.01)	.24 (.06)	23%
5	Control	Favored	.82 (.03)	.58 (.04)	24%
		Disfavored	.11 {.03)	.32 (.03)	21%
	Ostracism	Favored	.81 (.05)	.53 (.04)	28%
		Disfavored	.14 (.05)	.36 (.03)	22%

Note. Proportions calculated over all responses (Excluding No Responses). Means over subjects (SE). The alignment effect is the difference between the proportion of favored responses following favored vs disfavored partner perspectives/ proportion of disfavored responses following disfavored vs favored partner perspectives.

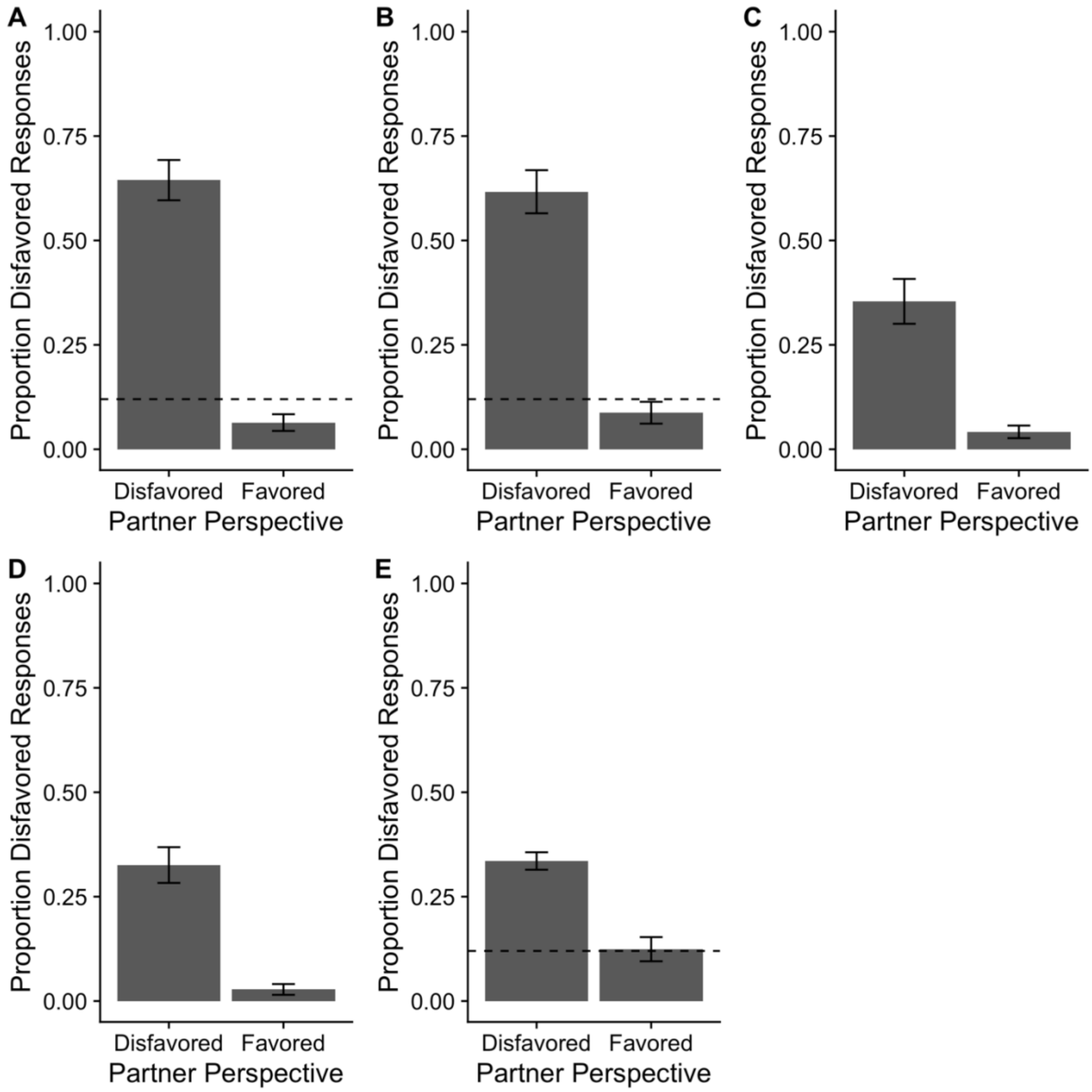

Figure 20. The mean proportion of disfavored responses (e.g., pony) as a function of Partner Perspective for all Experiment 1 (A), Experiment 2 (B), Experiment 3 (C), Experiment 4 (D) And Experiment 5 (E). Error bars indicate ±1 standard error. Dashed horizontal line corresponds to the mean proportion of disfavoured responses pre-schoolers spontaneously produced in the pre-test.

4.2.3. Discussion

Experiment 1 showed that three-to-four-year-old speakers spontaneously adopted the same perspective as their conversational partner had previously used (i.e., followed their partner's referential precedent): Children tended to use a favored or disfavored perspective when naming an object depending on whether the experimenter had used a favored or disfavored perspective for that object two turns earlier. This effect was very strong, with children using the same perspective as their partner on three-quarters of trials overall, and – strikingly – on two-thirds of trials where their partner had used a perspective that was usually strongly disfavored. Children's tendency to adopt the same perspective as their partner was not influenced by whether they had been exposed to videos depicting third party ostracism versus control videos (although there was some suggestion that they were unexpectedly more likely overall to adopt a disfavored perspective after control videos). As we shall see, there were no significant effects of this manipulation in any of our other experiments.

These results suggest that three-to-four-year-olds are flexible in the perspectives that they can recognize and adopt during language use. They are compatible with audience design, memory-based or priming-based mechanisms, but not social-affiliative mechanisms. That is, children might have used the same perspective as the experimenter had previously used because they modelled her mental states and believed that she preferred (and could understand) that perspective on a specific referent (or perhaps on a category of referents); or because the context, experimenter, and target picture automatically acted as memory cues triggering the recall of the previously used perspective; or because the experimenter's original expression activated relevant lexical representations that retained activation and influenced subsequent language production.

In Experiment 2, we set out to replicate Experiment 1's finding that children spontaneously adopted a partner's otherwise disfavored perspective, and to begin to discriminate the possible underlying mechanisms of this behavior. We therefore examined whether children aligned with a partner's perspective when referring to a different referent from the same category as the original referent (i.e., different token of a horse). That is, the experimenter is naming a picture of a horse as "pony," and children are naming a different picture of a horse (whereas in experiment 1, they were naming the same horse). If so, it would suggest that children could spontaneously adopt their partner's perspective on a category of object, and not simply on a specific referent. If this tendency was weaker than in Experiment 1, it would be compatible with a memory-based mechanism in which alignment was triggered by exposure to memory cues (Experiment 1: context, partner, category, and token; Experiment 2: context, partner, and category), under the assumption that repetition of a token should be a stronger cue than a token from the same category. If this tendency was however as strong as in Experiment 1, it would be compatible with a priming mechanism in which exposure to a linguistic stimulus would facilitate subsequent re-use for any appropriate referent. It could also be compatible with partner-modelling, under the assumption that children would infer a partner's favored perspective at the level of a category rather than a token.

4.3. Experiment 2

4.3.1. Method

Participants. A power analysis based on data from Experiment 1 showed that a sample of 24 participants was sufficient to achieve more than 80% power to detect the Partner Perspective effect found in Experiment 1. Twenty-four children (female = 11; mean age = 46.8 months; range = 36-59 months) from the Edinburgh area participated in Experiment 2. One further child was excluded, as they were unable to complete the task. Children were tested in the Developmental Lab at the University of Edinburgh. As in Experiment 1, we did not record SES information although participants were predominantly White from middle-class families.

Design and Materials. The design and materials were identical to Experiment 1, except that the prime picture in the experimenter's card set was a different token from the same category as the target picture in the participant's card set (target pictures were identical to Experiment 1; see Figure 20).

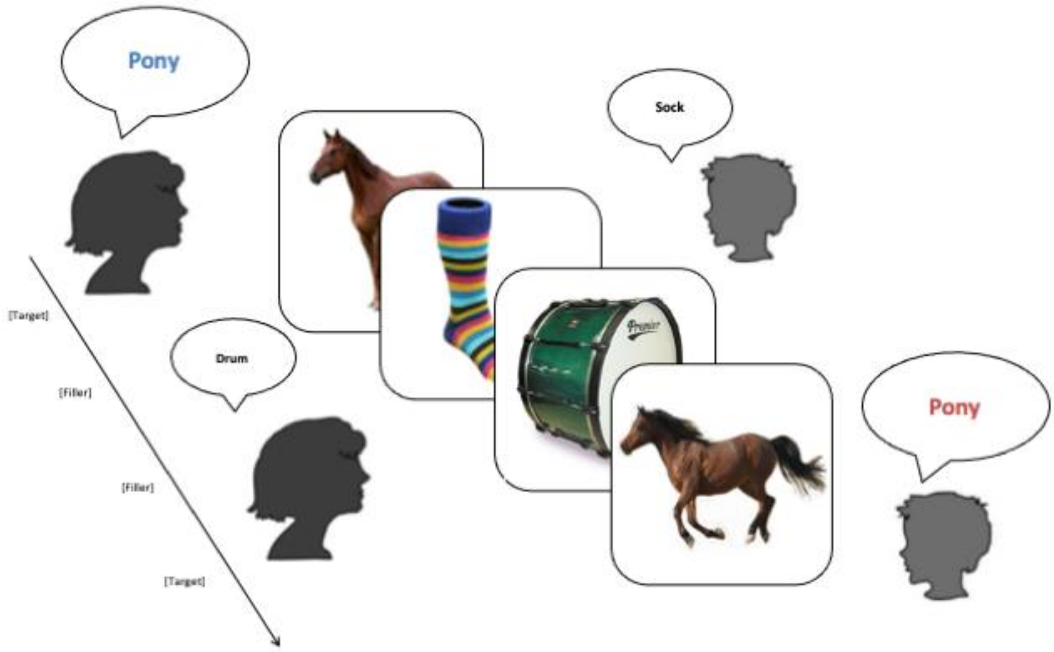

Figure 21. Example trial in Experiment 2 (Disfavored Partner-perspective condition). The prime and target picture show different tokens of the same category.

Procedure. The procedure was identical to Experiment 1.

Data Analysis. The data were coded and analyzed in the same way as Experiment 1. However, social-affiliative condition had no effect in any of our models, and for this and future analyses, we report a model with only Partner perspective as a fixed effect. We used the maximal random effects structure justified by our design (Barr, Levy, Scheepers & Tily, 2013; see Supplementary Materials for a full model including Social-affiliative condition and the two-way interaction between Partner perspective and Social-affiliative condition).

4.3.2. Results

Across conditions, children aligned with the experimenter on 72% of trials (favored partner-perspective: 81%; disfavored partner-perspective: 62%). Replicating Experiment 1, children were more likely to use a disfavored perspective after the experimenter had used a disfavored perspective than a favored perspective (62% vs. 9%; $B = 4.57$, $SE = 1.07$, $z = 4.28$, $CI = [2.47, 6.66]$, $p < .001$, BF = 4485.47). Thus, children tended to adopt the same perspective as their partner on a *category* of referents.

4.3.2.1. Combined Analysis of Experiments 1 and 2

To determine whether naming different tokens of the same category influenced the extent of children's alignment, we combined the data from Experiment 1 and Experiment 2, and included Partner perspective, Token type (same [Experiment 1] vs. different [Experiment 2]), and a two-way interaction (Partner perspective:Token type) as fixed effects in our model.

Participants were more likely to use a disfavored perspective after the experimenter used a disfavored perspective than a favored perspective ($B = 4.15$, $SE = .60$, $z = 6.95$, $CI = [2.98, 5.31]$, $p < .001$, BF = 68054.4). There was no main effect of Token type ($B = .11$, $SE = .54$, $z = .21$, $CI = [-.95, 1.17]$, $p = .84$, BF = .043) and no significant interaction between Partner perspective and Token type ($B = -.17$, $SE = 1.06$, $z = -.16$, $CI = [-2.24, 1.91]$, $p = .88$, BF = .042). Thus, children tended to adopt the same perspective as their partner on a category to the same extent as they did on an individual referent.

4.3.3. Discussion

Experiment 2 showed similar results to Experiment 1: Three-to-four-year-old speakers tended to spontaneously adopt the same perspective as their conversational partner had previously used, even when a usually disfavored perspective. Moreover, this tendency occurred when they named a referent from the same category as their partner's referent. Hence children tended to adopt the same perspective with respect to a category of referents rather than individual referents. Our combined analysis showed that this tendency was equivalent in magnitude to that found when children referred to the same referent as their partner (alignment effect: Experiment 1: 58% vs. Experiment 2: 53%).

These results are compatible with both audience design and priming mechanisms. They are less compatible with memory-based mechanisms, which would predict stronger alignment in the presence of more cues to retrieving the relevant memory (i.e., the previously used precedent), and hence stronger alignment in Experiment 1, where the referent was the same in the original episode and the retrieval episode, than in Experiment 2, where the category was the same but the referents were different.

4.4. Experiment 3

Experiments 1 and 2 demonstrated that 3-4 year old children's referential choices are flexible and sensitive to the perspectives expressed in their partners' language use. They also provide some evidence against an underlying memory-based mechanism for these effects. However, although they show that children can spontaneously adopt a partner's perspective even when this differs from the default perspective that most children would adopt, they do not show directly that children can switch flexibly in their language production from an established perspective to a new perspective.

In Experiment 3, we therefore tested whether children aligned with their partner's perspective on a category of referents even after they had earlier explicitly established and adopted a different perspective on the target referent. Hence, Experiment 3 had the same structure as Experiment 2, but children additionally named the target objects prior to playing the 'snap' game, thus setting a referential precedent. The pre-game naming task also allowed us to unequivocally determine each individual child's favored perspective for each object (rather than using the population preference established in pre-tests), and to use this individual preference to inform our Partner-perspective manipulation. More specifically, we recorded each child's favored perspective during the pre-game naming task, and adjusted the experimenter's script during the picture-matching game accordingly.

If children tended to align with their partner's perspective under these circumstances, it would provide strong evidence that they were able to spontaneously switch from an established, favored perspective to a partner's (otherwise disfavored) perspective in response to their partner's language use. In other words, it would show that they were able to abandon their own referential precedent and successfully adopt an alternative perspective.

4.4.1. Method

Participants. Twenty-four children (female = 11; mean age = 46.5 months; range = 36-56 months; predominantly White from middle-class families, although this was not recorded) participated. Four children were further excluded as they were either bilingual (N=3) or failed to complete the task (N=1). Children were tested in the Developmental Lab at the University of Edinburgh.

Design and Materials. The design and materials used were identical to Experiment 2, except for the following differences. First, we created a picture book for the naming task. This contained 24 pictures that would subsequently appear in the 'snap' game (12 target pictures from the experimenter's card set and 12 filler pictures from the participant's card set). Second, the experimenter's scripts specified each experimental item for Partner-perspective condition (favored vs. disfavored perspective), but did not identify the specific perspective to be used, as this was individually determined by the perspective used by the child in the pre-game naming task. Following the naming task, the script was completed with the appropriate perspective (e.g., if the child used the *horse* perspective in the pre-game naming task, the script was specified as *horse* in the favored Partner-perspective condition, and as *pony* in the disfavored Partner-perspective condition; conversely, if the child used the *pony* perspective in the pre-game naming task, the script was specified as *pony* in the favored Partner-perspective condition, and as *horse* in the disfavored Partner-perspective condition).

Procedure. The procedure was identical to Experiment 2, except for the following differences. At the beginning of the session, children named pictures from the picture book with E1 while E2 noted the perspective that the child used for each object. E1 used these notes to complete the script appropriately (i.e., using the child's perspective for items in the favored Partner-perspective condition, and using the alternative perspective for items in the disfavored Partner-perspective condition) while absent from the room, while E2 and the child watched the social-affiliative videos together.

Data Analysis. The data were coded and analyzed as in Experiment 2. See Supplementary Materials for a full model including social-affiliative condition and the two-way interaction between Partner perspective and Social-affiliative condition.

4.4.2. Results

In the pre-naming task, children used the favored name [i.e., favored according to the original pretest] 60% of the time and the disfavored name [i.e., disfavored according to the original pretest] 25% of the time. For the other 15% of cases, children used a different word (e.g., they used *television* rather than the favored *TV* or disfavored *Telly*). These results validate our pre-test, showing that children were more than twice as likely to use a word identified as favored in the original pretest over a word identified as disfavored in the original pretest.

Across conditions, children aligned with the experimenter on 63% of trials (favored partner perspective: 90%; disfavored partner perspective: 35%). Children were more likely to use a disfavored perspective in the picture-matching game after the experimenter used a disfavored perspective than a favored perspective, even though doing so required them to abandon a favored perspective that they had previously established with the same partner (35% vs. 4%; $B = 2.49$, $SE = .52$, $z = 4.78$, $CI = [1.38, 5.45]$, $p < .001$, BF = 721.26).

4.4.3. Discussion

As in Experiments 1 and 2, three-to-four-year-old speakers tended to spontaneously adopt the same perspective on objects as their conversational partner had previously used, and as in Experiment 2, this tendency involved a perspective on a category of objects, rather than on individual referents. Moreover, they did so even when this required them to override a referential precedent that they had already established for that referent (although to a lesser extent than when they had not previously established such a referential precedent: 35% Disfavored perspective alignment effect, compared to 64% in Experiment 1 and 62% in Experiment 2). Hence Experiment 3 demonstrates not only that three-to-four-year-olds are able to adopt an otherwise disfavored perspective, but also that they are able to abandon a previously adopted, favored perspective in order to do so. In other words, they are able to actively switch between alternative perspectives in their language production.

This pattern of results is compatible with alignment arising from priming mechanisms (in which children were primed for both perspectives - for the disfavored perspective by the experimenter's use of that perspective in the 'snap' task, and for the favored perspective by their own use of that perspective in the pre-game naming task) and with alignment arising from audience design mechanisms (in which children took the experimenter's use of the disfavored perspective in the 'snap' task, as well as her acceptance of the favored perspective that they themselves used in the pre-game naming task - i.e., their referential precedent - as evidence for her likely understanding of both perspectives).

4.5. Experiment 4

In Experiment 4, we aimed to replicate the finding that children would spontaneously abandon a previously adopted (favored) perspective to adopt their partner's (disfavored) perspective. We also aimed to discriminate possible underlying mechanisms by manipulating the identity of the partner with whom the child established the initial precedent versus the partner who subsequently introduced an alternative perspective (and with whom the child could align). In Experiment 3, one experimenter (E1) acted as the child's partner in both phases. In Experiment 4, in contrast, one experimenter (E1) acted as partner in the pre-game naming task, and a second experimenter (E2) - not present during the naming task - acted as partner in the 'snap' game. Thus as in Experiment 3, children named the target referents before playing the 'snap' game - but unlike in Experiment 3, they did not thereby establish a referential precedent with their partner for the 'snap' game.

If children's alignment to their partner's perspective arises from priming mechanisms, we should see a similar pattern as in Experiment 3: Children should show competition between alternative perspectives activated by previous use (their own favored perspective in the pre-game naming task, and the experimenter's [disfavored] perspective in the 'snap' game), irrespective of whether the partner had been present when they used their favored perspective. That is, we should find no difference in alignment between Experiments 3 and 4. However, if alignment is affected by audience design mechanisms, then children should be more likely to align with their partner in Experiment 4 than Experiment 3: In Experiment 3, children had previous evidence that their partner could understand their favored perspective (i.e., there was a referential precedent), whereas in Experiment 4, children have no previous evidence about which perspectives their partner can entertain (i.e., there is no referential precedent), and should therefore rely on the perspective that she has just used within the 'snap' game as evidence about her likely understanding.

4.5.1. Method

Participants. Twenty-four children (female = 9; mean age = 43.3 months; range = 36-57 months; predominantly white from middle-class families, although this was not recorded) participated. One child was further excluded as they failed to complete the task (N=1). Children were tested in the Developmental Lab at the University of Edinburgh.

Design and Materials. The design and materials used were identical to Experiment 3.

Procedure. The procedure was the same as Experiment 3, except that children named the pictures and watched the assigned videos with E2 while E1 was absent from the room. After the videos, E2 left the room and the participant played the snap game with E1.

Data Analysis. The data were coded and analyzed as in previous experiments. See Supplementary Materials for a full model including Social-affiliative condition and the two-way interaction between Partner-perspective and Social-affiliative condition.

4.5.2. Results

Across conditions, children aligned with the experimenter on 61% of trials (favored partner perspective: 89%; disfavored partner perspective: 33%). Children were more likely to use a disfavored perspective after the experimenter used a disfavored perspective than a favored perspective, even though this required them to abandon a favored perspective that they had previously established with a different partner (33% vs 3%; *B* = 21.58, *SE*= 6.83, *z*= 3.16, *CI*=[8.21, 34.96], *p*<.01, BF = 59.59)[27].

4.5.2.1. Combined Analysis of Experiments 3 and 4.

To examine the effect of interacting with an old vs. new partner on children's likelihood to align, we combined the data from Experiments 3 and 4. We conducted a BGLMER model that included Partner perspective and Partner (Old Partner [Experiment 3] vs. New Partner [Experiment 4]) as fixed effects.

Children were more likely to use a disfavored perspective after the experimenter used a disfavored perspective than a favored perspective (*B* =3.74, *SE*= .86, *z*=-4.35, *CI*=[2.06, 5.43], *p*<.001, BF = 8042.73). However, we found no main effect of Partner (*B* = .51, *SE*= .66, *z*= .78, *CI*=[-.78, 1.81], *p*=.44, BF=.06): Children were no more likely to use a disfavored perspective (and so abandon an established favored perspective) when interacting with a new partner than an old partner. Nor was there a significant interaction between Partner Perspective and Partner (*B* = -.76, *SE*= 1.33, *z*= -.58, *CI*=[-3.36, 1.84], *p*=.57, BF=.05): Children's tendency to adopt their partner's disfavored perspective was not affected by whether they were interacting with an old partner (with whom they had previously established the favored perspective as a referential precedent) or a new partner (with whom they had not previously established the favored perspective as a referential precedent).

4.5.2.2. Combined analysis of Experiments 1-4

Experiments 1-4 provided strong evidence that three-to-four-year-olds tended to spontaneously adopt the same perspective on objects as their conversational partner. The

[27] This model had random effect correlations of -1; a model with forced zero correlations showed the same overall result.

pattern of effects in these experiments was compatible with an underlying priming mechanism, but less compatible with memory-based, audience design, or social-affiliative mechanisms. In particular, none of the experiments showed any suggestion that exposure to third-party ostracism influenced children's likelihood to adopt their partner's perspective, despite evidence for ostracism-induced enhanced alignment in the 'snap' game in older children (Hopkins & Branigan, 2020). However, that study suggested that such effects might be relatively small. To increase our power to detect any such effects, and so to provide further confidence in our conclusions, we therefore carried out a further analysis on the combined data from all four experiments. We included Partner Perspective, Social Affiliative condition and the two-way interaction between these as fixed effects in our model. Overall, participants were more likely to use a disfavored perspective after the experimenter used a disfavored perspective than a favored perspective (B =3.29, SE= .39, z=8.34, CI=[2.52, 4.06], p<.001, BF = 4940.72). However, there was still no effect of social-affiliative condition (B =-.25, SE= .32, z=-.78, CI=[-.88, .38], p=.43, BF = .0001) and no interaction (B =.16, SE= .64, z=-.25, CI=[-1.09, 1.41], p=.80, BF = .0001).

4.5.3. Discussion

Experiment 4 further confirmed that three-to-four-year-old speakers tended to spontaneously align with a conversational partner's (otherwise disfavored) perspective on a category, and that they did so even after having previously established a favored perspective for the target referent. Furthermore, it showed that this tendency occurred to the same extent with a new partner (with whom they had not established the precedent for the favored perspective; alignment effect 30%) as with an old partner (with whom they had established the precedent for the favored perspective; alignment effect 31%; Experiment 3). This pattern is compatible with an underlying priming mechanism that facilitated previously encountered expressions (and associated perspectives). It does not support an audience design mechanism under which children would choose a perspective to maximize their partner's likely comprehension, based on evidence from their partner's previous behavior, i.e., acceptance of the favored perspective during the pre-game naming task vs. use of the disfavored perspective during the 'snap' game. (We note that it also provides further evidence against a memory-based mechanism, which would predict weaker alignment in Experiment 3 than in Experiment 4: In Experiment 3, the experimenter's presence during the pre-game naming task should have led to her being associated with the favored perspective that the child used in that task, as well as with the disfavored perspective that she herself used during the 'snap' game', thereby weakening her effectiveness as a memory cue for the disfavored perspective.)

4.6. Experiment 5

Experiments 1-4 demonstrated spontaneous, responsive, and flexible perspective-taking in three-to-four-year-old speakers' interaction with a conversational partner in a simple picture-matching game. These results demonstrate that pre-schoolers are able to spontaneously adopt disfavored perspectives (and in doing so, to abandon previously established favored perspectives) under at least some circumstances. They also provide some evidence that priming mechanisms may play an important underpinning role in this behavior. But it is not clear to what extent such effects might generalize beyond the highly structured and minimal context presented by the 'snap' game to more open-ended, complex and naturalistic settings.

Various aspects of the 'snap' game might in principle constrain the conclusions that we can draw from our results. First, the nature of the game may have promoted the possible influence of lexical priming mechanisms and made it less likely that audience design and social-affiliative mechanisms would play a significant role. The 'snap' game did not require mutual understanding for task success – naming the pictures was incidental to successful performance, since the game simply required players to identify matching pictures (which could be done on the basis of the visual stimuli alone). Moreover, the game was competitive, in the sense that the goal was to be first to identify matching pictures. Therefore, there was no strong motivation for children to ensure that their partner understood their utterances. Relatedly, social-affiliative mechanisms might have been less important in this ‘competitive’ context than in a cooperative task where affiliative goals are enhanced (Chartrand & Bargh, 1999). (Of course, the ‘snap’ game was cooperative to the extent that both players had to mutually respect the rules of the game, observe appropriate turn-taking etc.)

Additionally, the ‘snap’ game does not reflect the demands of typical conversations. Although the task is interactive and involves two players, it is highly structured and cognitively undemanding, requiring children to produce single word utterances in response to individual objects (i.e., naming), as opposed to more typical language use such as producing complex grammatical structures that convey a proposition and can bring about a partner’s action on the world. Furthermore, the minimal linguistic and non-linguistic context would also serve to maximize the potential influence of lexical priming. Clearly, to be confident that young children robustly engage in spontaneous and flexible perspective-taking in response to a partner’s language use, it is important to demonstrate that our effects generalize to more ecologically valid contexts (see Verga & Kotz, 2018).

Experiment 5 addressed these issues by examining children’s propensity to spontaneously adopt a partner’s perspective in a novel referential communication paradigm – the ‘moving house’ game – in which the child and experimenter took turns to give each other a series of instructions to move objects from a removal truck to locations in a house/garden. In three pairs of rounds, the experimenter told the child where to place a series of six objects from the truck into their house/garden (prime rounds); the child and experimenter then switched roles, and the child directed the experimenter where to place a further series of six objects into another house/garden (target rounds). Experimental items appeared in both prime and target rounds and were objects compatible with more than one perspective (e.g., *horse/pony*). As in Experiment 2, we tested whether children adopted the same perspective as their partner had previously used for a different token from the same category (and when children had not previously established a precedent).

Importantly, the game was explicitly cooperative and task success relied upon mutual understanding: Success depended on both players placing their items correctly on the basis of the instructions that they received from their partner, so that children had to work together with the experimenter and ensure that the experimenter correctly understood to which item they were referring. As such, the game involved both social-affiliative and communicative goals. Furthermore, the task was considerably more cognitively demanding and less structured than the 'snap' game, thus better mimicking real conversation. Children had to identify the correct object and intended location (from a picture showing the desired end state) and formulate an appropriate instruction that would induce their partner to put the correct object in the correct location.

The 'moving house' game therefore allowed us to investigate the extent to which our findings might generalize to more naturalistic settings, and moreover to contexts in which audience design mechanisms and social-affiliative mechanisms might be more likely to be relevant. In this respect, we might expect to find a particularly strong tendency for children to adopt their partner's perspective if audience design and social-affiliative mechanisms contribute to alignment. However, Experiments 1-4 suggested that priming mechanisms may play an important role in alignment. We might expect such mechanisms to be less effective in this setting than in the 'snap' game, because both an extended time delay and a series of cognitively demanding actions intervened between initial exposure to the partner's (disfavored) perspective and the child's opportunity to adopt that perspective. This potential reduction in priming-induced alignment might offset any increased tendency to align based on audience design and social-affiliative mechanisms. However, this reduction would occur equally across conditions. We would therefore still be able to detect any social-affiliative modulation of children's tendency to adopt a partner's perspective, in an increased likelihood to align after exposure to third party ostracism compared to the control condition.

4.6.1. Method

Participants. We increased our sample size for this experiment for two reasons: First, previous research suggests that the effects of an ostracism manipulation on children's lexical alignment are small (Hopkins & Branigan, 2020); second, characteristics of the 'moving house' paradigm (e.g., cognitive complexity of the task, extended interval between the experimenter's prime and the child's target) might weaken the effects of partner perspective compared to the 'snap' game. We were however unable to run a traditional power analysis due to a lack of data using this new paradigm and so instead aimed to test as many children as possible. As with our previous experiments, we used Bayes Factors to quantify the strength of evidence for an alternative vs. the null hypothesis. Participants were sixty-two children (female = 30; mean age = 47.4 months; range = 36-60 months); predominantly White and from middle-class families, although this was not recorded. Thirteen further children were excluded because they were bilingual (N=12) or failed to complete the task (N=1). Children were tested in nurseries in the Edinburgh area or the Developmental Lab at the University of Edinburgh.

Design and Materials. Experimental items were taken from the 'snap' game; 4[28] items that were inappropriate in a 'removal truck' context were replaced with other items from Branigan *et al*. (2016a). As in the 'snap' game, the 12 experimental items comprised two printed pictures (a prime and a target) that allowed two alternative perspectives (*horse* vs.

---

[28] *belly*, *bobble* and *dots* were replaced with *stairs, rabbit* and *ship*

pony), and a scripted prime name (favored vs. disfavored). As in Experiments 2-4, the prime and target pictures involved different tokens from the same category. We prepared two paired lists, each containing one version of the experimental items in a Latin Square design, as well as six filler pictures to appear in prime rounds and six filler pictures to appear in target rounds (each used only once in the experiment).

We also prepared eight A3-sized printed house scenes (four for prime rounds and four for target rounds, with one from each set being used for practice) and a printed removal truck. Each house included four different rooms as well as other locations in the garden (e.g., a shed). For target rounds, we created A4-sized booklets that showed a miniature version of the experimenter's house and depicted the intended location of each item, so that the child could tell the experimenter where to put items from the removal truck. Each page of the booklet highlighted one item to be moved.

Each experimental prime item was associated with a prime-round house and a target-round house; item order was fixed for each house, with each experimental item appearing in the same ordinal position in the prime and target rounds. However, we counterbalanced the order in which the pairs of houses were presented. Each experimental round consisted of four experimental items and two filler items. In total, there were three prime rounds and three target rounds.

Procedure. As in all previous experiments, testing was conducted by an experimenter (E1) who was blind to the social-affiliative condition to which the child had been assigned. After children watched the videos with E2, E1 entered the room and explained how to play the 'moving house' game and completed the practice round with the child. E1 showed the child their practice house and gave them the removal truck with the items arranged on it. E1 first told the child where each item should go into the house (e.g. *Put the dog in the bedroom*), after which the child moved the item from the removal truck into the house (see Figure 21). After all the items had been moved, the experimenter and child switched roles. The child then told E1 where each item should be moved into the house, using the booklet as a guide. E2 stayed in the room to help the child move items from the moving truck to the house.

After E1 was satisfied that the child understood how to play the game, they began the experimental rounds. During a prime round, the experimenter referred to half of the experimental items using the favored perspective, and the other half using the disfavored perspective. After all items had been moved, the roles were reversed, and the target round began. Each session lasted between 15 and 20 minutes.

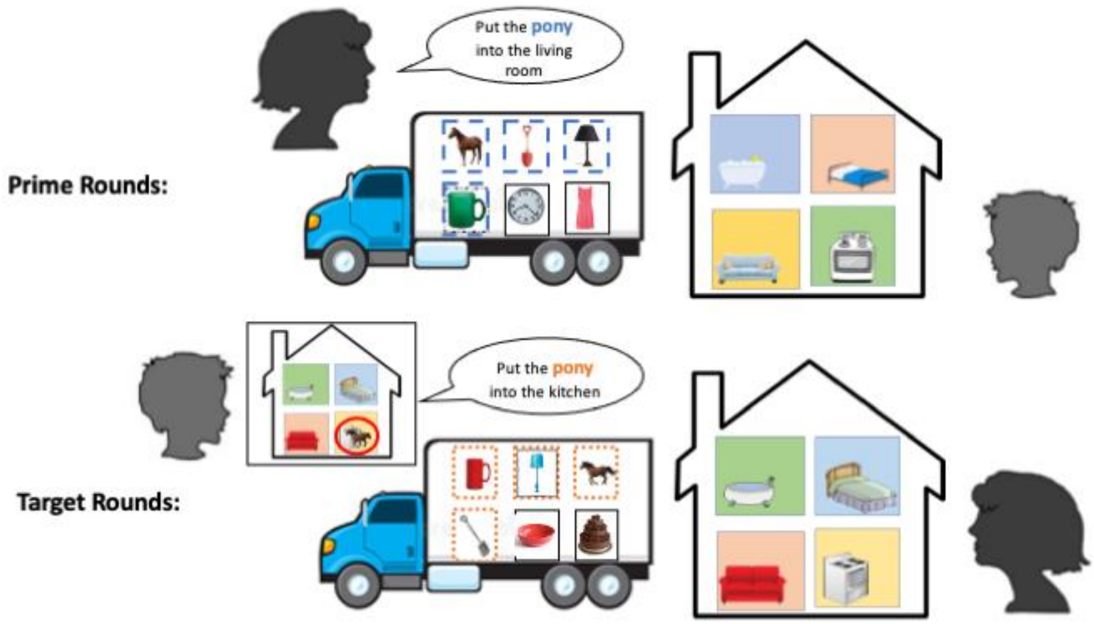

Figure 22. Example trial from the Moving House game (Disfavored Perspective condition). During a prime round, the experimenter told the child where to move each of six items from the truck to the house. The experimenter referred to half of the experimental items (dashed boxes) using the favored perspective, and the other half using the disfavored perspective.

Once all items in the truck had been moved, the experimenter and child reversed roles and the target round began.

Data Analysis. The data were coded and analyzed as in Experiments 1-4. As the 'moving house' game took much longer than the 'snap' game to complete, we also included Round Number in our model to control for the possibility that affiliative motivation might increase with increasing interaction (Foltz, Gaspers, Meyer, Thiele, Cimiano & Stenneken, 2015). Thus, we report a full model that includes Partner perspective, Social-Affiliative condition, Round and the two-way interaction between Partner perspective and Social-affiliative condition as fixed effects. We used the maximal random effects structure justified by our design (Barr, Levy, Scheepers & Tily, 2013).

4.6.2. Results

Overall, participants aligned with their partner on 59% of trials (favored partner perspective: 81%; disfavored partner perspective: 34%). Children were more likely to use a disfavored perspective when instructing the experimenter where to place objects after the experimenter had used a disfavored perspective for another token from the same category in the previous round than after the experimenter had used a favored perspective (33% vs. 12%; B=3.09, SE=.74, z=4.19, CI=[1.64, 4.54], p<.001, Bayes Factor = 17835.42). There was no effect of Social-affiliative condition (B = .60, SE=.49 , z= 1.21, CI=[-.40, 1.57], p=. 23, Bayes Factor = .08) or Round Number on children's overall likelihood of using the disfavored perspective (B = .03, SE=.07, z= .38, CI=[-.11, .17], p=.70, Bayes Factor = .0000005). Furthermore, the tendency to align with the experimenter's perspective was not modulated by Social-affiliative condition (B = -.75, SE=.90 , z= -.83, CI=[-2.51, 1.02], p=.41, Bayes Factor = .05).

4.6.3. Discussion

Experiment 5 showed that three-to-four-year-olds' tendency to adopt a conversational partner's (otherwise disfavored) perspective on a category generalized to a context involving a cooperative game in which children instructed their partner where to place a series of objects. As in previous experiments, this tendency was spontaneous and responsive to their partner's language use. These results suggest that referential alignment is not restricted to simple and highly structured tasks in which referential communication is unnecessary for task success and conversational partners have a competitive goal: In the Moving House task, children had to determine the correct location for each object, and cooperatively communicate both the identity of the relevant object and its location to the experimenter, in order to achieve joint success. Moreover, alignment occurred over a substantial number of intervening turns and associated time delay.

The magnitude of alignment was substantial (21%), but smaller than under similar conditions in the 'snap' game (Experiment 2: 53%), and was not modulated by exposure to ostracism. These results do not determine the underlying mechanisms of the alignment effect, but suggest that a context that promotes mutual understanding and social affiliation does not necessarily increase alignment, which in turn suggests that audience design and social-affiliative contributions to alignment may be relatively weak in young children. Relatedly, the reduced magnitude of alignment is compatible with underlying priming mechanisms, in which activation of primed representations decays over time and with intervening language use (Nakamura, Ohta, Ozaki & Matsushima, 2006; Wheeldon & Monsell, 1992).

4.7. General Discussion

Flexible perspective-taking is a hallmark of adult language use that is critical to successful communication, exemplified in referential alignment, or adults' robust tendency to adopt the same perspective on objects as their conversational partners. We investigated whether young children are similarly able to engage in flexible perspective-taking during language use. Using an established picture-naming game (Experiments 1-4) and a novel 'moving house' game (Experiment 5), we tested whether three-to-four-year-olds tended to adopt the same perspective as a conversational partner, and explored the mechanisms underlying such behavior.

Across five experiments, three-to-four-year-olds tended to use the same perspective when referring to an object as their partner had used in a previous turn, even when this was a perspective they would not normally use. Such referential alignment occurred spontaneously, and over a range of objects. In Experiment 1, children were more likely to adopt a disfavored perspective when naming an object after hearing their partner use a disfavored perspective for the same referent. In Experiment 2, children adopted a partner's disfavored perspective when naming a different referent from the same category. Experiment 3 showed that children adopted a partner's disfavored perspective even after having established their own precedent (i.e., favored perspective) for the referent with their partner; and Experiment 4 showed that this effect was not partner-specific. Experiment 5 showed that children's tendency to use a partner's disfavored perspective endured (at least) six conversational turns. Taken together, these results extend our understanding of young children's perspective-taking in communication.

4.7.1. Perspective-taking in preschoolers' referential production

First, our findings are informative about the occurrence and extent of flexible perspective-taking in this age group. Previous research offers mixed evidence about whether three-to-four-year-olds can recognize and adopt multiple perspectives on the same referent, with some studies suggesting that they cannot (Doherty & Perner, 1998; Doherty, 2000; Perner, Stummer, Sprung & Doherty, 2002; Markman & Wachtel, 1988; Markman, 1989), and others suggesting that they can given appropriate pragmatic contexts and/or explicit cues that make the alternative perspectives more salient (E. V. Clark, 1997; Deak, Yen, & Pettit, 2001; see also Nadig & Sedivy, 2005, in older children). Moreover, previous findings have suggested that young children's ability to flexibly switch away from a previously adopted perspective to another perspective is limited and slow to develop, and that this limitation is more marked in production than comprehension (Graham et al., 2014; Koymen et al., 2014; Matthews et al., 2010).

Our results show that three-to-four-year-olds are able to recognize and adopt multiple perspectives on the same referent, and use these perspectives in their own production. Thus they provide evidence for preschoolers' cognitive flexibility in both conceptualization and communication. Critically, our results suggest that this behavior is not contingent on different pragmatic contexts (e.g., in response to different contrast sets; Clark, 1997) or explicit elicitation of alternative perspectives (e.g., questions cueing function versus shape; Deák et al., 2001). Given the same referent and the same referential domain (i.e., set of potential referents), children in our study spontaneously adopted different perspectives in response to the perspective that their partner had recently used. Furthermore, they did so not only for the same referent, but also for a different referent from the same category, suggesting that their ability to adopt a partner's perspective was not narrowly constrained. Thus our work provides novel evidence that preschoolers' referential choices are responsively sensitive to a conversational partner's perspectives and associated referential expressions in the same way as adults and school-aged children (Branigan et al., 2016; Brennan & Clark, 1996; Garrod & Anderson, 1987; Garrod & Clark, 1993; Hopkins & Branigan, 2020).

Our work also suggests that such responsive communicative perspective-taking occurs robustly and pervasively in this age group. Experiments 1 and 2 showed that children used the experimenter's perspective on around three quarters of trials. Notably, they did so on around two thirds of trials where the experimenter used a perspective that was strongly disfavored within the population (i.e., that children would normally use less than around a third of the time, based on the pre-test). Even when children pre-established their individual preferences (Experiments 3 and 4), they continued to adopt the experimenter's disfavored perspectives on around a third of trials.

Even more strikingly, these experiments show that three-to-four-year-olds are able not only to inhibit a prepotent perspective, but also to spontaneously disengage from a favored and established perspective in order to adopt a disfavored perspective. In Experiments 1 and 2, the favored perspective was prepotent and available when children prepared to name the target object, and must therefore have been inhibited when children selected the disfavored perspective (following the experimenter's use). In Experiments 3 and 4, children selected their favored perspective during the pre-game naming task, and they must therefore have inhibited and de-selected this established perspective in order to select the disfavored perspective (following the experimenter's use) during the snap game. This finding is interesting given the suggestion that underdeveloped executive control functions – specifically, the ability to inhibit a favored perspective – might underlie young children's perspective-taking difficulties in referential communication tasks (Nilsen & Graham, 2009). Our results suggest that, under certain conditions, children are able to successfully inhibit underlying (default) perspectival preferences on objects. We return to this point below.

Our findings show that children's tendency to adopt the same perspective as their conversational partner is broad-based. First, it occurs across a range of referents within an interaction (in our experiments, across the 12 varied objects that constituted our experimental items). Second, it occurs in varied communicative contexts. We observed referential alignment in a highly structured task (i.e., the 'snap' game), where children named individual objects two turns after the experimenter had done so (Experiments 1-4). We also observed referential alignment in a more complex, cognitively demanding task (the Moving House game) where children had to plan and produce a series of sentences involving two referents (the object to be moved and its intended location) several conversational turns after the experimenter had done so (and after undertaking a series of actions in response to the experimenter's utterances; Experiment 5). Under these conditions, children still tended to adopt the experimenter's disfavored perspective on around a third of trials. This suggests that the effects we observed in Experiments 1-4 were not simply an artefact of a highly structured task (see also Hopkins, Yuill, & Keller, 2016).

4.7.2. Mechanisms underlying referential alignment in young children's communication

As well as providing novel evidence that pre-schoolers spontaneously adopt the same perspective as their partners during language use, our results cast light on the mechanisms that might underlie this behavior. Most research on young children's communicative perspective-taking has been framed in terms of models of conversational partners' knowledge states and in particular the knowledge that they believe themselves to share (i.e., audience design and common ground). But as we noted, existing data are compatible with alternative interpretations, and research on adults' communicative perspective-taking – and specifically referential alignment – has proposed alternative mechanisms that might contribute individually or collectively to speakers' tendency to adopt their partners' perspectives.

Although our results do not explicitly quantify the contribution of specific mechanisms to perspective-taking in children's communication, they are suggestive about which mechanisms might have been most relevant to referential alignment in the settings described. First, our results do not provide any evidence that memory-based mechanisms contributed to alignment effects, by which children would tend to recall previously used perspectives when cued by contexts, objects, and partners. Children were no more likely to adopt the same perspective as their partner when a highly relevant cue was present (the same token as the partner had originally named; Experiment 1) than when that cue was absent (a different token; Experiment 2). Nor were they more likely to re-use their own established perspective when a highly relevant cue was present (the same partner as was present when the perspective was initially established; Experiment 3) than when that cue was absent (a different partner; Experiment 4).

Second, our results do not suggest children's tendency to adopt their partner's perspective in these experiments was informed by audience design. The results of Experiments 1 and 2 are compatible with audience design mechanisms, under which children might have adopted their partner's perspective because they inferred that this was the partner's favored perspective for the referent (Experiment 1) or its category (Experiment 2), and so used this perspective to ensure successful communication. However, the results of Experiments 3-5 are less compatible with this interpretation.

In Experiment 3, children had previously established a referential precedent with their partner, whereas in Experiment 4 they had not. Under an audience design account, children should be sensitive to the existence of a referential precedent, and should tend to maintain this precedent where possible (in keeping with the evidence that their partner accepted this perspective, and also their expectation that people should refer to objects in a consistent way; Matthews et al., 2010). They should therefore be more likely to (re-)use their own favored perspective (and less likely to adopt their partner's perspective) in Experiment 3 than in

Experiment 4, where they had not previously established a referential precedent with their partner and therefore should have relied on the perspective that she had just used as evidence about her likely understanding. However, children aligned with their partner to the same extent irrespective of whether or not they had previously established a (different) precedent with them.[29] Finally, Experiment 5 involved a cooperative context in which mutual understanding was critical to task success (unlike Experiments 1-4), and so audience design mechanisms might be expected to promote an increased tendency to adopt their partner's perspective. But in fact, children referentially aligned to a substantially lesser extent than in Experiments 1 and 2, and to approximately the same extent as in Experiments 3 and 4 (in which alignment involved abandoning an established precedent, and so might be expected to be attenuated compared with a situation in which it did not involve abandoning an established precedent as in Experiment 5). Taken together, our results do not provide any evidence that children modelled their partner's likely knowledge and selected a perspective accordingly.

Additionally, our findings did not suggest that social-affiliative mechanisms contributed to children's tendency to adopt their partner's perspective. Evidence from individual experiments, and from the combined analysis of Experiments 1-4, shows that children were no more likely to referentially align after exposure to videos depicting third-party ostracism than after control videos. While it is possible that our affiliation manipulation was ineffective, we note that it has successfully induced enhanced non-linguistic imitation in

[29] Note that these results are also inconsistent with an alternative possibility, in which the partner's failure to follow the child's established precedent in Experiment 3 might have been taken by children as evidence that the partner found this perspective problematic and thus increased their tendency to adopt their partner's (new) perspective in order to promote mutual understanding. In that case, children should have been more likely to adopt their partner's perspective in Experiment 3 than Experiment 4.

other studies in young children (Over and Carpenter, 2009b; Song, Over, & Carpenter, 2015; see also Watson-Jones et al., 2014 for a similar, successful manipulation). Furthermore, evidence that a first-person ostracism manipulation induces heightened referential alignment in older children (Hopkins & Branigan, 2020) suggests that a more direct experience of ostracism might be required to enhance linguistic imitation relative to non-linguistic imitation.

Overall, our pattern of results is compatible with children's choice of perspective being influenced by underlying psycholinguistic priming mechanisms that facilitate particular perspectives via activation within the lexicon, without making reference to others' mental states. Under this account, hearing *pony* would automatically activate the wordform *pony* in the child's lexical system, in turn activating the lemma pony, and thereby the concept PONY (Levelt, Roelofs, & Meyer, 1999), giving rise to the relevant perspective. Following use, all of these representations would then retain activation (which would decay over time), facilitating their subsequent re-use. When the child subsequently had to name their target picture, these representations would be easier to access, increasing the likelihood that they would be selected. Crucially, note that children only did so when these representations (and the underlying perspective that they expressed) were congruent with the target picture: Children never used their partner's perspective inappropriately, e.g., referring to a non-equine referent as *horse.* In this case, when children initiated the production process by selecting a perspective to encode, the normally disfavored perspective PONY would be pre-activated and so more likely to be selected when naming the target picture than the normally favored perspective HORSE. The associated lemma and wordform representations would also be pre-activated, facilitating their processing during subsequent stages of grammatical and phonological encoding (Levelt et al., 1999).

This interpretation is supported by the finding that children were as likely to adopt their partner's perspective whether they were referring to the same referent as their partner (Experiment 1) or a different token of the same category (Experiment 2): In both cases, the activated perspective was congruent with the target. It is also supported by the finding that children were less likely to align perspectives when they had previously established a (favored) perspective in the pre-game naming task, irrespective of the partner with whom they had established that perspective (Experiments 3 and 4). Through pre-naming, children activated representations associated with their favored perspective, which retained activation following initial use. When children had to name targets during the 'snap' game, these self-primed representations competed strongly for selection with representations primed by their partner's utterance, manifesting as a reduced tendency to align with the partner's perspective. Finally, a priming explanation is compatible with our finding of reduced alignment in Experiment 5, where both the intervening time interval and intervening language use would be expected to reduce residual activation of representations associated with the experimenter's utterance, thereby rendering them weaker competitors for selection compared with the favored perspective. (See also Ostashchenko et al., 2019, for converging evidence in comprehension).

4.7.3. Implications for young children's communication

Our findings suggest that, in specific dialogue contexts, preschoolers are able to adopt a common perspective on referents with their conversational partners via automatic and potentially resource-free priming mechanisms, rather than by explicitly modelling their partner's mental states. Such mechanisms might therefore underpin successful communication at an early age in ways that might support the subsequent development of more complex and cognitively demanding communication skills.

Yet this conclusion also implicates priming mechanisms as a potential source of *mis*communication in other contexts. In our experiments, children tended to use the same perspective as their partner when referring to the same referents or different tokens from the same category. But the referential domain never contained potential competitors (e.g., other referents to which a *horse* perspective might apply); hence children's referential alignment could not be detrimental to communication. However, the same might not be true of other contexts. Garrod and Clark (1993) showed that when pairs of younger (seven-to-eight-year-old) and older (nine-to-ten-year-old and eleven-to-twelve-year-old) children adopted common referring expressions in a maze-navigation task, the magnitude of this tendency correlated with communicative success in older children only. That is, although younger children frequently used the same words as their partners to describe the maze, they did not necessarily communicate about it successfully: They often aligned on referring expressions (e.g., *line*) without aligning on underlying reference (e.g., a horizontal arrangement of nodes). In these cases, the perspective cued by the primed expression did not correspond to the perspective originally taken by their partner. Seven-to-eight-year-olds, unlike older children, were poor at detecting and repairing such misunderstandings.

Such findings suggest a potential role for executive function (specifically inhibitory control) in children's communicative perspective-taking. Recent work has suggested that inhibitory control skills contribute to children's referential communication: Young children may fail to inhibit a prepotent contextually-induced favored (visual) perspective in order to adopt their partner's perspective (Nilsen & Graham, 2009). Our experiments showed that young children were able to inhibit their strongly favored (and hence prepotent) perspective after exposure to their partner's perspective. But Garrod and Clark's (1993) findings suggest that inhibitory control might play an another role in children's communication, by suppressing automatic but communicatively maladaptive alignment. Children may sometimes need to diverge from the representations primed by a partner's language, in order to (re)establish mutual understanding; failure to do so could create pragmatic challenges in more complex, communicatively demanding situations than those studied here. And the ability to effectively suppress a tendency to re-use a partner's language may in turn depend upon relevant audience design skills that make use of common ground and feedback to track partner's likely and actual understanding, and so identify when alignment is inappropriate, and where necessary initiate and pursue relevant repair strategies that may depend on explicit partner modelling. These skills appear to develop slowly (Anderson, Clark, & Mullin, 1991, 1994).

In sum, we have shown that three-to-four-year-olds are able to recognize and adopt multiple perspectives on objects, and that they do so spontaneously, flexibly and responsively during interaction with conversational partners. We further propose that young children may be able to achieve successful communication without engaging in cognitively demanding audience design, and instead through priming mechanisms that ultimately support the adoption of a partner's perspective and so automatically promote mutual understanding without the need for explicit partner modelling (Pickering & Garrod, 2004). As such, adult partners' language could help to scaffold children's communicative language use, by facilitating – and so strengthening – a range of linguistic representations and associated perspectives, which may in turn support children's developing knowledge of referential conventions within their speech community. However, priming mechanisms alone might not result in successful communication in more complex, challenging discourse contexts; the ability to accommodate a broad range of complex referential domains and respond appropriately to interactive feedback will depend on the development of other cognitive skills.

5. Study 4: Investigating Referential Maintenance Over Time and Contexts

5.1. Introduction

Any object or individual can be referred to in many different ways, reflecting different conceptions, or perspectives. For example, a given horse can be referred to as *the horse, the brown one, my favourite pony, Seabiscuit, that thing in the farm,* and so on. Each of these expressions correspond to a different perspective. To ensure successful communication, speakers must produce referring expressions (and utterances more generally) that are both unambiguous and informative. One way adults reduce the ambiguity of their referential expressions during conversation is by creating local conventions – or referential precedents – about how an object should be referred to (Brennan & Clark, 1996; Clark & Wilkes-Gibb, 1986; Garrod & Anderson, 1987; Metzing & Brennan, 2003). Once a referential precedent has been established, adult speakers continue to maintain it: that is, they often use and expect to hear their partner use this specific description in subsequent interactions, even if it is in a context where that description is overinformative (i.e., *lexical entrainment*; Brennan & Clark, 1996; Brown-Schmidt, 2009; Metzing & Brennan, 2003; Shintel & Keysar, 2007). Importantly, speakers continue to maintain referential precedents even when this entails adopting a perspective that they would not usually favour (e.g., using *pony* where they would normally strongly favour using *horse*; Branigan, Pickering, Pearson, McLean & Brown, 2011; Tobar, Rabagliati, & Branigan, 2020).

Adult's referential communication skills have often been interpreted as evidence of perspective-taking (e.g., Horton & Keysar, 1996). Indeed, according to mediated accounts of referential maintenance, such as the collaborative account, referential precedents are driven by our beliefs about our partner. When deciding how to describe an object, speakers will choose a term they believe their partner is most likely to understand by modelling their partner's knowledge against their own (Clark, 1996). These beliefs can be shaped by judgements about their addressee's likely knowledge (Fussell & Krauss, 1992; Isaacs & Clark, 1987) or by their prior interactions with that addressee (Clark & Marshall, 1981). For example, if a speaker has used a specific expression (e.g., *pony*) with a specific partner in previous interactions, that expression becomes part of the common ground - the mutual knowledge and beliefs – between conversation partners and influences how they produce or understand referring expressions for subsequent interactions. The speaker knows their partner will understand what that specific expression (*pony*) means and so will reuse it in subsequent interactions. Thus, according to this account, referential precedents constitute referential *pacts,* i.e., an implicit agreement between conversation partners to adopt a given perspective on an object (and thus use the same name) and its maintenance is understood to be cooperative (Brennan & Clark, 1996; Clark, 1996). As such, conversational partners would expect each other to follow and maintain precedents in a way consistent with Grice's (1975) principle of cooperativity. It would be highly uncooperative to shift away from a previously agreed upon referential pact (and shared perspective) and use a different term. Indeed, adults are slower to identify a referent if their partner uses a term different to the agreed referential pact (e.g., uses *a dog* instead of *a cocker spaniel*), even though it is an acceptable description of that object (Metzing & Brennan, 2003; Shintel & Keysar, 2007). However, if they are interacting with a new partner who they did not share a referential pact with, they will refer to an object in an optimally informative way (Brennan & Clark, 1996) and there is no difference in their reaction times for identifying an object when their new partner uses a different term compared to when their familiar partner uses the referential precedent (Metzing & Brennan, 2003).

However, unmediated accounts of referential maintenance, such as the cue-based or priming-based accounts, take a more egocentric view and assume that a referential precedent inhibits the use of other terms, and its maintenance is not necessarily a reflection of perspective-taking abilities. For example, cue-based accounts explain referential precedent maintenance through lower-level cue-based memory mechanisms. Under this account, a referential expression is associated and stored with different objects, partners and contexts. Thus, a partner acts as a cue in subsequent interactions which activates a specific referential expression and inhibits the use of other terms (e.g., Horton & Gerrig, 2005). Alternatively, priming-based accounts (e.g., Interactive Alignment Model; Garrod & Pickering, 2004), assume that speakers' tendency to reuse their partner's referential precedent is a simple response to their partner's linguistic behaviour: simply hearing a term (e.g., *pony*) increases the activation of that term and inhibits other responses (e.g., *horse*). However, priming-based accounts cannot explain the partner-specific effects seen in adults. Whilst mediated and unmediated accounts of referential maintenance assume different underlying mechanisms, they are not necessarily mutually exclusive (Branigan et al., 2011).

Children find referential communication especially difficult to do. Even young school age children, who have mastered most of the structural features of their language, frequently produce unclear and ineffective descriptions when they are trying to describe a specific referent to an addressee (e.g., describing an object as *the red one* even though there are other objects of the same colour present; Deutsch & Pechmann, 1982; Glucksberg & Krauss, 1967; Glucksberg et al, 1966; Matthews, Lieven & Tomasello, 2007). Whilst children's difficulty in referential communication is well established, it is still unclear why this difficulty persists so late in development.

Chapter 5

The most prominent explanation for young children's poor performance in referential communication has focused on perspective-taking, and specifically difficulties in recognising or accommodating others' perspectives during language use. In order to produce an informative referential expression, a speaker must monitor their partner's perspective against their own and use this information to plan their utterances (Clark, 1992; 1996). For example, for a speaker to felicitously refer to *the horse*, she must be sure that her addressee is aware of only one horse. In early research, children under the age of 7 were assumed to be egocentric, and so they could not recognise or adopt a perspective other than their own favoured perspective (Piaget, 1926). Accordingly, they were claimed to be incapable of tailoring their communicative acts to the mental states of their partner, and so produced utterances that appeared to not take their partner's perspective into account (Piaget, 1926; Deutsch & Pechmann, 1982). However, more recent work has cast doubt on the assumption that young children are incapable of recognising that others may favour perspectives that are different from their own: Despite their tendency to produce egocentric utterances, study after study has shown that young children are able to reason about the mental states – the knowledge and beliefs – of others and adapt their communicative acts in accordance with this (Akhtar, Carpenter, & Tomasello, 1996; Liebal, Carpenter & Tomasello, 2010; Matthews et al., 2007; Matthews, Lieven & Tomasello, 2010; Moll, Richter, Carpenter & Tomasello, 2008). For example, even 18-month-olds are able to track their shared experiences with specific individuals, and then use this information when deciding what to declaratively point to (Liebel et al., 2010). These findings suggest that young children are able to recognise that their partner may have perspectives different from their own, and have led to proposals that their production of egocentric utterances may instead reflect an inability to manage the cognitive burden of integrating this knowledge with the already complex task of formulation (de Cat, 2015; Epley, Morewedge & Keysar, 2004a; Nadig & Sedivy, 2002; Nilsen & Graham, 2009).

Indeed, recent theories have suggested that children's general cognitive limitations, such as their developing inhibitory skills, may explain children's inability to use their perspective-taking skills in the utterances that they produce. Indeed, both children and adults have egocentric biases in the perspectives that they adopt when producing and understanding referential expressions. That is, they have a default favoured perspective that is always activated irrespective of context. Adults can inhibit this default perspective in a much more efficient manner than children, and as a result will communicate and process referring expressions in a way that is consistent with their partner's perspective (*inhibition-of-perspective* account; e.g., Brown-Schmidt, 2009; de Cat, 2015; Epley, Keysar, Van Boven & Gilovich, 2004b; Nilsen & Graham, 2009; Wardlow, 2013). Consistent with this, during comprehension both children and adults show automatic egocentric interpretations of their partner's utterances, but adults are much quicker and more effective at correcting their interpretation than children are (e.g., Epley et al, 2004b). Likewise in production, adults are more likely to produce egocentric utterances (i.e., those consistent with their default perspective) than utterances that take their partner's perspective into account in cases where it is more difficult to override their biases, such as when they are under cognitive load or time pressure than when they are not (because adjusting to your partner's perspective requires time and mental effort; e.g., Epley et al., 2004*a*; Ferreira, Slevc & Rogers, 2005). However, it is unclear to what degree inhibitory skills determine children's ability to produce utterances that accommodate their partner's perspective: Although Nilsen and colleagues (2009) showed that 4-5-year-old children who have better inhibitory skills are more likely to override their egocentric biases when interpreting their partner's statements (see also Khu, Chambers & Graham 2020 for similar results), they did not find strong evidence for this relationship in children's *production* of sentences (Nilsen & Graham, 2009; Nilsen, Varghese, Xu & Fecica, 2015).

The studies discussed so far have typically used referential communication paradigms which young children may find especially challenging and thus require more mental effort. As a result, they may perform similarly to adults who are under cognitive load and thus produce more egocentric utterances (Epley et al., 2004*a*). For example, these tasks often ask participants to describe a target object amongst distractor objects that are identical to the target apart from on one dimension (e.g., a big horse is the target, and a small horse is the distractor). To succeed in these tasks (i.e., to produce a non-egocentric utterance; *the big horse*) the child must monitor the context for any potential ambiguity (i.e., pro-active monitoring), identify and distinguish the target object from the distractors, and encode any distinguishing features into their chosen referring expression, all whilst inhibiting more canonical under-informative expressions (e.g., *the horse* in multiple horse contexts). It is possible that the complexity of these tasks mask evidence of perspective-taking during production in young children. Indeed, children tend to produce more informative responses in cases where there are fewer distractor items, and thus fewer responses to inhibit (Davies & Kreysa, 2018; Nadig & Sedivy, 2002).

One way to investigate perspective-taking during referential communication (especially during production) is to focus on children's abilities to adopt their partner's perspective on an object (or maintain referential precedents) throughout an interaction. If children are sensitive to referential precedents and expect their partner to maintain them, this would show that they can use perspective-taking skills during referential communication. Furthermore, if children themselves maintain a referential precedent, it would provide further evidence that children can inhibit their own favoured perspective (*horse*) and instead adopt their partner's favoured perspective (*pony*).

The ability to maintain a referential precedent is evident in much older children. For example, typically developing seven-to-twelve-year-olds navigating a maze tend to re-use referential expressions that their partners have just used (e.g., *line* vs *row;* Garrod & Clark, 1993), and both typically developing and autistic eight-to-ten-year-olds tend to name pictures in a picture-matching game using the same perspective - even when usually disfavored - as an adult experimenter previously used (Branigan, Tosi, & Gillespie-Smith, 2016; Hopkins & Branigan, 2020; Hopkins, Yuill, & Branigan, 2017; 2021).

However, early findings suggest that pre-schoolers struggle to maintain a referential precedent as they are unable to adopt multiple perspectives – and therefore names – on the same referent, and instead favour their own default perspective during both comprehension and production. For example, three-to-five-year-olds show a mutual exclusivity bias during word-learning, and they assume that new words they hear refer to new objects (Markman & Wachtel, 1988; Markman, 1989). Furthermore, in studies using an alternative naming paradigm whereby participants name an object (e.g., *horse*) and judge whether a partner's use of an alternative name (e.g. *pony*) is acceptable, three-year-olds consider their partner's use of a different name to be incorrect (Doherty & Perner, 1998; Perner, Stummer, Sprung & Doherty, 2002). Likewise in production, children's ability to flexibly adopt multiple perspectives on a referent appears limited: whilst three-year-olds can recognize and use different perspectives when given pragmatically appropriate contexts and/or cues that explicitly elicit multiple perspectives (e.g., calling a dinosaur-shaped crayon both *a dinosaur* and *a crayon* when probed to use an alternative perspective; Clark, 1997; Deák & Maratsos, 1998; Deák, Yen & Pettit, 2001; Mervis, Golinkoff, & Bertrand, 1994), they struggle to spontaneously switch between perspectives (i.e., to abandon an established precedent and adopt a new perspective). In Deák et al.'s (2001) study, three-year-olds were less likely to adopt a new perspective on an object after they already taken another perspective on it than when they had not.

However, more recent studies have examined pre-schoolers' sensitivity to referential precedents, with the idea being that developmental data can help distinguish between accounts explaining referential precedent maintenance in terms of perspective-taking and accounts explaining them in terms of more domain-general processes (Graham, Sedivy & Khu, 2014; Koymen, Schmerse, Lieven & Tomasello, 2014; Lindsay, Hopkins & Branigan, in prep; Matthews et al., 2010). These studies have typically looked at the effect an old or new partner has on children's understanding or production of referential precedents. For example, Matthews and colleagues (2010) showed that both 3- and 5-year-olds were sensitive to referential precedents that their partner had established during comprehension. In their study, an experimenter asked children to move objects with two possible names (*truck* vs. *lorry*). For half of the trials, the same experimenter then asked them to move the same object using the same name (i.e., they maintained the referential precedent; *truck-truck*) or a different name (i.e., they broke the referential precedent; *truck-lorry*). The other half of the time, a different experimenter asked them to move the same object, but used the same name that the first experimenter had (*truck*) or a different name (*lorry*). Both 3- and 5-year-olds showed an adult-like pattern in that they were slower to pick up objects when the same experimenter used a different name (i.e., broke the referential precedent) than the same name. However, there was no difference in times to pick up the object when the different experimenter used the same or different name. 4-year-olds also show a similar effect but for modified noun phrases (e.g., *striped ball* vs. *purple ball;* Graham et al., 2014). These results suggest that children are sensitive to referential precedents (that is, they expect their partner to maintain a precedent) and have been interpreted as evidence that children's understanding of referential precedents is rooted in perspective-taking.

Likewise, partner-specific effects have also been found during production, but these effects are much more limited in pre-schoolers. For example, Koymen and colleagues (2014) showed that both 4- and 6-year-olds would successfully establish partner-specific referential precedents with a peer when that precedent was a proper noun (e.g., *Peter*; see also Birch & Bloom, 2002; Diesendruck & Markson, 2001; Diesendruck, 2005). However, when children were asked to establish a referential precedent for an object (e.g., *a woman's shoe*), 4-year-olds failed to maintain the precedent with the same partner once that expression became overinformative (i.e., because there was only one shoe in the array of pictures to describe), and instead shifted to using a simpler expression (e.g., *shoe*), whereas 6-year-olds behaved in an adult-like way and continued to reuse the referential precedent (e.g., *a woman's shoe*). These results have been interpreted as showing that children can establish referential precedents but that younger children, who have limited abilities using perspective-taking during production, are less consistent in their ability to do so. Furthermore, more recent work in our own lab (Chapter 4) has found that pre-schoolers will not only adopt their partner's perspective, but will also continue to use that perspective in their own language use, even if that perspective is not their usual choice (i.e., they maintained their partner's referential precedent; Lindsay et al., in prep). They continued to do this - although to a lesser extent - even after they had previously named the object using their favoured perspective, suggesting that even pre-schoolers will adopt their partner's perspective and maintain that referential precedent. However, we failed to find any partner-specific effects, as children adopted their partner's perspective to the same extent regardless of whether they had previously named the objects with the same or a different partner. This could suggest the mechanisms underlying referential precedent maintenance are not necessarily wholly driven by perspective-taking mechanisms.

All in all, our recent work, as well as the studies discussed so far, suggest that even 3-4-year-olds can and do maintain referential precedents, even if that precedent is not their default (and favoured) perspective. Our recent work further suggests that such referential maintenance holds for categories of referents, and not a specific exemplar. However, there are inconsistencies in the production data, with pre-schoolers maintaining a referential precedent in some instances, and failing to in others. In this study, we investigate referential maintenance in pre-schoolers further by examining the circumstances by which children adopt their partner's perspective and discuss what this might tell us about the mechanisms underlying such effects. Most work on referential precedent maintenance has manipulated partner identity to investigate the extent to which referential maintenance is based on perspective-taking. That is, they examine whether children will maintain the same term with the same partner compared to a different partner, with the idea being that if referential maintenance is based on our beliefs about what a partner can understand, they should maintain that precedent with the same partner (who has shown evidence of understanding that term), but not necessarily with someone else. Here, we take a different tack and hold the partner constant and instead change the context (i.e., the task). If referential maintenance is based on beliefs about what a partner can understand, children should continue to use a precedent once it has been established, even in a different context. However, if perspective-taking mechanisms are affected by failures to inhibit an egocentric (i.e., favoured) perspective, then children should perform similarly across contexts.

To test this, 3-4-year-olds completed two referential communication tasks with the same partner. Both tasks had previously been used successfully with this age group, and they showed referential maintenance within each task. We measured whether children would maintain a referential precedent and whether this referential maintenance was persistent across tasks. For the first task, we used a ‘snap’ paradigm in which a child and experimenter alternated turning over cards and naming pictures of familiar objects in a card-matching game (see Branigan, McLean & Jones, 2005). On experimental trials, the experimenter named an object compatible with two perspectives (reflected in two alternative names; e.g., *horse/pony*) using either the favoured or disfavoured perspective, with favoured/disfavoured perspectives having been established by a pre-test. Following two filler trials, the child named the same target object. For the second task, we used the same “Moving House” task that we have used previously (see Lindsay et al., in prep) whereby children and the experimenter took turns giving each other a series of instructions about where to place objects in a house. The experimenter gave instructions involving objects that did not appear in the first task; in the next block, the child gave instructions involving the different referents of objects seen in Task 1 during the snap game. For both tasks, our dependent measure was children’s likelihood of using the less favoured perspective (i.e., the disfavoured name) when referring to an object after hearing the experimenter use the disfavoured vs. favoured name in Task 1. If referential maintenance relies on perspective-taking mechanisms, children should be more likely to use a disfavoured name after their partner has used a disfavoured name in Task 1. This referential maintenance should then extend to Task 2: children should be more likely to continue using that disfavoured name in the second task if their partner used a disfavoured name in the first task. However, if children’s issue during referential communication is due to a failure in inhibiting their own egocentric perspective, then they should fail to maintain a referential precedent across both tasks.

5.2. Method

Participants. Twenty-one native English-speaking children (female= 11; mean age = 44.6 months; range=37-56 months) from the Edinburgh area participated. One further child was excluded as they were unable to complete the task. Children were tested in the Developmental Lab at the University of Edinburgh. None of the children had any known language or developmental disorders. Ethical approval for this experiment was obtained from the Psychology Research Ethics Committee at the University of Edinburgh.

Materials. For the 'snap' game used in Task 1, we chose twelve experimental items from the experimental items used by Branigan et al. (2016) that comprised pairs of objects from the same category (a prime and target) that had two alternative names (e.g., *horse* vs. *pony*), and a scripted prime name (favoured vs. disfavoured). The prime and target pictures used different tokens of the same object (see Fig 1). These items had already been pre-tested to establish that children in our population could both understand and describe them using two different names, but they had a strong preference for one alternative (i.e., the favoured item). We prepared two paired lists (experimenter/participant), each containing one version of the experimental items in a Latin Square design, plus 24 filler items (12 participant filler items; 12 experimenter filler items). Item order was randomised with the following constraints: two filler items appeared between the experimenter's prime card and the participant's target card; and four snap items distributed among trials.

For the "moving house" game used in Task 2, we wanted to see if children would maintain the referential precedent that they had developed with the experimenter. Therefore, participants named the *same* experimental items as in the snap game, but we used *different* tokens of those items (to ensure that if they did maintain a referential precedent, a memory account would not explain it). Because we were not priming participants in the moving house game (Task 2), the experimenters named 12 different filler items that did not appear in the snap game.

We prepared 8 A3-sized printed house scenes (4 for experimenter rounds and 4 for participant rounds, with one from each set being used for a practice) and printed removal trucks. Each house included four different rooms (living room, kitchen, bathroom and bedroom) plus other locations outside (e.g., a shed). For participants' rounds, we prepared A4-sized booklets that showed a miniature version of the experimenter's house and depicted the intended location of each of the experimental items, so the child could tell the experimenter where to put the items from the removal truck. Each page of the book highlighted one of the items that had to be moved. Thus, item order was fixed for each house, with each experimental item appearing in the same position in each round. However, we counterbalanced the order the houses appeared. Each round consisted of four experimental rounds (where the participants named the objects to be moved) and four filler items (in which the experimenter told participants which objects should be moved and where they should go).

Procedure. Participants first complete Task 1 and played a game of snap with an experimenter (see Figure 22 for full procedure). The experimenter placed a pre-arranged pile of cards on the table picture face down and told the participant that they were going to play a game of snap, where they will take turns naming pictures on cards, whilst looking for 'snap' cards (i.e., those that match). The experimenter turned over the top card and named it (following a script) – this constituted the prime. The child then turned over their top card and named it – this constituted the target response. The game continued like this, with the experimenter and participant alternating until all the cards had been named. If the participant and experimenter cards matched, the first player to shout "snap" won the cards in play.

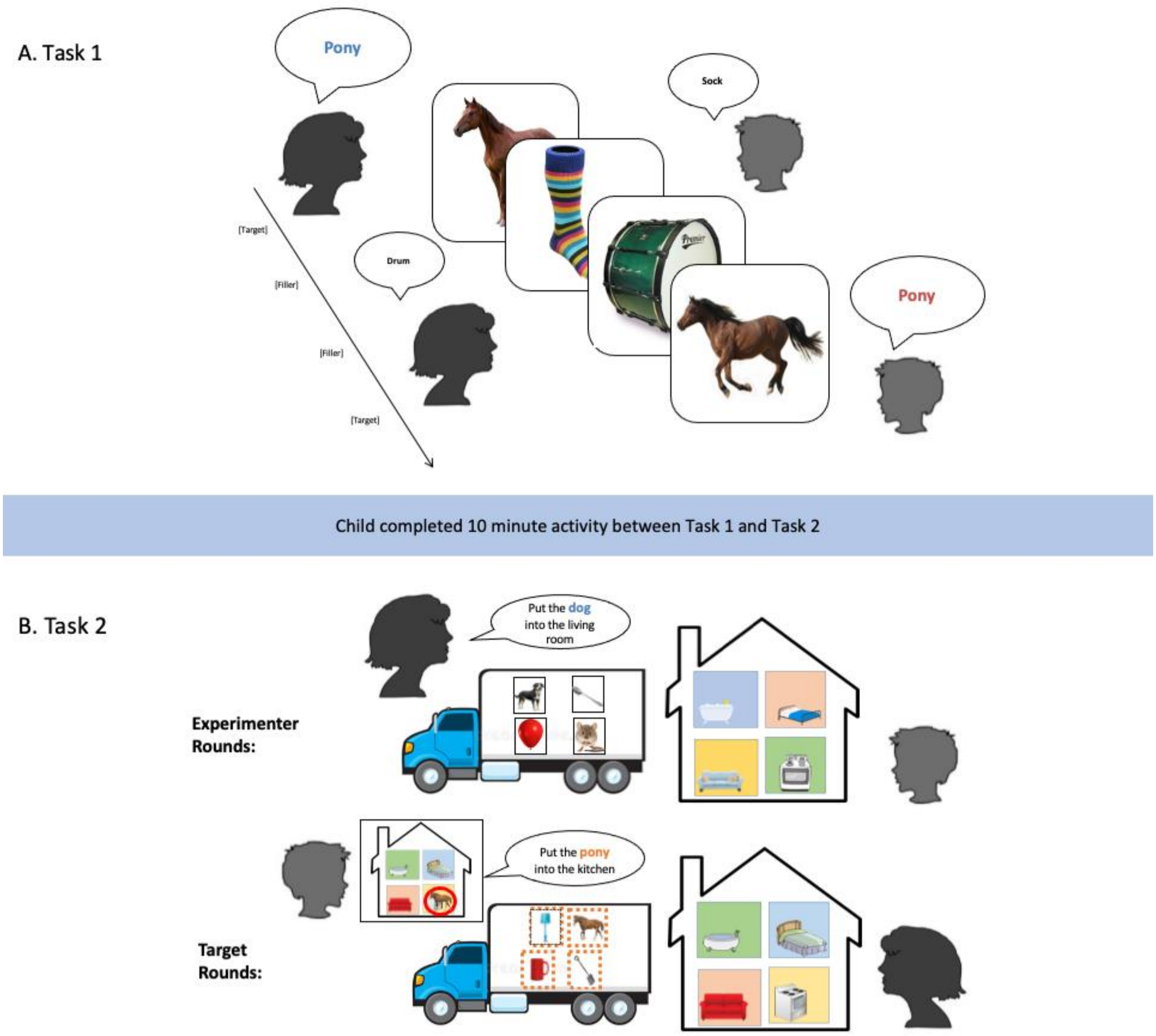

Figure 23. Example trial in each task. A) Example trial of the snap game (Task 1). The prime and target picture show different tokens of the same category. After a 10 minute delay where children completed an activity separate from the current study, they began the second task. B) Example round of the Moving House game (Task 2). During experimenter rounds, the experimenter told the child where to move each of the four items from the truck to the house. These items did not appear in the first task and the experimenter did not relabel any tokens that appeared in the first task. Once all items in the truck had been moved, the experimenter

and child reversed roles and the target round began. The items used in target rounds were different tokens of the same items used in the snap game.

After the snap game finished, the participant and experimenter completed a different task (colouring in) for 10 minutes. The experimenter and participants then completed Task 2 and played the moving house game. The experimenter explained how to play the moving house game, and then they played a practice round, where the experimenter gave the participant an A3 house and a removal truck with items arranged on it. The experimenter then asked the participant to move each item into a different room of the house (e.g., *put the car into the shed*). After all the items had been moved, the experimenter and participant switched roles. The participant then told the experimenter where each item should be moved into the house, using the booklet as a map.

Once the experimenter was satisfied the participant understood how to play, they began the main rounds. During the experimental rounds, participants told the experimenter where items should be moved. Crucially, the items to be moved were the same as those in the previous snap game but were different tokens of those items from the snap game. During filler rounds, the experimenter told participants where each item should be moved. These items were completely different from those in the snap game.

Coding and Analysis. Target responses were coded as either favoured (e.g., *horse*), disfavoured (e.g., *pony*), or other (e.g., *horsie*). Trials were excluded from our analysis if participants did not produce a response. The data was analysed in two ways. We first examined target responses from the snap game (Task 1). In this analysis, we were interested in whether children would use a disfavoured name after hearing a disfavoured prime. Thus, we coded responses where children used a disfavoured name as 1, and all other responses as 0. Our second analysis was concerned with whether children continued to use a disfavoured name in the moving house game (Task 2). As in our snap game analysis, we coded disfavoured responses as 1 and all other responses as 0.

We used generalised linear mixed-effects models using the lme4 package (0) in RStudio (version 1.2). All models included participants and items as random effects (Baayen, Davidson & Bates, 2008), and incorporated maximal random effects structure (Barr, Levy, Scheepers & Tily, 2013), except when otherwise specified. In cases where the model would not converge, correlations among random effects were fixed to zero (Bates, Kliegl, Vasishth, & Baayen, 2015). Unless otherwise specified, all predictors were contrast coded (-0.5, 0.5) and centered. For our GLMM models, we report coefficient estimates (b), standard errors (SE) and z and p values for each predictor; 95% confidence intervals (CI) are from the confint function (method=Wald). Due to our low sample size, we also calculated Bayes Factors (BF) by using the models' Bayesian Information Criteria (BIC) values. The use of Bayes Factors has been argued to offer advantages over traditional power analyses (Dienes, 2014) and also allowed us to quantify the strength of evidence for the alternative hypothesis versus the null hypothesis. Following Dienes (2014), we interpret a Bayes Factor of greater than 3 as strong evidence for the alternative hypothesis over the null, less than 1/3 as strong evidence for the null hypothesis over the alternative, and anything between these values as support for neither the alternative nor the null. We estimated a Bayes Factor as $e^{(BIC\_null - BIC\_experimental)/2}$ from the BIC values of both the experimental and null models (Wagenmakers, 2007).

5.3. Results

Table 20 reports the frequency of different responses for the Snap Game round and the Moving House Game round. Figure 22 shows the proportion of trials where children used a disfavoured name as a function of whether they heard a favoured or disfavoured prime (in Task 1) for both tasks.

Our first analysis focuses on the snap game (i.e. Task one) and asks whether children were more likely to use a disfavoured name after a disfavoured prime. Our second analysis focuses on the moving house game (Task 2).

Table 20. Frequency of each response type by prime condition for the snap game round and moving house game round.

		Prime Type	
Prime Type	Response	Favoured	Disfavoured
Task 1	Favoured	96	49
	Disfavoured	11	58
	Other	17	13
	No Response	2	6
Task 2	Favoured	80	65
	Disfavoured	16	35
	Other	18	16
	No Response	12	10

Table 21 shows the proportion of disfavoured responses for when the experimenter used a favoured or disfavoured prime in both tasks. When the experimenter used a disfavoured name (i.e., disfavoured prime), children used a disfavoured response on 46% of trials [SD=.31]. When the experimenter used a favoured name (i.e. favoured prime), children used a disfavoured response on 9% of trials [SD=.11]. Our mixed effects model indicated that children were more likely to use a disfavoured response after a disfavoured prime than a favoured prime (B = 3.20, SE= .91, z= 3.52, CI= [1.42, 4.98], p<.001). This replicates our previous findings demonstrating that children are able to reuse their partner's referential precedent in conversation, and suggests that children are not simply using their own favoured perspective to describe pictures as the inhibition-of-perspective account would suggest.

Table 21. Proportion of disfavoured responses by prime condition for the snap game and moving house game. Proportions calculated over all responses (excluding no responses). Means over subjects (SE). The alignment effect is the difference between the favoured and disfavoured prime conditions.

	Prime Type		
Task	Favoured	Disfavoured	Alignment Effect
Task 1 (Snap)	.09 (.02)	.46 (.07)	.37
Task 2 (Moving House Game)	.14 (.03)	.30 (.05)	.16

Task 2. Our next analysis is concerned with whether children will continue to reuse the referential precedent their partner established in Task 1. If children maintain referential precedents for perspective-taking reasons, they should continue using their partner's favoured name in Task 2, even if this is a disfavoured perspective.

To test this, we first looked at the effect of prime on children's likelihood of using a disfavoured name in the second Task after a favoured vs. disfavoured prime in Task 1 using a logistic model with the form Disfavoured Response in Task 2 ~ Task 1 Prime. Our model found a marginal effect of prime (B = 2.84, SE= 1.51, z= 1.88, CI= [-.12, 5.79], p=.06, BF = 0.0006). However, the corresponding Bayes factor is less than 1/3, suggesting that this is a true null effect. This suggests that children were less likely to use their partner's perspective in Task 2, if that perspective was disfavoured than if it was favoured.

However, we wanted to confirm that there was a significant difference in children's likelihood of using a disfavoured name across tasks. To test this, we pooled our Task 1 and Task 2 data together and analysed whether children were more likely to use a disfavoured name after a disfavoured prime in Task 1 across both tasks using a logistic model with the form Disfavoured Response ~ Task 1 Prime * Task. In this model, we are specifically looking for a main effect of prime: the perspective-taking account predicts that children would be more likely to use a disfavoured name after a disfavoured prime across both tasks.

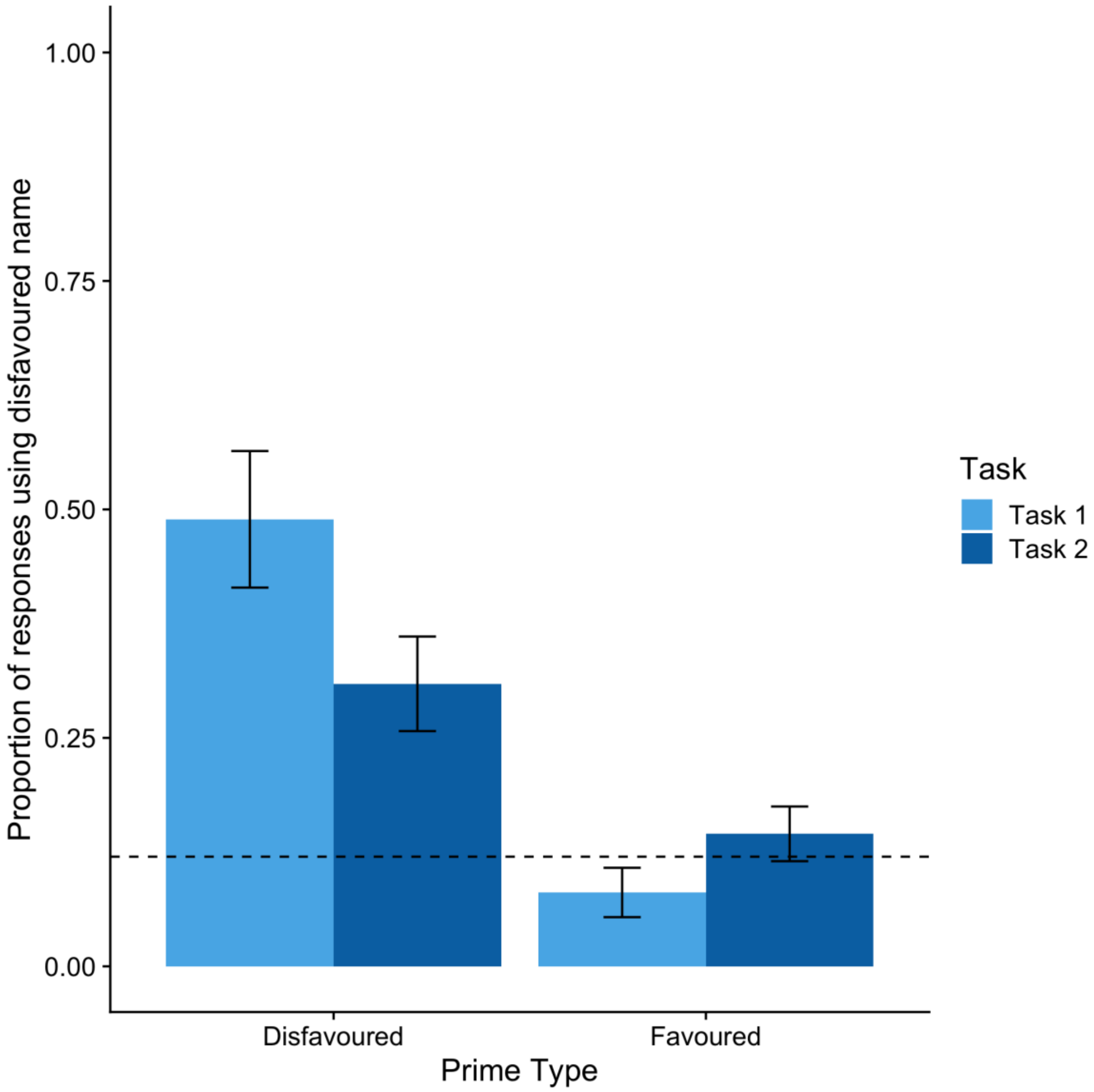

Figure 24. Proportion of responses using disfavoured name in the moving house task as a function of prime type and Task. Bars indicate ±1 standard error.

Our model indicated that children were more likely to use a disfavoured name in general if the experimenter used a disfavoured prime in Task 1 than a favoured prime (38% vs. 11.5%; B = 1.99, SE= .30, z= 6.65, CI= [1.43, 2.61], p<.001, BF = 66501592840). There was no effect of task on children's use of a disfavoured name (B = .14, SE= .53, z= .26, CI= [-.44, 0.66], p=.79; BF = .05) but there was a significant interaction between task 1 prime and task (*B* = 1.63, *SE*= .58, *z*= 2.81, *CI*= [.59, 2.80], *p*<.01, BF = 4.97): Children were more likely to use the disfavoured name after a disfavoured task 1 prime in Task 1 than in Task 2 (46% vs 30%: *B* = .98, *SE*= .32, *z*= 3.06, *CI*= [.35, 1.61], *p*<.01, BF = 8.40). However, children were marginally more likely to use a disfavoured name after a favoured task 1 prime in Task 2 than in Task 1 (*B* =-.87,, *SE*= .52 , *z*=-1.69, *CI*= [-1.89, .14], *p*=.09, BF =.30). Overall, this suggests that children were less likely to adopt the same perspective as their partner in the second task.

5.4. Discussion

Children's poor referential communication abilities have often been attributed to difficulties inhibiting their own favoured perspectives during language use. However, studies investigating children's sensitivities to referential precedents have provided inconsistent results, with some studies suggesting children can use perspective-taking mechanisms (by adopting their partner's favoured perspective on an object; i.e., maintain a referential precedent) during language use, and others not. In this study, we further investigated children's abilities to adopt a partner's perspective (and thus maintain a referential precedent) by testing whether children would continue to adopt a partner's perspective across different tasks (i.e., contexts). If children are using perspective-taking mechanisms, we hypothesised they would continue to adopt their partner's perspective across both tasks. However, if children fail to inhibit their own perspectives, we predicted children would adopt their own favoured perspective, but this would hold for both tasks (i.e., a null result).

We found that children could maintain a referential precedent, but only within the first task: children were more likely to use a disfavoured name in the first task after a disfavoured prime than a favoured prime. This replicates our previous finding (Lindsay et al., in prep). However, we found a much weaker effect in the second task. Whilst this effect was marginally significant and the corresponding Bayes Factor suggests that it is a truly null effect, it should be noted that Bayes Factors approximated using BIC (as we have done in this study) are generally more conservative and so it is possible that if we calculated Bayes Factors in a different way (e.g., with Bayesian regression models and stronger priors), we may find this effect to be a true one. However, the effect of disfavoured prime on children's likelihood of maintaining a referential pact does not seem to be long lasting: when we examined children's likelihood of using a disfavoured perspective as a function of task and task one's prime, we found a significant interaction between task and prime: children were more likely to use a disfavoured perspective after a disfavoured prime in the first task than in the second task, providing further evidence that children used the disfavoured name to a lesser extent after a disfavoured task prime in the second task. Overall, this suggests that children can maintain a referential precedent that their partner established immediately after their partner uses a specific name, but this effect is not long lasting (although, we cannot distinguish whether this was affected by the time delay or change in context).

We have a small sample size in this study and it is important that we increase our sample size before we draw firm conclusions. However, we can make some tentative conclusions. Our results from the first task replicate our previous results (Lindsay et al., in prep): Pre-schoolers adopted their partner's perspective on a referent and used it in their own production. This provides further evidence that 3-4-year-olds cab flexibly adopt multiple perspectives on a referent. As in our previous experiments, this occurred without explicit probing (e.g., by using questions to cue shape or function; Deák et al., 2001), within the same pragmatic context (e.g., rather than in response to different contrast sets; Clark, 1997), and extended across a whole category of a referent rather than on a specific exemplar from a category. Thus, we provide further evidence for pre-schoolers' cognitive flexibility in both conceptualisation and communication. Furthermore, these results demonstrate communicative perspective-taking abilities in pre-schoolers: children in our study were able to inhibit their own favoured perspective in the first task and adopted their partner's perspective during production in order to maintain a referential precedent their partner had established. This result is consistent with previous work demonstrating that children can establish a referential precedent with their partner (Graham et al., 2014; Koymen et al., 2014; Matthews et al., 2010; Lindsay et al., in prep). Overall, task one's findings suggests that preschoolers' short-term referential choices are responsively sensitive to a conversational partner's perspectives and associated referential expressions in the same way as adults and school-aged children (Branigan et al., 2016; Brennan & Clark, 1996; Garrod & Anderson, 1987; Garrod & Clark, 1993; Hopkins & Branigan, 2020; Lindsay et al., in prep).

However, the effects from Task One were short-lived and much weaker in the second task: pre-schoolers in our study did not maintain the referential precedent their partner set in the first task to the same extent as they did in the first task. Thus, after a delay and within a different context, children tended to revert to using their favoured perspective (we see this in the opposite direction in task two as children tended to use a disfavoured name more after a favoured task one prime). This result is inconsistent with previous findings suggesting that pre-schoolers are less likely to adopt a new perspective on a referent if they have already taken a perspective on it (Deak et al., 2001). However, it echoes Koymen et al's (2014) result showing that 4-year-olds failed to maintain a referential precedent that they had set with their partner. In Koymen et al's study, the referential precedent was established in trials 4-6. As in our study, there was a delay in the experiment between establishing the precedent and referring to that same referent (i.e., they created the precedent in trial 6, and referred to the same referent in trial 9), but also a change in context: the images used in the visual array to establish the precedent were different to when they referred back to that referent (e.g., different kinds of shoes when establishing the referent, but a shoe amongst other objects when referring back to the referent). Overall, these results, alongside our findings from the second task, suggest that pre-schoolers can maintain a referential precedent that their partner established but are less likely to maintain this precedent across time and over contexts. What can this tell us about children's referential production abilities?

5.4.1. An Alternative Referential Production Hypothesis: Inhibition-of-Production Account

Much of the work concerned with referential communication has focused on perspective-taking and specifically whether speakers use their partner's perspective or not during utterance planning. Children's poor performance in referential communication tasks has often been attributed to failures with inhibiting their own perspective (*inhibition-of-perspective* account; e.g., de Cat, 2015; Epley et al., 2004*a*; Nilsen & Graham, 2009). Even work examining the mechanisms underlying referential precedent maintenance assumes pre-schoolers inconsistencies in maintaining referential precedents is due to a limited ability to use perspective-taking skills (Koymen et al., 2014). However, our results do not support such a conclusion as we find evidence of communicative perspective-taking in the first task (as pre-schoolers were able to adopt their partner's perspective) but a much lesser use of such perspective-taking abilities in the second task (whereby pre-schoolers tended to produce their favoured perspective). How can we explain such inconsistencies in our results?

Many studies investigating the inhibition-of-perspective account has attempted to find a relationship between inhibitory skills and an ability to produce (and understand) utterances that take their partner's perspective into consideration: Whilst Nilsen and colleagues (2009) have found this relationship in children's abilities to understand their partners' utterances, they have failed to find strong evidence for such a relationship in children's *production* of sentences (Nilsen & Graham, 2009; Nilsen, Varghese, Xu & Fecica, 2015). An alternative possibility is that children's developing inhibitory skills might directly affect their language processing, which in turn affects their ability to produce utterances that take their partner's perspective into consideration (*inhibition-of-perspective* account).

Inhibitory skills play an important role in adults' sentence production, including lexical retrieval, where research suggests that speakers initially activate multiple alternatives and then inhibit competing alternatives to select the appropriate word (e.g., Dell, 1986, 1988; Dell, Oppenheim & Kittredge, 2008; Levelt, 1999). Thus, when describing an object in typical referential communication tasks, speakers must inhibit the prepotent, dominant and more frequent perspectives (e.g., *the horse*) so they can produce an utterance that takes their partner's perspective into account (*pony*). However, in cases where a partner's perspective becomes more activated (for example, after hearing their partner name an object), then it would be easier for a speaker to inhibit their favoured perspective. This kind of account therefore assumes a bi-directional relationship between word forms and their associated perspectives. Simply hearing a word (*pony)* activates the word form *pony* which subsequently activates the perspective PONY which then inhibits more dominant and favoured perspectives like HORSE.

The inhibition-of-production account therefore predicts that pre-schoolers – whose inhibitory skills are still developing – would struggle to inhibit their own favoured perspective during production. Indeed, we can explain classic findings demonstrating children's inability to produce informative referential expressions with this hypothesis. For example, Pechmann and Deutsch (1982) asked children (2- to 9-year-olds) to describe an object they liked best out of an array of 8 objects. In order to uniquely identify a specific referent, children had to identify multiple dimensions (e.g., size, shape, colour), thus giving children more choice in deciding how to describe the referent as well as increasing the complexity of formulation processes. They found that in this task, 6-year-olds produced inadequate descriptions (i.e., descriptions that would not specify a unique referent) 50% of the time, whilst 2-year-olds produced such descriptions 94% of the time (Pechmann & Deutsch, 1982). However, this task requires children not only to inhibit responses associated with the other objects in the array (e.g., *the small green spoon*), but also to inhibit more dominant labels (e.g., *big yellow spoon* could simply be described as *spoon* because spoon is the more typical label).

We can explain our results in this experiment using this account: Pre-schoolers were able to adopt their partner's perspective in task one (e.g., *pony*) because they had heard their partner use that particular name just a short while ago. Hearing their partner's perspective activated that perspective and inhibited their own (the participant's) favoured perspective. As a result, it looked like they were able to maintain a referential precedent – even if this was a disfavoured perspective – and perform in an adult-like way. However, during the delay between the first and second task, the activation of the partner's perspective decreased which meant that it was more difficult for our pre-schoolers to inhibit their favoured perspective. As a result, they were less likely to use their partner's perspective in the second task compared to the first and instead produced a word that they found easier to access. Consistent with this, our previous work found that three-to-four-year-olds were more likely to maintain their partner's referential precedent after a short delay, but this effect was reduced after a pre-naming phase (whereby the favoured perspective was more activated) regardless of whether they had named that object with the same or different partner (Lindsay et al., in prep).

This kind of account, alongside our findings, support a cognitive load hypothesis which suggests that children find production difficult (e.g., Charest & Johnston, 2011), and as a result, would be more susceptible to factors that make production easier (such as environmental influences). Thus, children's difficulties in producing informative and perspectival utterances are because the overall act of speaking is just difficult for them.

5.4.2. Mediated vs. Unmediated Mechanisms Underlying Referential Precedent Maintenance

Whilst the goal of the study was not to distinguish between the theories underlying referential precedent maintenance in pre-schoolers, our results can add to the literature investigating these mechanisms. Whilst our findings from task one suggest that children can maintain a referential precedent (and adopt a perspective) that their partner establishes, they show that such referential maintenance is short-lived and does not generalise across tasks. This finding contradicts mediated accounts of referential maintenance, especially those that place a central role of perspective-taking mechanisms: If referential maintenance in pre-schoolers was solely down to perspective-taking mechanisms, children should have continued adopting their partner's perspective in the second task as to the same extent – unlike previous studies – we kept participants' partner constant throughout the whole experiment and only changed the context that referential precedent occurred in. These findings contradict previous studies demonstrating perspective-taking abilities in children's referential maintenance (e.g., Graham et al, 2014; Koymen et al., 2014; Matthews et al., 2010). However, these findings are also compatible with other non-perspective-taking interpretations: Pre-schoolers' slowed comprehension after a speaker failed to maintain a referential precedent (Matthews et al., 2010; Graham et al., 2014) is also compatible with automatic memory mechanisms between particular contexts, partners, objects and expressions. Furthermore, recent replications using more precise methods to measure reaction time have failed to find partner-specific effects during comprehension, instead showing that children were slower to select objects after hearing a different name than a same name, regardless of whether their partner was the same or not (Ostachenko et al., 2019). However, it is possible that the ten-minute delay in our experiment caused pre-schoolers to forget what their partner's favoured perspective was, resulting in them using a perspective they themselves favoured. Similarly, the added complexity of the second task whereby the child had to identify where a referent was in a house and produce a more complex utterance (e.g., *put the pony in the garden*), may have affected pre-schoolers abilities to inhibit their own favoured perspective. Whilst our previous work has shown that children can maintain a referential precedent using this more complex task, there was a reduction in their abilities to do so compared to the simpler snap task (Lindsay et al., in prep). Therefore, it is important for future work to test whether using a simpler task as the second task would provide more positive results (i.e., show referential maintenance).

Despite these limitations, our results support unmediated accounts of referential precedent maintenance: although we cannot quantify the exact contribution these mechanisms had on children's performance in the task, we can make some assumptions about those which are most relevant. First, our results provide some evidence that memory-based mechanisms contributed to maintenance of referential precedents, through which children would tend to recall previously used precedent when cued by contexts, objects, and partners. Pre-schoolers failed to maintain the referential precedent that was established in the first task into the second task. Children may have associated the referential precedent with the first task, and so when we moved to the second task – which used a different context as well as used a different token of the referent (i.e., a different horse) – children lost two of their cues for the established referential precedent. However, it should be noted that in the first task, children did name a different token from the experimenter, yet they were still able to maintain the referential precedent, replicating our previous work (Lindsay et al., in prep). Furthermore, children did still have the partner as a cue to recall the referential precedent in the second task. It is possible that children rely on multiple cues in order to maintain a precedent (that is, only having the same partner during a task is not a satisfactory cue). However, to further examine the contribution of memory-based mechanisms on children's referential maintenance, it would be important to examine how children behave when the context and partner are the same in both tasks.

Our results are also compatible with pre-schoolers' referential maintenance being influenced by priming-based mechanisms that facilitate particular perspectives via activation within the lexicon. Under this account, hearing *pony* automatically activates the wordform *pony* in the child's lexical system, which in turn activates the lemma pony, and thereby the concept PONY (Levelt, Roelofs, & Meyer, 1999), resulting in that perspective subsequently being produced. Following use, all these representations would retain activation, which would eventually decay over time. It should be noted that whilst this sounds similar to the inhibition-of-production account, it is slightly different: priming-based accounts do not assume a specific (favoured) perspective is activated all the time (which then needs inhibited) and instead assume that a linguistic behaviour (e.g., producing a specific word) is the by-product of prior processing of linguistic material. In our study, the ten-minute delay between the first and second task would have reduced activation of the primed perspective (which happened to be the disfavoured perspective), resulting in a reduction of that perspective being used in the second task. (See also Ostashchenko et al., 2019, for converging evidence in comprehension).

In sum, we have shown that three-to-four-year-olds fail to maintain referential precedents in subsequent interactions with the same partner after a significant period of time has passed (ten minutes) and within a different context. We have proposed that such a finding could be indicative of difficulties inhibiting more dominant (and favoured) word forms during lexical retrieval, resulting in a failure to inhibit a favoured perspective. Furthermore, we have argued that our findings support unmediated accounts of referential maintenance, although we cannot quantify the contributions of specific mechanisms. Future study investigating these conclusions is required due to our limited sample size and design.

6. General Discussion

In the developmental literature, there has been a considerable amount of work explaining how children learn language. However, very little work considered how children then use that language to become active participants within their linguistic community. Do children formulate words and sentences in the same way as adults? And are the mechanisms underlying language production the same in pre-schoolers and adults? In this thesis, we set out to answer these questions, with a specific focus on the mechanisms underlying pre-schoolers' syntactic and referential planning.

The following sections first provides an overview of the findings from the four studies presented in Chapters 2 to 5 (section 6.1) before interpreting these findings in relation to theories of syntactic planning, morpho-phonological encoding, and referential communication in more detail, before finally discussing the implications of our findings on models of the developing production system (Section 6.2).

6.1. Summary of empirical findings

6.1.1. Pre-schoolers can use bottom-up information to aid syntactic planning

In the first set of experiments (Study 1; Experiments 1a & 2a), we investigated how pre-schoolers developed an utterance plan which would guide the order in which words were incrementally retrieved. To do this, we measured pre-schoolers' and adult controls' eye movements as they named two pictures of objects (as a conjoined noun phrase in Experiment 1a and locative sentence in Experiments 2a) one of which had been preceded by a subliminal cue. We showed that pre-schoolers' starting point – and subsequent word order choice – was predicted by their first fixations which in turn was influenced by our subliminal cue.

These results suggest that pre-schoolers' sentence planning is affected by visual attention for both simple and more complex sentences. Furthermore, pre-schoolers' sentence planning was susceptible to low-level bottom-up factors for simple sentences that do not involve semantic roles.

Overall, these findings are the first to suggest that pre-schoolers can use a planning strategy whereby they begin their sentence with the most salient or accessible character first and build a sentence plan based on this starting point (i.e., they can use a lexically incremental planning strategy).

6.1.2. Pre-schoolers coordinate their eye gaze with speech, just like adults

Our next study (Study 2; Experiments 1b & 2b) investigated how pre-schoolers retrieved the word forms for their sentence. We measured pre-schoolers' and adult controls' eye movements as they named objects in simple and more complex sentences. We found that pre-schoolers, like adults, tended to preferentially fixate on each object in the order they mentioned them for both simple and complex sentences, suggesting they retrieved each word form incrementally. In addition, pre-schoolers had longer gaze durations for their final word, indicating that it took longer for them to retrieve this word. This could indicate that pre-schoolers are less efficient at planning upcoming words whilst speaking.

Our results are the first to demonstrate that children's eye movements and synchronised with their speech output. Overall, these findings suggest that the mechanisms underlying word retrieval are similar in pre-schoolers and adults, but pre-schoolers are less efficient in their abilities to do so.

6.1.3. Pre-schoolers can referentially align with a partner, but this effect is not long-lasting

Our final set of experiments (Study 3 & 4; Experiments 3-8) explored the mechanisms underlying pre-schoolers' referential communication (specifically referential alignment), and asked whether pre-schoolers' choice of perspective and their associated referential expressions were affected by the perspective their partner previously used. We found that pre-schoolers were more likely to use the same perspective that their partner had previously used, even when using this perspective meant overcoming their default preference, and even after they had already taken a different perspective on that object. Our results suggest that priming mechanisms play an important role in pre-schoolers' referential communication, as we found no partner-specific effects in children's likelihood of adopting their partner's perspective (Experiment 6). Furthermore, pre-schoolers failed to maintain their partner's perspective over time and failed to generalise their partner's perspective to different contexts. This suggests that linguistic context can facilitate lexical selection (i.e., choosing a perspective), but the impact of such context can disappear quickly.

6.2. General implications and future directions

6.2.1. Implications for models of sentence planning

Speakers must develop a sentence plan that guides the order by which words are incrementally retrieved. Adults can use either a structurally incremental planning strategy, whereby they develop an overall structure for the sentence plan, which is imposed before speaking (Griffin & Bock, 2000; see Papafragou & Grigoroglou, 2019), or a lexically incremental planning strategy, whereby they begin speaking about the most salient or accessible character or object first and build a sentence plan that is consistent with this starting point (e.g., Gleitman et al., 2007).

Our results from Study 1 suggest that pre-schoolers can use a lexically incremental planning strategy for sentences that involve minimal semantic roles. In both experiments, pre-schoolers tended to name the cued picture first, demonstrating that their sentence planning was influenced by bottom-up factors that affect their attention. This occurred for both of our sentence types, suggesting that the complexity of the to-be-produced sentence did not affect the sentence planning strategy pre-schoolers used.

Overall, our results, in conjunction with previous studies showing that pre-schoolers' starting points are affected by linguistic biases (Brough et al., 2018; Ibbotson et al., 2013), suggest that pre-schoolers can use both bottom-up and top-down factors when planning their sentences. This would indicate that pre-schoolers can use both a lexically incremental and a structurally incremental planning strategy to plan a to-be-produced utterance, much like adults can (e.g., van de Velde et al, 2014).

How do pre-schoolers decide on the strategy they will use? Previous research has shown that adults are more likely to use a lexically incremental planning strategy when describing events that are hard to interpret (e.g., van de Velde et al., 2014) or when naming objects in a scene (e.g., Coco et al., 2014), but they tend to use more of a structurally incremental strategy as the default (e.g., Griffin & Bock, 2000; but see Konopka et al., 2019). Based on our results, however, we suggest that children could use a more lexically incremental strategy for easier sentences (as we showed in Experiments 1 and 2) but move to a more structurally incremental strategy when producing complex sentences that require them to consider thematic roles (as seen in other studies; Brough et al., 2018; Ibbotson et al., 2013). This hypothesis is consistent with recent work investigating sentence planning mechanisms in native and non-native adult speakers (Konopka et al., 2019). When adult speakers produced a transitive sentence in their first language (L1), they tended to use more of a lexically incremental strategy, but when they switched to using their second language (L2; and is the language they have less experience using), they also switched the planning strategy they used, and moved to more of a structurally incremental planning strategy. However, when they were familiar with the event, the L1-L2 difference disappeared, and L2 speakers also planned in a more lexically incremental way. Konopka et al (2019) argued that the reason L2 speakers showed this effect is because they preferred to encode the overall message (i.e., conceptualisation) before they began formulating their utterance (i.e., selecting syntactic structure; selecting words). That is, message-level encoding and sentence-level encoding did not interact in speakers when they were speaking their L2. It is possible that children also show this separation between conceptualisation and formulation. Indeed, our study required very little message-level encoding as our sentences had minimal semantic content, and so it may have been easier for children to use a lexically incremental strategy in our study than in other studies where children had to describe an event (therefore requiring more message-level encoding than our study).

Konopka et al (2019) argue that L2 speakers engage in more of a structurally incremental planning strategy than a lexically incremental one because planning in L2 is more difficult than planning in L1 and it means they will have more of their sentence prepared before speaking. As a result, speakers do not have to engage in message-level and sentence-level encoding concurrently whilst speaking, and instead can focus on linguistic encoding processes like lexical retrieval, therefore limiting the amount of processing resources required during articulation. It is possible that children behave in a similar way to adults for this reason as well. Indeed, many studies suggest that pre-schoolers' errors during production may be due to limits in their available processing resources rather than to limits in their linguistic knowledge (e.g., Budd, Hanley & Griffiths, 2011; Charest & Johnston, 2011).

So far, we have proposed that pre-schoolers can use a lexically incremental planning strategy for easy sentences because they do not need to allocate processing resources to more complex message-level planning like they do for more complex transitive sentences. But how do children become adult-like and use a lexically incremental planning strategy for more complex sentences? It is possible that more experience with verbs and their associated thematic roles (e.g., agents, patients, etc) may make it easier for children to engage in concurrent message-level and sentence-level encoding. Indeed, more experience may automate the processes involved in message-level encoding, meaning less processing resources are required to generate thematic roles for a verb. This is something future studies should investigate, perhaps by manipulating the number of times a child names a picture with a particular verb or not.

6.2.2. Implications for models of morpho-phonological encoding

When naming multiple objects, adult speakers tend to retrieve word forms incrementally (e.g., Meyer et al., 1998). That is, they will primarily focus on retrieving the word form for the first object and begin focussing on retrieving the next word form after they begin naming the first. Eye-tracking studies have also shown that adults flexibly engage in parallel planning, whereby they begin processing of the next-named object whilst they are still retrieving the first object name. Such parallel planning ensures that a speaker will produce their utterance fluently.

Our results from Study 2 found evidence of incremental planning in pre-schoolers: like adults, pre-schoolers tended to mention each picture in the order they fixated on them, regardless of whether they were planning a simple sentence (as in Experiment 1b) or a more complex sentence (as in Experiment 2b). These findings suggest that, by the pre-school years, the overall processes involved in lexical retrieval during multi-word utterances are similar in children and adults. Considering that children tend to only begin speaking within their second year, this means that the ability to incrementally retrieve word forms occurs within the two years after they begin speaking. This then leads us to ask when do children develop the ability to incrementally retrieve word forms in multi word utterances? And how does this skill develop?

Throughout this thesis, we have assumed that incremental retrieval should be difficult, as it involves parallel planning: speakers need to concurrently plan upcoming chunks whilst speaking. Therefore, it should require processing resources. This perspective would therefore predict that the ability to incrementally retrieve word forms should be a later acquired skill which would develop as children's capacity to manage cognitive load develops (e.g., as children develop greater working memory capacity). However, it is possible that the mechanism underlying incremental retrieval develops at the same time as children begin producing their first words. In fact, this could explain why children begin producing multi-word utterances shortly after they begin producing their first words in their second year. Incremental retrieval in its simplest form requires the speaker to plan in small chunks. This would be beneficial for young children, as they would only have to plan one word at a time, which would therefore minimise the number of processing resources required during multi-word production, with the implication being that children will only begin retrieving each word as they begin producing the previous word. That is, when producing a sentence like "*the cat and the dog*," children will only begin retrieval of the second content word dog *after* they begin producing the first content word cat. Indeed, children have a much slower speech rate than adults, and it is possible that a reason for this is that they are planning each word one at a time.

According to this perspective then, what needs to develop is the ability to parallel plan during production. That is, children need to develop the ability to begin planning of the next named word whilst they are still planning or speaking the first. Our results from Study 2 indicate that parallel processing may be something that pre-schoolers can do, but they are much more inefficient in their ability to do so. Indeed, we found that children tended to shift their gaze to the second named picture before they began speaking in both experiments. Despite this, their gaze durations for the second named picture were much longer than for the first named picture, indicating that it took longer for them to retrieve the second word whilst they were speaking. However, it should be noted that our results are not conclusive, as we did not manipulate whether children could complete in extra-foveal processing or not (as Morgan & Meyer, 2005 have done previously) and future studies should investigate pre-schoolers' abilities to plan in parallel (i.e., whether they can process extra-foveal objects, and if they can, whether they are particularly bad at processing extra-foveally when they are speaking).

If parallel planning during incremental retrieval is something that develops later, how does this ability develop? It is possible that word form retrieval uses most of the child's processing resources, meaning that they do not have enough resources available for lexical retrieval of the next word. Therefore, more experience accessing and retrieving words may automate the process of lexical retrieval, which will then limit the processing resources required to retrieve a word form. As a result, children would be able to begin parallel planning. This account would therefore predict to find evidence of parallel planning for more frequent, earlier acquired words than less frequent, later acquired words. However, parallel planning may also develop alongside more domain-general mechanisms, such as the development of executive functions or greater working memory. Future studies should investigate these possibilities.

6.2.3. Implications for models of referential communication

During conversation, adult speakers tend to adopt the same perspective as their partner, and this is reflected in their referential choice (i.e., referential alignment or referential maintenance). This tendency has been explained in terms of perspective-taking (audience design mechanisms), but memory-based and priming-based mechanisms also play an important role in adults' referential maintenance (Branigan et al., 2011).

Our results from both Study 3 and Study 4 suggest that pre-schoolers can also adopt their partner's perspective and use this perspective in their own language use. These results suggest that pre-schoolers' abilities to flexibly adopt their partner's perspective may be less restricted than previously thought (e.g., Koymen et al, 2014). This particular finding has important theoretical and methodological implications. From a methodological standpoint, our results suggest that it is important to study pre-schoolers' perspective-taking abilities using interactive methods that mimic conversation as they seem to show greater abilities than when testing using other non-interactive methods (e.g., Deak et al, 2000). Theoretically, these results suggest that interactive contexts may be important for scaffolding perspective-taking behaviours that influence referential choices. As a result, future studies investigating referential communication should use interactive paradigms.

We can also draw some inferences about the mechanisms that might underlie pre-schoolers referential alignment. Our results indicate that priming-based mechanisms play an important role in pre-schoolers' abilities to adopt and reuse their partner's perspective in their own language use. In Study 3, pre-schoolers' tendency to reuse their partner's perspective decreased after they had previously named a referent with their own favoured perspective (regardless of whether their partner was aware of their favoured perspective or not), indicating that there was competition between their partner's perspective and their own favoured (and default) perspective in their lexicon. Furthermore, this tendency decreased over time in Study 4 indicating that activation for their partner's perspective decayed over time, resulting in them reverting to use their own favoured perspective (and therefore name).

From a cognitive load perspective, priming-based mechanisms are low-level resource-free and automatic, and allow for pre-schoolers to take their partner's perspective without using more complex perspective-taking mechanisms. Furthermore, they may facilitate lemma selection (including perspective selection) because their partner's perspective has greater activation than their own favoured, default perspective. Pre-schoolers therefore require fewer processing resources to select a word to describe a referent when reusing their partner's word choice. Thus, pre-schoolers' adoption of their partner's perspective is a way of reducing cognitive load during production, but the resulting behaviour is one that facilitates conversation and enhances mutual understanding.

However, priming mechanisms alone do not necessarily promote mutual understanding during conversation, and it is important that children develop the skills necessary to ensure that a conversation is running smoothly. Indeed, Garrod and Clark (1993) showed that whilst younger school-age children (7-8-year-olds) would align with their partner, their alignment did not result in successful communication nor mutual understanding. Adults can use perspective-taking mechanisms to ensure a conversation is running smoothly. But how do children develop these skills? It is possible that priming mechanisms support the development of more complex communicative perspective-taking skills, such as audience design. That is, pre-schoolers adopt their partner's perspective using automatic, low-level, resource-free priming-based mechanisms. The more experience a child has adopting their partner's perspective may act as a learning signal whereby children begin to understand that their partner has a perspective that is different from their own. This learning signal will then aid in the development of audience design mechanisms, where the child will start taking their partner's perspective into account. Future studies could investigate this idea further by measuring whether children's alignment effect scores (i.e., the proportion of trials whereby children adopt their partner's perspective over the total number of trials) is predicted by their scores in a task designed to measure their perspective-taking skills (e.g., theory of mind tasks).

In this thesis, we have considered alignment (or maintaining a referential pact) from communicative perspective, specifically focusing on how alignment aids communication between two partners and the mechanisms driving this behaviour. However, we have not considered how referential pacts fit in the wider social world. It is possible that maintaining a referential pact is an example of following a social norm. Under this account, referential pacts carry normative force. That is, once a partner has referred to an object in a particular way (e.g., *pony*), that is how that particular object should continue to be referred to in this particular context, regardless of partner. This could potentially have wider impacts in a community, resulting in community preferences for a specific word, and indeed, adults can extrapolate community-level lexical choices from individual encounters with a partner from a different community (Tobar, Rabagliati & Branigan, 2021).

By the age of 2 and 3, children are able to pick up social norms quickly and easily (see Rakoczy & Schmidt, 2013 for review). Furthermore, they often protest any violations of social norms, for example, when playing a novel board game (Rakoczy, Warneken, & Tomasello, 2008) or during joint pretending (Racoczy, 2008). There is evidence that children treat referential pacts as having normative force and protest violations to them. For example, Matthews et al (2010) found that pre-schoolers were not only slower to identify referents when a partner broke a referential pact, but a third of children also protested the use of other terms – regardless of whether the partner was the one who developed the referential pact (the same partner condition) or not (the different partner condition). Such protestations could reflect children treating referential pacts as a social norm; they are protesting the violation of social norms to the original partner, and informing the different partner of the social norm (that is – in this context, this object is called x). Overall, this suggests that there are many factors influencing children's use of referential pacts and future research could investigate the interplay between these factors in children's maintenance of referential pacts.

6.2.4. Implications for models of the developing language production system

Taking all of our findings together, our results suggest that the mechanisms underlying pre-schoolers' production are similar to adults'. We found that, like adults, pre-schoolers can flexibly engage in different sentence planning strategies; like adults, pre-schoolers can incrementally retrieve word forms during production; and they can flexibly adopt their partner's perspective during interaction like previous studies have found in adults (e.g., Branigan et al., 2011).

However, our results also indicate that pre-schoolers' production system is less efficient than adults': pre-schoolers' starting points were more influenced by our cue manipulation in Experiments 1a and 2a compared to adults, indicating that they used bottom-up factors to facilitate lexical retrieval (whereas adults' starting points were more affected by their left bias, despite their first fixations being influenced by the cue); pre-schoolers had longer gaze durations for the second named picture than the first named picture in Experiments 1b and 2b, indicating that it took longer for them to retrieve the second word whilst they were speaking; furthermore, pre-schoolers relied more on low-level, automatic and resource-free priming mechanisms when adopting their partner's perspective, whereas adult studies have shown that speakers can also use more complex communicative perspective-taking mechanisms when adopting their partner's perspective (e.g., Brennan & Clark, 1996; Branigan et al., 2011).

Why might the developing production system be less efficient than the fully developed adult system? On the one hand, children may have limited processing resources, which affects their abilities during production. Production involves the coordination of a series of complex cognitive processes, and so children may not have the processing resources available to them to complete these processes as efficiently as adults. As a result, pre-schoolers' performance during production is much more inefficient, which results in them taking longer to retrieve words whilst speaking, or they may be susceptible to influences of context that facilitate retrieval processes. For example, they will immediately reuse a perspective their partner has previously used, but this effect will decay over time as the activation of their partner's perspective decreases (Experiment 8); or they will begin their utterance with a cued object name because the cue directs attention to one object which facilitates retrieval of that name (Experiment 1a & Experiment 2a). Indeed, adults also show similar patterns when they are under cognitive load, as shown in dual-task studies. Cognitive load causes adults to produce more errors and disfluencies, or produce shorter, less complex utterances, indicating that production becomes less efficient (Hartsuiker & Barkhuysen, 2006). Furthermore, adults' performance in the secondary task is often worse, indicating that production requires processing resources (e.g., Ferreira & Pashler, 2002). Theoretical accounts of dual-task performance have often been attributed to either a structural bottleneck (Pashler, 1994), whereby the processes of the secondary task can only begin once the corresponding stages of the primary task (i.e., speaking) are completed which results in worse performance in the secondary task, or a limited capacity central processor (Kahneman, 1973), whereby the speaker only has a limited capacity to allocate to each task, resulting in poor overall performance in both tasks. We can relate these two accounts to pre-schoolers' production processes. Let's focus on morpho-phonological encoding and imagine that a child is going to name a cat and a dog. This task requires them to retrieve two word forms, which means they have to go through the processes of lexical retrieval twice. According to a version of the structural bottleneck account, the child can only begin central processing of the second word once central processing of the first word has ended. This would therefore predict that children will not be able to parallel plan. Similarly, the limited capacity account would assume that the child can only allocate a certain amount of processing resources to each word. This would mean that children *could* plan in parallel, but their overall production performance would be affected (e.g., longer retrieval times when planning in parallel). Our results cannot distinguish between these accounts as we do not have clear evidence as to whether pre-schoolers were able to plan in parallel (in Study 2). However, more recent work using a dual-task paradigm in older children (7-14-year-olds) would indicate that children's inefficient production processes may be due to capacity limits (Sasisekaran & Donohue, 2015).

Regardless of the reason underlying pre-schoolers' inefficient production, our results show that pre-schoolers' performance is similar to adults under pressure or cognitive load. One of the main findings in the adult production literature is that the fully developed production system is flexible. For example, adults will use a lexically incremental planning strategy when they are struggling to encode an event but will use a structurally incremental approach for simple events (van de Velde et al., 2014); adults' planning scope is flexible and can range from a word (e.g., a dog) to a phrase (e.g., a dog and a cat) and this can be affected by time pressures or cognitive load (Ferriera & Swets, 2002; Swets et al., 2013; Wagner et al., 2010; Wheeldon et al., 2013). Adults can flexibly engage in parallel planning or not (e.g., Malpass & Meyer, 2010); and finally, adults can also make use of different mechanisms during referential communication depending on who they are speaking to and the context they are in (e.g., they will use audience design mechanisms when interacting with a partner who may not understand their favoured (and default) perspective; Branigan et al., 2011). Adults' flexibility during production has often been explained in terms of ease of retrieval. That is, adults will use a specific production strategy or process that will make production (e.g., selection or retrieval) easier or faster. Indeed, the goal of production is often to speak as quickly as possible (for example during turn-taking in conversation) and so it is important that speakers have a range of production strategies to hand so they can use them. It is unclear from our studies whether children can flexibly switch between different production strategies. Future studies could investigate whether the developing production system is as flexible as the fully developed production system.

We now consider the developmental process of the production system. Our results indicate that by the pre-school years, the production system is almost adult-like. Considering that children only begin producing their first words during their second year, our results therefore imply that the mechanisms underlying production develops between the second and the fourth year. How do these mechanisms develop in such a short time? And what might be the developmental processes involved in these?

A theory of automaticity has been used to explain the development of production processes in second language acquisition (Segalowitz, 2003), but has rarely been considered to explain the development of production processes during first language acquisition (see- Logan, 1997 for a review who relates automaticity to reading development). This theory posits that with more practice and experience, tasks can be performed without consciously devoting mental effort to them, resulting in these processes taking less time to complete (Moors & De Houwer, 2006). It is possible that more practice producing automates the processes of production and strengthens linguistic representations. This in turn makes the whole production process faster and would allow for less processing resources being required during production. An instance-based theory (Logan, 1988; 1990; 1992) posits that the learning mechanisms underlying automaticity is episodic memory whereby experience with a specific instance lays down a memory trace which can be retrieved the next time an individual completes a task. Therefore, more experience and practice with an instance allows for more automatic performance. Indeed, instance-based theory can be linked to theories of production development whereby infants learn to produce certain syntactic frames with certain lexical items (i.e., item-based learning; Tomasello, 2009). This theory would then predict that children would be better (e.g., faster; show more evidence of parallel planning) for specific frame, or words they have had more practice producing. Our results from Study 2 (Experiment 1b), whereby children produced early acquired words (e.g., *cat*, *dog;* which they would have had more practice producing) in simple sentence frames, suggested that children may have been engaging in parallel planning. It is possible that because children have had more practice producing these words, they had developed the ability to parallel plan. But for more complex, later acquired words, they may not have yet developed the ability to parallel plan.

Alternatively, increases in verbal working memory capacity through more exposure to linguistic experience could result in the adult-like performance we see in pre-schoolers (MacDonald & Christiansen, 2002). According to this account, increases in working memory allows for increases to the attentional resources that can be split between tasks, meaning that children can behave in an adult-like way. The results from our studies indicate that pre-schoolers' production mechanisms are less efficient than adults, and this could be due to limited working memory capacities. Future studies could investigate whether these learning mechanisms aid in production development.

6.3. Methodological Implications of this thesis

Finally, this thesis has shown it is possible to investigate the developing production system using more precise methods, such as eye-tracking, rather than wholly relying on speech errors. By using more robust and precise methods, research on the developing production system may make important theoretical advancements, recapitulating the adult literature thirty years ago.

However, it is important to replicate and reproduce these findings in future research, especially in response to the replicability crisis (Open Science Collaboration, 2015; see also Gilbert, King, Pettigrew & Wilson, 2016). The replicability crisis may even be exacerbated in developmental research as developmental experiments often have lower statistical power due to having smaller sample sizes and less experimental trials than adult experiments whilst also having more stubborn who may not follow study instructions as well as adults would (Frank et al., 2017). Furthermore, there are systematic differences between developmental labs which could affect a study's findings. Whilst we have reproduced our findings in this thesis by conducting multiple experiments per study, it is important other labs and researchers can also replicate our findings.

The open science movement is a response to the replicability crisis and encourages researchers to be more transparent and open about their research practices (whilst discouraging questionable research practices which can lead to improper statistical inferences, such as p-hacking), whilst also promoting labs to replicate and reproduce research (e.g., ManyBabies project in developmental research). There are many open science practices that researchers can easily adopt in practice (see Crüwell et al., 2019 for review) such as having their materials, data, and statistical analyses code publicly available, or by replicating their findings in multi-experiment papers. One important way to be more open in research is to preregister a study before running it. Preregistration forces the researcher to be more transparent about any analytical decisions that need to be made prior to any actual analysis whilst also ensuring the researcher cannot present exploratory results as confirmatory (Wagenmakers, Wetzels, Borsboom, van der Maas & Kievit, 2012). On a more practical level, preregistration ensures the researcher is clear about exactly how they will analyse the data before they have even run the study. This is especially useful for more novel studies using paradigms or methods that have not been used before. One regret I have for this thesis is not preregistering it as I believe it would have made the overall analysis – especially of the eye-tracking studies – much easier, especially as I was using a paradigm that had not been used in children before.

Conclusion

In this thesis, we aimed to explore the mechanisms underlying pre-schoolers' production. We focused on three aspects of production: syntactic planning, morpho-phonological encoding, and referential communication. We adapted paradigms and methods that have successfully been used to investigate the mechanisms underlying production in adults, and found that pre-schoolers can use bottom-up factors to aid sentence planning; they can retrieve word forms incrementally; and they can flexibly adopt their partner's perspective in their own language use.

Taken together, the results from our studies suggest that the underlying mechanisms of pre-schoolers' language production are strikingly similar in important ways to those that have been found in adults (using very similar paradigms). However, our pattern of results indicate that pre-schoolers' production abilities may be less efficient than adults, and as a result, their behaviour may be similar to that of adults under processing load.

Overall, our results provide novel insight into the mechanisms underlying pre-schoolers' online production. Future studies should investigate how these mechanisms develop. We have proposed that children become more automatic in their abilities to produce (through instance-based learning) or through changes in working memory capacity.

7. References

Abrahams, L., Hartsuiker, R. J., De Fruyt, F., & Bajo, M. T. (2019). Structural alignment and its prosocial effects in first and second languages. Acta Psychologica, 199, 102906. https://doi.org/10.1016/j.actpsy.2019.102906

Akhtar, N., & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. Developmental psychology, 33(6), 952.

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. Child Development, 67(2), 635-645.

Allum, P. H., & Wheeldon, L. (2009). Scope of lexical access in spoken sentence production: Implications for the conceptual–syntactic interface. Journal of Experimental Psychology: Learning, Memory, and Cognition, 35(5), 1240.

Allum, P. H., & Wheeldon, L. R. (2007). Planning scope in spoken sentence production: the role of grammatical units. Journal of Experimental Psychology: Learning, Memory, and Cognition, 33(4), 791.

Almor, A. (2008). Why does language interfere with vision-based tasks?. Experimental Psychology, 55(4), 260-268.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. Cognition, 73(3), 247-264.

Anderson, a H., Clark, a, & Mullin, J. (1991). Introducing information in dialogues: forms of introduction chosen by young speakers and the responses elicited from young listeners. Journal of Child Language, 18(3), 663–687. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1761619

Anderson, a H., Clark, a, & Mullin, J. (1994). Interactive communication between children:

learning how to make language work in dialogue. Journal of Child Language, 21(2), 439–463. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7929689

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of memory and language, 59(4), 390-412.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. Joint attention: Its origins and role in development, 131, 158.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language, 68(3), 255-278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. arXiv preprint arXiv:1506.04967.

Bates, E., Bretherton, I., & Snyder, L. S. (1991). From first words to grammar: Individual differences and dissociable mechanisms (Vol. 20). Cambridge University Press.

Bates, E., MacWhinney, B., Caselli, C., Devescovi, A., Natale, F., & Venza, V. (1984). A cross-linguistic study of the development of sentence interpretation strategies. Child development, 341-354.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). *Lme4: Linear mixed- effects models using "Eigein" and S4* (R package version 1.1-23). Retrieved from http://CRAN.R-project.org/package=lme4

Belke, E., & Meyer, A. S. (2007). Single and multiple object naming in healthy ageing. Language and Cognitive Processes, 22(8), 1178-1211.

Benedict, H. (1979). Early lexical development: Comprehension and production. Journal of child language, 6(2), 183-200.

Birch, S. A., & Bloom, P. (2002). Preschoolers are sensitive to the speaker's knowledge when learning proper names. Child development, 73(2), 434-444.

Bloom, L. (1973). One word at a time: The use of single word ut-terances before syntax. The Hague: Mouton.

Bloom, L. (1993). The transition from infancy to language: Acquiring the power of expression. Cambridge University Press.

Bloom, L. (1998). Language acquisition in its developmental context.

Bloom, P. (1990). Subjectless sentences in child language. Linguistic inquiry, 21(4), 491-504.

Bloom, P. (1990). Syntactic distinctions in child language. Journal of Child Language, 17(2), 343-355.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. Psychological review, 89(1), 1.

Bock, J. K. (1986). Syntactic persistence in language production. Cognitive psychology, 18(3), 355-387.

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. Cognition, 21(1), 47-67.

Bock, K. (1995). Sentence production: From mind to mouth.

Bock, K., & Levelt, W. J. (1994). Language production: Grammatical encoding (pp. 945-984). Academic Press.

Bock, K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. The interface of language, vision, and action: Eye movements and the visual world, 191-211.

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. Scientific reports, 5(1), 1-11.

Boiteau, T. W., Malone, P. S., Peters, S. A., & Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. Journal of experimental psychology: General, 143(1), 295.

Branigan, H. P., & McLean, J. F. (2016). What children learn from adults' utterances: An ephemeral lexical boost and persistent syntactic priming in adult–child dialogue. Journal of Memory and Language, 91, 141-157.

Branigan, H. P., McLean, J. F., & Jones, M. (2005, April). A blue cat or a cat that is blue? Evidence for abstract syntax in young children's noun phrases. In Proceedings of the 29th annual Boston University conference on language development (pp. 109-121). Cascadilla Press.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. Cognition, 75(2), B13-B25.

Branigan, H. P., Pickering, M. J., McLean, J. F., & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. Cognition, 104(2), 163–197. https://doi.org/10.1016/j.cognition.2006.05.006

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: evidence from dialogs with humans and computers. Cognition, 121(1), 41–57. https://doi.org/10.1016/j.cognition.2011.05.011

Branigan, H. P., Tosi, A., & Gillespie-Smith, K. (2016). Spontaneous Lexical Alignment in Children With an Autistic Spectrum Disorder and Their Typically Developing Peers. Journal of Experimental Psychology: Learning, Memory, and Cognition, 42(5), 1–11. https://doi.org/10.1037/xlm0000272

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. Journal of experimental psychology: Learning, memory, and cognition, 22(6), 1482.

Brough, J., Branigan, H., Gambi, C., & Rabagliati, H. (2018). Incremental Strategies in

Children’s Language Production. In AMLaP 2018. Berlin, Germany.

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. Psychonomic bulletin & review, 16(5), 893-900.

Brown-Schmidt, S., & Konopka, A. E. (2015). Processes of incremental message planning during conversation. Psychonomic bulletin & review, 22(3), 833-843.

Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. Journal of Memory and Language, 54(4), 592-609.

Budd, M. J., Hanley, J. R., & Griffiths, Y. (2011). Simulating children’s retrieval errors in picture-naming: A test of semantic/phonological model of speech production. Journal of Memory and Language, 64(1), 74-87.

Bunger, A., Trueswell, J. C., & Papafragou, A. (2012). The relation between event apprehension and utterance formulation in children: Evidence from linguistic omissions. Cognition, 122(2), 135-149.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. Journal of statistical software, 80(1), 1-28.

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, i-174.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. Psychological review, 113(2), 234.

Charest, M., & Johnston, J. R. (2011). Processing Load in Children's Language Production: A Clinically Oriented Review of Research. Canadian Journal of Speech-Language Pathology & Audiology, 35(1).

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. In Journal of Personality and Social Psychology (Vol. 76). https://doi.org/10.1037/0022-3514.76.6.893

Chatterjee, A. (2001). Language and space: Some interactions. Trends in cognitive sciences, 5(2), 55-61.

Chatterjee, A., Southwood, M. H., & Basilico, D. (1999). Verbs, events and spatial representations. Neuropsychologia, 37(4), 395-402.

Chokron, S., & De Agostini, M. (2000). Reading habits influence aesthetic preference. Cognitive Brain Research, 10(1-2), 45-49.

Christman, S., & Pinger, K. (1997). Lateral biases in aesthetic preferences: Pictorial dimensions and neural mechanisms. Laterality: Asymmetries of Body, Brain and Cognition, 2(2), 155-175.

Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. Psychometrika, 78(4), 685-709.

Clark, E. V. (1973). Non-linguistic strategies and the acquisition of word meanings. Cognition, 2(2), 161-182.

Clark, E. V. (1997). Conceptual perspective and lexical choice in acquisition. Cognition, 64(1), 1–37. https://doi.org/10.1016/S0010-0277(97)00010-3

Clark, E. V. (2007). Young children's uptake of new words in conversation. Language in Society, 36(2), 157-182.

Clark, E. V. (2014). Pragmatics in acquisition. Journal of Child Language, 41(S1), 105-116.

Clark, H. H. (1992). Arenas of language use. University of Chicago Press.

Clark, H. H. (1996). Using language. Cambridge university press.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K.

Joshe, B. Webber, & I. A. Sag (Eds.), Elements of discourse understand- ing. Cambridge: Cambridge University Press.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. Cognitive psychology, 37(3), 201-242.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. Cognition, 22, 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Coco, M. I., & Keller, F. (2014). Classification of visual and linguistic tasks using eye-movement features. Journal of vision, 14(3), 11-11.

Coco, M. I., & Keller, F. (2015). Integrating mechanisms of visual guidance in naturalistic language production. Cognitive processing, 16(2), 131-150.

Coco, M. I., Malcolm, G. L., & Keller, F. (2014). The interplay of bottom-up and top-down mechanisms in visual guidance during object naming. The Quarterly Journal of Experimental Psychology, 67(6), 1096-1120.

Cook, A. E., & Meyer, A. S. (2008). Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(4), 886.

Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it. Cognition, 175, 77-95.

Corps, R. E., Gambi, C., & Pickering, M. J. (2020). How do listeners time response articulation when answering questions? The role of speech rate. Journal of Experimental Psychology: Learning, Memory, and Cognition, 46(4), 781.

Corrigan, R. (1978). Language development as related to stage 6 object permanence development. Journal of Child Language, 5(2), 173-189.

Costa, A., & Caramazza, A. (2002). The production of noun phrases in English and Spanish:

Implications for the scope of phonological encoding in speech production. Journal of Memory and Language, 46(1), 178-198.

Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., ... & Schulte-Mecklenbeck, M. (2019). Seven easy steps to open science: An annotated reading list. Zeitschrift für Psychologie, 227(4), 237.

Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. Journal of Memory and Language, 57(2), 195-209.

Dapretto, M., & Bjork, E. L. (2000). The development of word retrieval abilities in the second year and its relation to early vocabulary growth. Child Development, 71(3), 635-648.

Davies, C., & Kreysa, H. (2018). Look before you speak: children's integration of visual information into informative referring expressions. Journal of Child Language, 45(5), 1116-1143.

de Cat, C. (2015). The cognitive underpinnings of referential abilities. In S.E. Allen & L. Serratrice (Eds.), The acquisition of reference (pp. 263–283). Amsterdam, Netherland: John Benjamins.

De Villiers, J. G., & De Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. Journal of psycholinguistic research, 2(3), 267-278.

Deák, G. O., & Maratsos, M. (1998). On having complex representations of things: Preschoolers use multiple words for objects and people. Developmental Psychology, 34(2), 224.

Deak, G. O., Yen, L., & Pettit, J. (2001). By any other name: when will preschoolers produce several labels for a referent?. Journal of Child Language, 28(3), 787-804.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. Psychological review, 93(3), 283.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. Journal of memory and language, 27(2), 124-142.

Dell, G. S., & O'Seadhgha, P. G. (1992). Stages of lexical access in language production. Cognition, 42, 287–314.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. Psychological review, 104(1), 123.

Dell, G. S., Oppenheim, G. M., & Kittredge, A. K. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. Language and cognitive processes, 23(4), 583-608.

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. Cognition, 11(2), 159–184. https://doi.org/10.1016/0010-0277(82)90024-5

Dickson, W. P. (1982). (1982). Two decades of referential communication research: A review and meta-analysis. In C. J. Brainerd & M. Pressley (Eds.), Verbal processes in children. (pp. 1–33). New York: Springer.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. Frontiers in psychology, 5, 781.

Diesendruck, G. (2005). The principles of conventionality and contrast in word learning: an empirical examination. Developmental Psychology, 41(3), 451.

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: a pragmatic account. Developmental psychology, 37(5), 630.

Dobel, C., Diesendruck, G., & Bölte, J. (2007). How writing system and age influence spatial representations of actions: A developmental, cross-linguistic study. Psychological science, 18(6), 487-491.

Doherty, M., & Perner, J. (1998). Metalinguistic awareness and theory of mind: Just two words for the same thing?. Cognitive development, 13(3), 279-305.

Dromi, E. (1987). Early lexical development. University Press.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. Journal of personality and social psychology, 87(3), 327.

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. Journal of experimental social psychology, 40(6), 760-768.

Fairs, A., Bögels, S., & Meyer, A. S. (2018). Dual-tasking with simple linguistic tasks: Evidence for serial processing. Acta psychologica, 191, 131-148.

Fargier, R., & Laganaro, M. (2016). Neurophysiological modulations of non-verbal and verbal dual-tasks interference during word planning. PloS one, 11(12), e0168358.

Fay, D., & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. Linguistic inquiry, 8(3), 505-520.

Fenson, L., Dale, P., Reznick, S., Bates, E., Thal, D., Reilly, J.,& Harthung, J. (1990). MacArthur Communicative Devel-opment Inventories: Technical manual. San Diego: San Di-ego State University

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. Psychological science, 9(3), 228-231.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. Journal of Memory and Language, 30(2), 210-233.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. Journal of Memory and Language, 46(1), 57-84.

Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. Journal of memory and language, 35(5), 724-755.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(6), 1187.

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions?. Cognition, 96(3), 263-284.

Foltz, A., Gaspers, J., Meyer, C., Thiele, K., Cimiano, P., & Stenneken, P. (2015). Temporal effects of alignment in text-based, task-oriented discourse. Discourse Processes, 52(8), 609-641.

Ford, M., & Holmes, V. M. (1978). Planning units and syntax in sentence production. Cognition, 6(1), 35-53.

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., ... & Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. Infancy, 22(4), 421-435.

Friedrich, M., & Friederici, A. D. (2005). Phonotactic Knowledge and Lexical-Semantic Processing in One-year-olds: Brain Responses to Words and Nonsense Words in Picture Contexts. Journal of Cognitive Neuroscience, 17, 1785–1802.

Fromkin, V. A. (1971). The Non-Anomalous Nature of Anomalous Utterances. Language, 27-52.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. Psychological science, 23(8), 931-939.

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: effects of speakers' assumptions about what others know. Journal of Personality and Social Psychology, 62(3), 378–391. https://doi.org/10.1037/0022-3514.62.3.378

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee?. Journal of Memory and Language, 62(1), 35-51.

Garrett, M. (1980). Levels of processing in sentence production. In Language production Vol. 1: Speech and talk(pp. 177-220). Academic Press.

Garrett, M. F. (1975). The analysis of sentence production. In Psychology of learning and motivation (Vol. 9, pp. 133-177). Academic Press.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. Cognition, 27(2), 181-218.

Garrod, S., & Clark, A. (1993). Language and Cognitive Processes The development of dialogue co-ordination skills in schoolchildren The Development of Dialogue Co-ordination Skills in Schoolchildren. Language and Cognitive Processes, 8(February 2013), 37–41. https://doi.org/10.1080/01690969308406950

Gelman, S. A. (2009). Learning from others: Children's construction of concepts. Annual review of psychology, 60, 115-140.

Gentner, D. (1983). Why nouns are learned before verbs: Lin-guistic relativity versus natural partitioning. In S. Kuczaj(Ed.), Language development: Vol. 2. Language, thought, and culture (pp. 301–334). Hillsdale, NJ: Erlbaum

Gerken, L. (1996). Prosodic structure in young children's language production. Language, 683-712.

Gerken, L., & McIntosh, B. J. (1993). Interplay of function morphemes and prosody in early language. Developmental psychology, 29(3), 448.

Gershkoff-Stowe, L. (2001). The course of children's naming errors in early word learning. Journal of Cognition and Development, 2(2), 131-155.

Gershkoff-Stowe, L. (2002). Object naming, vocabulary growth, and the development of word retrieval abilities. Journal of Memory and Language, 46(4), 665-687.

Gershkoff-Stowe, L., & Smith, L. B. (1997). A curvilinear trend in naming errors as a function of early vocabulary growth. Cognitive Psychology, 34(1), 37-71.

Gershkoff-Stowe, L., Connell, B., & Smith, L. (2006). Priming overgeneralizations in two- and four-year-old children. Journal of Child Language, 33(3), 461-486.

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on “estimating the reproducibility of psychological science”.Science,351, 1036.

Giles, H., Coupland, J., & Coupland, N. (1991). Contexts of accommodation: Developments in applied sociolinguistics. In Contexts of accommodation (pp. 157–186). https://doi.org/10.1017/CBO9780511663673

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. Journal of memory and language, 57(4), 544-569.

Glucksberg, S., & Krauss, R. M. (1967). What do people say after they have learned how to talk? Studies of the development of referential communication. Merrill-Palmer Quarterly of Behavior and Development, 13(4), 309-316.

Glucksberg, S., Krauss, R. M., & Weisberg, R. (1966). Referential communication in nursery school children: Method and some preliminary findings. Journal of experimental child psychology, 3(4), 333-342.

Göbel, S. M., McCrink, K., Fischer, M. H., & Shaki, S. (2018). Observation of directional storybook reading influences young children’s counting direction. Journal of experimental child psychology, 166, 49-66.

Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. Journal of child language, 17(1), 171-183.

Goldin-Meadow, S., Seligman, M. E., & Gelman, R. (1976). Language in the two-year old. Cognition, 4(2), 189-202.

Goldman-Eisler, F. (1968). Psycholinguistics: Experiments in spontaneous speech.

Goldrick, M., Ferreira, V. S., & Miozzo, M. (Eds.). (2014). The Oxford Handbook of Language Production. Oxford University Press.

Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the intermodal preferential looking paradigm to study language acquisition: What have we learned?. Perspectives on Psychological Science, 8(3), 316-339.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: the case for a developmental lexical principles framework [*]. Journal of child language, 21(1), 125-155.

Gopnik, A., & Meltzoff, A. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. Child development, 1523-1531.

Gopnik, A., & Meltzoff, A. N. (1992). Categorization and naming: Basic-level sorting in eighteen-month-olds and its relation to language. Child Development, 63(5), 1091-1103.

Graham, S. a, Sedivy, J., & Khu, M. (2014). That's not what you said earlier: preschoolers expect partners to be referentially consistent. Journal of Child Language, 41(1), 34–50. https://doi.org/10.1017/S0305000912000530

Grice, H. P. (1975). " Logic and Conversation". In Syntax and Semantics. Leiden, The Netherlands: Brill. doi: https://doi.org/10.1163/9789004368811_003

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. Cognition, 82(1), B1-B14.

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. Psychonomic bulletin & review, 10(3), 603-609.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. Psychological science, 11(4), 274-279.

Griffin, Z. M., & Oppenheimer, D. M. (2006). Speakers gaze at objects while preparing intentionally inaccurate labels for them. Journal of Experimental Psychology: Learning, Memory, and Cognition, 32(4), 943.

Griffin, Z. M., & Spieler, D. H. (2006). Observing the what and when of language production for different age groups by monitoring speakers' eye movements. Brain and Language, 99(3), 272-288.

Halliday, M. A. K. (1975). Learning how to mean. In Foundations of language development (pp. 239-265). Academic Press.

Hansson, K., Nettelbladt, U., & Nilholm, C. (2000) Contextual influence on the language production of children with speech/language impairment. International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists, 35(1), 31–47. https://doi.org/http://dx.doi.org/10.1080/1368282 00247232

Hart, B., & Risley, T. R. (1995). Meaningful differences in the everyday experience of young American children. Paul H Brookes Publishing.

Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject-verb agreement. Language and Cognitive Processes, 21(1-3), 181-204.

Henderson, J., & Ferreira, F. (2004). The interface of language, vision, and action: Eye movements and the visual world. Psychology Press.

Hirsh-Pasek, K., & Golinkoff, R. (1991). Language comprehension: A new look at some old. Biological and behavioral determinants for language development, 301.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), Methods for assessing children's syntax (pp. 105–124). The MIT Press.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., Yust, P.K., & Suma, K. (2015). The contribution of early communication quality to low-income children's language success. Psychological science, 26(7), 1071-1083.

Hoff, E. (2010). Context effects on young children's language use: The influence of conversational setting and partner. First Language, 30(3–4), 461–472. https://doi.org/10.1177/0142723710370525

Hopkins, Z. L., & Branigan, H. P. (2020). Children show selectively increased language imitation after experiencing ostracism. Developmental Psychology.

Hopkins, Z. L., Yuill, N., & Branigan, H. P. (2017). Inhibitory control and lexical alignment in children with an autism spectrum disorder. Journal of Child Psychology and Psychiatry, 58(10), 1155–1165. https://doi.org/10.1111/jcpp.12792

Hopkins, Z. L., Yuill, N., & Branigan, H. P. (2021). Autistic children's language imitation shows reduced sensitivity to ostracism. Journal of Autism and Developmental Disorders, 1-13.

Horton, W. S. (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. Language and cognitive processes, 22(7), 1114-1139.

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. Cognition, 96(2), 127-142.

Horton, W. S., & Gerrig, R. J. (2016). Revisiting the memory-based processing approach to common ground. Topics in Cognitive Science, 8(4), 780-795.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground?. Cognition, 59(1), 91-117.

Hwang, H., & Kaiser, E. (2009). The effects of lexical vs. perceptual primes on sentence production in Korean: an on-line investigation of event apprehension and sentence formulation. In Talk presented at the 22nd CUNY conference on sentence processing, Davis, CA.

Ibbotson, P., Lieven, E. V., & Tomasello, M. (2013). The attention-grammar interface: eye-gaze cues structural choice in children and adults. Cognitive Linguistics, 24(3), 457-481.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. Cognition, 92(1-2), 101-144.

Ingram, D. (1978). Sensorimotor intelligence and languagedevelopment. In A. Lock (Ed.), Action, gesture, and sym-bol. New York: Academic Press

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. Journal of experimental psychology: general, 116(1), 26.

Jaeger, J. J. (1992). 'Not by the chair of my hinny hin hin': Some general properties of slips of the tongue in young children. Journal of Child Language, 19(2), 335-366.

Jaeger, J. J. (2005). Kids' slips: What young children's slips of the tongue reveal about language development. Psychology Press.

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., & Stern, D. N. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. Monographs of the society for research in child development, i-149.

Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. Journal of experimental psychology: learning, Memory, and cognition, 20(4), 824.

Jongman, S. R., Roelofs, A., Scheper, A. R., & Meyer, A. S. (2017). Picture naming in typically developing and language-impaired children: the role of sustained attention. International journal of language & communication disorders, 52(3), 323-333.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. Psychological review, 99(1), 122.

Kahneman, D. (1973). Attention and effort (Vol. 1063, pp. 218-226). Englewood Cliffs, NJ: Prentice-Hall.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. Acta psychologica, 86(2-3), 199-225.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. Cognitive science, 11(2), 201-258.

Khu, M., Chambers, C. G., & Graham, S. A. (2020). Preschoolers Flexibly Shift Between Speakers' Perspectives During Real-Time Language Comprehension. Child Development, 91(3), e619–e634. https://doi.org/10.1111/CDEV.13270

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. Journal of Memory and Language, 66(1), 143-162.

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. Cognitive Psychology, 73, 1-40.

Konopka, A. E., Meyer, A., & Forest, T. A. (2018). Planning to speak in L1 and L2. Cognitive psychology, 102, 72-104.

Koymen, B., Schmerse, D., Lieven, E., & Tomasello, M. (2014). Young children create partner-specific referential pacts with peers. Developmental Psychology, 50(10), 2334–2342. https://doi.org/10.1037/a0037837

Krauss, R. M., & Glucksberg, S. (1969). The Development of Communication: Competence as a Function of Age. Child Development, 40(1), 255–266.

Kubose, T. T., Bock, K., Dell, G. S., Garnsey, S. M., Kramer, A. F., & Mayhugh, J. (2006). The effects of speech production and speech comprehension on simulated driving performance. Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 20(1), 43-63.

Kuchinsky, S. E., & Bock, K. (2010, March). From seeing to saying: Perceiving, planning, producing. Paper presented at the 23rd meeting of the CUNY human sentence processing conference, New York, NY.

Kuchinsky, S. E., Bock, K., & Irwin, D. E. (2011). Reversing the hands of time: Changing the mapping from seeing to saying. Journal of Experimental Psychology: Learning, Memory, and Cognition, 37(3), 748.

Kunar, M. A., Carter, R., Cohen, M., & Horowitz, T. S. (2008). Telephone conversation impairs sustained visual attention via a central bottleneck. Psychonomic bulletin & review, 15(6), 1135-1140.

Lakin, J. L., & Chartrand, T. L. (2003). USING NONCONSCIOUS BEHAVIORAL MIMICRY TO CREATE AFFILIATION AND RAPPORT. Psychological Science, 14(4), 334–339.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). Atheory of lexical access in speech production. Behavioural and Brain Sciences, 22, 1–75.

Levelt, W. J., & Meyer, A. S. (2000). Word for word: Multiple lexical access in speech production. European Journal of Cognitive Psychology, 12(4), 433-452.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. Behavioral and brain sciences, 22(1), 1-38.

Liebal, K., Carpenter, M., & Tomasello, M. (2010). Infants' use of shared experience in declarative pointing. Infancy, 15(5), 545-556.

Lifter, K., & Bloom, L. (1989). Object knowledge and the emergence of language. Infant Behavior and Development, 12(4), 395-423.

Lindsay, L., Gambi, C., & Rabagliati, H. (2019). Preschoolers optimize the timing of their conversational turns through flexible coordination of language comprehension and production. Psychological science, 30(4), 504-515.

Lindsay, L., Rabagliati, H., & Branigan, H. (2018). Catching Your Eye: Low-level perceptual cues influence preschoolers' sentence formulation. In Boston University Conference on Language Development. Boston, MA.

Lindsley, J. R. (1975). Producing simple utterances: How far ahead do we plan?. Cognitive Psychology, 7(1), 1-19.

Locke, J. L. (1986). The linguistic significance of babbling. In Precursors of early speech (pp. 143-160). Palgrave Macmillan, London.

Logan, G. D. (1988). Toward an instance theory of automatization. Psychological review, 95(4), 492.

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms?. Cognitive Psychology, 22(1), 1-35.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: a test of the instance theory of automaticity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18(5), 883.

Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. Reading & writing quarterly: Overcoming learning difficulties, 13(2), 123-146.

Maass, A., & Russo, A. (2003). Directional bias in the mental representation of spatial events: Nature or culture?. Psychological science, 14(4), 296-301.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996).

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. Word, 15(1), 19-44.

MacWhinney, B. (1977). Starting points. Language, 152-168.

Malpass, D., & Meyer, A. S. (2010). The time course of name retrieval during multiple-object naming: Evidence from extrafoveal-on-foveal effects. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(2), 523.

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. Journal of Experimental Psychology: Human Perception and Performance, 38(4), 843.

Mani, N., Durrant, S., & Floccia, C. (2012). Activation of phonological and semantic codes in toddlers. Journal of Memory and Language, 66(4), 612–622. https://doi.org/10.1016/j.jml.2012.03.003

Markman, E. M. (1989). Categorization and naming in children: Problems of induction. mit Press.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. Cognitive psychology, 20(2), 121-157.

Martin, R. C., Miller, M., & Vu, H. (2004). Working memory and sentence production: Evidence for a phrasal scope of planning at a lexical-semantic level. Cognitive Neuropsychology, 21, 625-644.

Matthews, D., Lieven, E., & Tomasello, M. (2007). How toddlers and preschoolers learn to uniquely identify referents for others: A training study. Child development, 78(6), 1744-1759.

Matthews, D., Lieven, E., & Tomasello, M. (2010). What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. Developmental psychology, 46(4), 749.

McDaniel, D., McKee, C., & Garrett, M. F. (2010). Children's sentence planning: Syntactic correlates of fluency variations. Journal of Child Language, 37(1), 59-94.

McKee, C., & Iwasaki, N. (2001). Lemma structure in language learning: Comments on representation and realization. LANGUAGE ACQUISITION AND LANGUAGE DISORDERS, 23, 125-146.

McKee, C., McDaniel, D., & Garrett, M. F. (2017). Children's performance abilities: language production. The handbook of psycholinguistics, 491-515.

McKee, C., McDaniel, D., & Snedeker, J. (1998). Relatives children say. Journal of Psycholinguistic research, 27(5), 573-596.

McKee, C., McDaniel, D., Garrett, M. F., Lozoraitis, C., & Mutterperl, M. S. (2013). Articulation rate: Effects of age, fluency, and syntactic structure. Revista de Logopedia, Foniatría y Audiología, 33(2), 55-63.

McShane, J. (1980). Learning to talk. Cambridge University Press.

Mervis, C. B., & Bertrand, J. (1994). Acquisition of the novel name–nameless category (N3C) principle. Child development, 65(6), 1646-1662.

Mervis, C. B., Golinkoff, R. M., & Bertrand, J. (1994). Two-year-olds readily learn multiple labels for the same basic-level category. Child Development, 65(4), 1163-1177.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. Journal of Memory and Language, 49(2), 201-213.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture–word interference experiments. Journal of memory and Language, 35(4), 477-496.

Meyer, A. S., & Lethaus, F. (2004). The use of eye tracking in studies of sentence generation. In J. Henderson & F. Ferreira. The interface of language, vision, and action: Eye movements and the visual world, 191-211.

Meyer, A. S., & Van der Meulen, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. Psychonomic Bulletin & Review, 7(2), 314-319.

Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. Psychonomic bulletin & review, 14(4), 710-716.

Meyer, A. S., Ouellet, M., & Häcker, C. (2008). Parallel processing of objects in a naming task. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(4), 982.

Meyer, A. S., Roelofs, A., & Levelt, W. J. (2003). Word length effects in object naming: The role of a response criterion. Journal of Memory and Language, 48(1), 131-147.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. (1998). Viewing and naming objects: Eye movements during noun phrase production. Cognition, 66(2), B25-B33.

Meyer, A., Wheeldon, L. R., & Konopka, A. (2012). Effects of speech rate and practice on the allocation of visual attention in multiple object naming. Frontiers in psychology, 3, 39.

Moll, H., Richter, N., Carpenter, M., & Tomasello, M. (2008). Fourteen-month-olds know what “we” have shared in a special way. Infancy, 13(1), 90-101.

Moors, A., & De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. Psychological bulletin, 132(2), 297.

Morgan, J. L., & Meyer, A. S. (2005). Processing of extrafoveal objects during multiple-object naming. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(3), 428.

Mortensen, L., Meyer, A. S., & Humphreys, G. W. (2008). Speech planning during multiple-object naming: Effects of ageing. Quarterly Journal of Experimental Psychology, 61(8), 1217-1238.

Murray, L., & Trevarthen, C. (1986). The infant's role in mother–infant communications. Journal of child language, 13(1), 15-29.

Myachykov, A., Garrod, S., & Scheepers, C. (2010). Perceptual priming of structural choice during English and Finnish sentence production. Language & cognition: State of the art, 54-72.

Myachykov, A., Tomlin, R. & Posner, M. (2005). Attention and empirical studies of grammar . , 22(2-4), 347-364. https://doi.org/10.1515/tlir.2005.22.2-4.347

Nachson, I., Argaman, E., & Luria, A. (1999). Effects of directional habits and handedness on aesthetic preference for left and right profiles. Journal of Cross-Cultural Psychology, 30(1), 106-114.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. Psychological Science, 13(4), 329-336.

Nelson, K. (1973). Structure and strategy in learning to talk. Monographs of the society for research in child development, 1-135.

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. Cognitive psychology, 58(2), 220-249.

Nilsen, E. S., & Graham, S. A. (2012). The Development of Preschoolers' Appreciation of Communicative Ambiguity. Child Development, 83(4), 1400–1415. https://doi.org/10.1111/j.1467-8624.2012.01762.x

Nilsen, E. S., Varghese, A., Xu, Z., & Fecica, A. (2015). Children with stronger executive functioning and fewer ADHD traits produce more effective referential statements. Cognitive Development, 36, 68-82.

Norbury, C. F. (2014). Sources of variation in developmental language disorders: evidence from eye-tracking studies of sentence production. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1634), 20120393.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716.

Oppermann, F., Jescheniak, J. D., & Schriefers, H. (2008). Conceptual coherence affects phonological activation of context objects during object naming. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(3), 587.

Ostashchenko, E., Deliens, G., Geelhand, P., Bertels, J., & Kissine, M. (2019). Referential processing in 3-and 5-year-old children is egocentrically anchored. Journal of Experimental Psychology: Learning, Memory, and Cognition, 45(8), 1387.

Over, H., & Carpenter, M. (2009a). Eighteen-month-old infants show increased helping following priming with affiliation: Research report. Psychological Science, 20(10), 1189–1193.

Over, H., & Carpenter, M. (2009b). Priming third-party ostracism increases affiliative imitation in children. Developmental Science, 12(3), F1-8. https://doi.org/10.1111/j.1467-7687.2008.00820.x

Papafragou, A., & Grigoroglou, M. (2019). The role of conceptualization during language production: evidence from event encoding. Language, Cognition and Neuroscience, 34(9), 1117-1128.

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. Psychological bulletin, 116(2), 220.

Pechmann, T., & Deutsch, W. (1982). The development of verbal and nonverbal devices for reference. Journal of experimental child psychology, 34(2), 330-341.

Perner, J., & Leekam, S. R. (1986). Belief and quantity: Three-year olds' adaptation to listener's knowledge. Journal of Child Language, 13(2), 305-315.

Perner, J., Stummer, S., Sprung, M., & Doherty, M. (2002). Theory of mind finds its Piagetian perspective: Why alternative naming comes with understanding belief. Cognitive development, 17(3-4), 1451-1472.

Petretic, P. A., & Tweney, R. D. (1977). Does comprehension precede production? The development of children's responses to telegraphic sentences of varying grammatical adequacy. Journal of Child Language, 4(2), 201-209.

Piaget, J. (1926). The language and thought of the child. New York, NY:Harcourt Press.

Piaget, J. (1959). Judgment and reasoning in the child. Paterson: Littlefield, Adams & Co.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. Journal of Memory and language, 39(4), 633-651.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. Behavioral and brain sciences, 27(2), 169-190.

Pine, J. M., Lieven, E. V., & Rowland, C. F. (1998). Comparing different models of the development of the English verb category.

Power, M. J. (1986). A technique for measuring processing load during speech production. Journal of Psycholinguistic Research, 15, 371-382.

Rabagliati, H., & Robertson, A. (2017). How do children learn to avoid referential ambiguity? Insights from eye-tracking. Journal of Memory and Language, 94, 15-27.

Rakoczy, H. (2008). Taking fiction seriously: young children understand the normative structure of joint pretence games. Developmental Psychology, 44(4), 1195.

Rakoczy, H., & Schmidt, M. F. (2013). The early ontogeny of social norms. Child

Development Perspectives, 7(1), 17-21.

Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: young children's awareness of the normative structure of games. Developmental psychology, 44(3), 875.

Rispoli, M., Hadley, P., & Holt, J. (2008). Stalls and revisions: A developmental perspective on sentence production.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. Cognition, 42(1-3), 107-142.

Roelofs, A. (2007). Attention and gaze control in picture naming, word reading, and word categorizing. Journal of Memory and Language, 57(2), 232-251.

Roelofs, A. (2008). Tracing attention and the activation flow in spoken word planning using eye movements. Journal of Experimental Psychology: Learning, Memory, and Cognition, 34(2), 353.

Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. Frontiers in psychology, 2, 307.

Rowland, C. F., Chang, F., Ambridge, B., Pine, J. M., & Lieven, E. V. (2012). The development of abstract syntax: Evidence from structural priming and the lexical boost. Cognition, 125(1), 49-63.

Sachs, J., & Devin, J. (1976). Young children's use of age appropriate speech styles in social interaction and role-playing. Journal of Child Language, 3, 81 98 ST-Young children's use of age appropriat.

Sasisekaran, J., & Donohue, C. (2016). Effects of concurrent task demands on language planning in fluent children and adults. Applied psycholinguistics, 37(6), 1311-1336.

Schneider, W. (2002). Memory Development in Childhood. In U. Goswami (Ed.), Blackwell Handbook of Childhood Cognitive Development (pp. 236–256). Oxford: Blackwell.

Schnur, T. T., Costa, A., & Caramazza, A. (2006). Planning at the phonological level during sentence production. Journal of psycholinguistic research, 35(2), 189-213.

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. Journal of memory and language, 29(1), 86-102.

Schriefers, H., Teruel, E., & Meinshausen, R. M. (1998). Producing simple sentences: Results from picture–word interference experiments. Journal of Memory and Language, 39(4), 609-632.

Segalowitz, N. (2003). Automaticity and Second Languages. In C. J. Doughty & M. H. Long (Eds.), The Handbook of Second Language Acquisition (pp. 382–408). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470756492.ch13

Shatz, M., & Gelman, R. (1973). The development of communication skills: modifications in the speech of young children as a function of listener. Monographs of the Society for Research in Child Development, 38(5), 1–38.

Shintel, H., & Keysar, B. (2007). You said it before and you'll say it again: Expectations of consistency in communication. Journal of Experimental Psychology: Learning, Memory, and Cognition, 33(2), 357.

Shipley, E. F., Smith, C. S., & Gleitman, L. R. (1969). A study in the acquisition of language: Free responses to commands. Language, 322-342.

Sjerps, M. J., & Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. Cognition, 136, 304-324.

Slevc, L. R. (2011). Saying what's on your mind: working memory effects on sentence production. Journal of experimental psychology: Learning, memory, and cognition, 37(6), 1503.

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. Cognition, 73(3), 205-246.

Smith, M., & Wheeldon, L. (2004). Horizontal information flow in spoken sentence production. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(3), 675.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. Cognitive psychology, 49(3), 238-299.

Sonnenschein, S., & Whitehurst, G. J. (1984). Developing referential communication: A hierarchy of skills. Child Development, 55(5), 1936-1945

Spalek, T. M., & Hammad, S. (2005). The left-to-right bias in inhibition of return is due to the direction of reading. Psychological Science, 16(1), 15-18.

Spieler, D. H., & Griffin, Z. M. (2006). The influence of age on the time course of word preparation in multiword utterances. Language and Cognitive Processes, 21(1-3), 291-321.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. Nature, 388(6640), 381-382.

Stemberger, J. P. (1989). Speech errors in early child language production. Journal of Memory and Language, 28(2), 164-188.

Strayer, D. L., & Drews, F. A. (2007). Cell-phone–induced driver distraction. Current Directions in Psychological Science, 16(3), 128-131.

Swets, B., Jacovina, M. E., & Gerrig, R. J. (2013). Effects of conversational pressures on speech planning. Discourse Processes, 50(1), 23-51.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. Psychological science, 13(5), 480-484.

Tobar-Henríquez, A., Rabagliati, H., & Branigan, H. P. (2020). Lexical entrainment reflects a stable individual trait: Implications for individual differences in language processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 46(6), 1091.

Tobar-Henríquez, A., Rabagliati, H., & Branigan, H. P. (2021). Speakers extrapolate community-level knowledge from individual linguistic encounters. Cognition, 210, 104602.

Tomasello, M. (1992). The social bases of language acquisition. Social development, 1(1), 67-87.

Tomasello, M. (2000). Do young children have adult syntactic competence? Cognition, 74(3), 209-253.

Tomasello, M. (2009). Constructing a language. Harvard university press.

Tomlin, R. S. (1995). Focal attention, voice, and word order: an experimental, cross-linguistic study. Word order in discourse, 30, 517-554.

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. Cognitive psychology, 23(4), 515-557.

Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. Cognition, 40(1-2), 21-81.

Valian, V., & Aubry, S. (2005). When opportunity knocks twice: two-year-olds' repetition of sentence subjects. Journal of Child Language, 32(3), 617-641.

Valian, V., Hoeffner, J., & Aubry, S. (1996). Young children's imitation of sentence subjects: Evidence of processing limitations. Developmental Psychology, 32(1), 153.

van Baaren, R. B., Holland, R. W., Steenaert, B., & van Knippenberg, A. (2003). Mimicry for money: Behavioral consequences of imitation. Journal of Experimental Social Psychology, 39(4), 393–398. https://doi.org/10.1016/S0022-1031(03)00014-3

Van de Velde, M., Meyer, A. S., & Konopka, A. E. (2014). Message formulation and structural assembly: Describing “easy” and “hard” events with preferred and dispreferred syntactic structures. Journal of Memory and Language, 71(1), 124-144.

van der Meulen, F. F., Meyer, A. S., & Levelt, W. J. (2001). Eye movements during the production of nouns and pronouns. Memory & Cognition, 29(3), 512-521.

van der Meulen, F., Hartl, H., & Tappe, H. (2003). Coordination of eye gaze and speech in sentence production. In Mediating between Concepts and Grammar (= Trends in Linguistics 152)

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. Psychonomic Bulletin and Review. https://doi.org/10.3758/BF03194105

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. Perspectives on Psychological Science, 7(6), 632-638.

Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(2), 423.

Wardlow, L. (2013). Individual differences in speakers’ perspective taking: The roles of executive control and working memory. Psychonomic bulletin & review, 20(4), 766-772.

Watson-Jones, R. E., Legare, C. H., Whitehouse, H., & Clegg, J. M. (2014). Task-specific effects of ostracism on imitative fidelity in early childhood ☆. https://doi.org/10.1016/j.evolhumbehav.2014.01.004

Watson-Jones, R. E., Whitehouse, H., & Legare, C. H. (2016). In-Group Ostracism Increases High-Fidelity Imitation in Early Childhood. Psychological Science. https://doi.org/10.1177/0956797615607205

Wheeldon, L. (2013). Producing spoken sentences: The scope of incremental planning. Speech production and perception: Speech planning and dynamics, 97-118.

Wheeldon, L. R., & Lahiri, A. (2002). The minimal unit of phonological encoding: prosodic or lexical word. Cognition, 85(2), B31-B41.

Wheeldon, L., Ohlson, N., Ashby, A., & Gator, S. (2013). Lexical availability and grammatical encoding scope during spoken sentence production. Quarterly Journal of Experimental Psychology, 66(8), 1653-1673.

Whitehurst, G. J., & Sonnenschein, S. (1981). The development of informative messages in referential communication: knowing when vs knowing how. In W. P. Dickson (Ed.), Children's communication skills (pp. 127–141). New York: Academic Press.

Wijnen, F. (1990). The development of sentence planning. Journal of Child Language, 17(3), 651-675.

Wijnen, F. (1992). Incidental word and sound errors in young speakers. Journal of Memory and Language, 31(6), 734-755.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. Journal of memory and language, 31(2), 183-194.

Williams, K. D., & Jarvis, B. (2006). Cyberball: a program for use in research on interpersonal ostracism and acceptance. Behavior Research Methods, 38(1), 174–180.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. Cognition, 13(1), 103-128.

Wühr, P., & Frings, C. (2008). A case for inhibition: Visual attention suppresses the processing of irrelevant objects. Journal of Experimental Psychology: General, 137(1), 116.

Wühr, P., & Waszak, F. (2003). Object-based attentional selection can modulate the Stroop effect. Memory & Cognition, 31(6), 983-994.

Yuan, S., & Fisher, C. (2009). “Really? She blicked the baby?” Two-year-olds learn combinatorial facts about verbs by listening. Psychological science, 20(5), 619-626.

8. Appendix: Experimental Materials used in Studies 1-4

8.1. Experimental Materials Used in Study 1 & 2

Table A1: Experimental items used in Studies 1 & 2. In Study 1, the cue appeared on

List	Item	Item 1 (Appeared on Left Side)	Item 2 (Appeared on Right Side)	Cue Side (For Study 1)
1	1	car	doll	right
	2	hand	cat	left
	3	fridge	stool	right
	4	lemon	necklace	left
	5	sheep	fish	right
	6	hoover	glasses	left
	7	trousers	pencil	right
	8	lamp	scarf	left
	9	shoulder	scissors	right
	10	hat	glove	left
	11	sofa	slipper	right
	12	spoon	nose	left
	13	bus	shirt	right
	14	ant	bag	left
	15	watch	bear	right
	16	monkey	tiger	left
	17	candle	onion	right
	18	belt	pear	left
	19	bee	clock	right
	20	horse	mouse	left
	21	balloon	arm	right
	22	spider	rabbit	left
	23	foot	sock	right
	24	cup	chair	left
	25	table	apple	right
	26	mirror	ruler	left
	27	comb	tape	right
	28	dog	eye	left
	29	plaster	carrot	right
	30	penguin	lion	left
2	1	doll	car	right
	2	cat	hand	left
	3	stool	fridge	right
	4	necklace	lemon	left

5	fish	sheep	right
6	glasses	hoover	left
7	pencil	trousers	right
8	scarf	lamp	left
9	scissors	shoulder	right
10	glove	hat	left
11	slipper	sofa	right
12	nose	spoon	left
13	shirt	bus	right
14	bag	ant	left
15	bear	watch	right
16	tiger	monkey	left
17	onion	candle	right
18	pear	belt	left
19	clock	bee	right
20	mouse	horse	left
21	arm	balloon	right
22	rabbit	spider	left
23	sock	foot	right
24	chair	cup	left
25	apple	table	right
26	ruler	mirror	left
27	tape	comb	right
28	eye	dog	left
29	carrot	plaster	right
30	lion	penguin	left

8.2. Experimental Materials Used in Study 3

Table A2. List of Experimental items used in Study 3.

Favoured Name	Disfavoured Name
Sofa	Couch
Daddy	Father
Mummy	Mother
T.V.	Telly
Cup	Mug
Horse	Pony
Lamp	Light
Spade	Shovel
Shoe (Exp 1-4)	Trainer (Exp 1-4)
Bobble (Exp 1-4)	Hairband (Exp 1-4)
Spots (Exp 1-4)	Dots (Exp 1-4)

Belly (Exp 1-4)	Tummy (Exp 1-4)
Toilet (Exp 5)	Loo (Exp 5)
Boat (Exp 5)	Ship (Exp 5)
Stairs (Exp 5)	Steps (Exp 5)
Rabbit (Exp 5)	Bunny (Exp 5)

8.3. Experimental Materials Used in Study 4

Table A3. Experimental Items used in Study 4

Favoured Name	Disfavoured Name
Sofa	Couch
Daddy	Father
Mummy	Mother
T.V.	Telly
Cup	Mug
Horse	Pony
Lamp	Light
Spade	Shovel
Shoe (Exp 1-4)	Trainer (Exp 1-4)
Bobble (Exp 1-4)	Hairband (Exp 1-4)
Spots (Exp 1-4)	Dots (Exp 1-4)
Belly (Exp 1-4)	Tummy (Exp 1-4)