Mobile Eye Tracking During Storybook Listening: Applying the Visual World Paradigm in the Investigation of Preschoolers' Online Discourse Processing

by

Abigail Toth

A thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science

Department of Linguistics
University of Alberta

Abstract

The current thesis assessed the application of the visual world eye tracking paradigm (VWP) as a tool for investigating online language processing in a naturalistic setting. Furthermore, it investigated how individual differences in vocabulary and working memory influence children's eye movements within in the visual scene with respect to spoken language input.

In the VWP an individual's eye movements are monitored as they receive spoken language input and view a visual scene. It works under the assumption that where the individual is looking indicates where their attention is focused and thus what they are processing at any given moment. The VWP has been used to investigate the online processing of various linguistic phenomena, including the processing of reference, particularly in the realm of online pronoun resolution. These studies have shown that upon hearing an ambiguous third person singular pronoun (i.e., 'he'), there is an increased proportion of looks to the subject of the preceding clause, suggesting that people are more likely to interpret the pronoun as co-referring with the subject (e.g., Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Järvikivi, Van Gompel, Hyönä, & Bertram 2005; Song & Fisher, 2005; 2007). For example when participants hear utterances such as 'The panda hit the parrot by the lake. He wanted to go home' while viewing a scene with both animals, they are more likely to look at the panda than the parrot when they hear '*he'*, indicating that there is a subject bias. However, almost all of the previous studies have looked at a series of isolated items (such as the example above), with only to referents in the visual scene (i.e., the subject and object). We do not know how the VWP works in more naturalistic settings, that is, when there is continuous linguistic discourse and multiple referents in the visual scene, as is the case in storybook listening.

Both children and adults listened to a five-minute long storybook while wearing eye-tracking glasses. The storybook contained multiple referring expressions, both names (e.g., ‘Bear’) and pronouns (e.g., ‘he’), and was designed to becoming increasingly more complex as it unfolded over time, beginning with just a single character and ending with a total of five characters. Using generalized additive mixed modeling (GAMMs), we analyzed the eye gaze data of 16 children and 12 adults with respect to the mention of 37 names and 10 pronouns embedded throughout the story. Overall we found that eye movements patterns differed for items (names and pronouns) that occurred during the first half of the story compared to items (names and pronouns) that occurred during the second half of the story, for both children and adults. Upon hearing a name during the first half of the story, both children and adults’ looks to the target referent increased. Upon hearing a pronoun during the first half of the story, children’s looks to the subject of the preceding clause increased. Adults, however, had the highest proportion of looks to the subject at the onset of the pronoun, suggesting they were able to use discourse cues to predict that the subject would be referred to. Hearing a name during the second half of the story had no influence on looks to the target referent, for both children and adults. Upon hearing a pronoun during the second half of the story, children’s looks to the subject of the preceding clause increased, however, this took much longer in the time course compared to pronouns that occurred during the first half of the story. Hearing a pronoun during the second half of the story had no influence on adults’ looks to the preceding subject. Furthermore, we found that children’s working memory (WM) capacity influenced their language mediated eye movements.

The findings of the current thesis demonstrate that there is not a uniform mapping between linguistic input and eye movements within the visual scene. It is likely that individuals only direct their eye gaze towards entities in the visual scene under particular language processing circumstances. As such, these findings call into question whether or not the visual world paradigm is an effective tool for investigating language processing in naturalistic settings. Further research is needed in order to better understand the relationship between eye gaze and spoken language processing in continuous discourse.

Preface

This thesis is an original work by Abigail Toth. The research project, of which this thesis is a part, received research ethics approval form the University of Alberta Research Ethics Board, Project name “Processing of Referring Expressions in Children with ASD”, Pro00075655, September 17th, 2018.

Acknowledgements

I would first like to extend my sincerest gratitude to Dr. Juhani Järvikivi, my primary supervisor. His confidence in my ability has helped me grow not only as an academic researcher, but also as an individual. He has challenged me to ask the big questions and given me the confidence to do so. I truly cannot express how big of an impact he has had on the trajectory of my life and I am forever grateful for the countless discussions we have had. I would also like to thank Dr. Monique Charest, my second supervisor. She was the perfect addition to the team, adding new perspectives, while also helping me remain focused on the specific research questions at hand. I am deeply thankful for the numerous hours she spent carefully reading over my thesis. I would also like to thank my external committee member, Elena Nicoladis, for her engaging discussion and feedback on my thesis. As well as, Evangelia Daskalaki for chairing my defense.

I am deeply thankful for my squad of fellow graduate students. Especially Kaidi, Figen and Filip, who have become lifelong friends to me. From statistical questions to wine and pizza nights, they have truly made the past two years a fun and enjoyable experience and I can only imagine how different my experience would have been without them. I would also like to thank Hannah for her presence in the lab and always being there to listen to my hand-coding sorrows. Furthermore, I would like to thank Romy and Kaleigh, the undergraduate research assistants who spent many hours helping me create and record the experimental stimuli.

I would also like to acknowledge the Social Sciences and Humanities Research Council of Canada (SSHRC), as well as the Words in the World project (a SSHRC partnered research training initiative) for funding my master's research.

Finally, I would like to thank my family and closest friends for always supporting me. They always keep me grounded and remind me to think ‘big picture’. My parents have always pushed me to fully apply myself and it is from them that I get my strong work ethic. I would specifically like to thank my dad for being my biggest supporter over the past two years, although there were challenges along the way, his level-headedness helped keep me focused and reminded me to not worry about things beyond my control. Finally, it would be out of character if I did not thank my dog Otto for being the best company during the writing process.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1. Eye Movements and Spoken Language Input

One advantageous method of investigating online language processing is to use the visual world eye tracking paradigm (VWP) (Allopenna, Magnuson, & Tanenhaus, 1998). In the VWP an individual's eye movements are monitored as they receive spoken language input and view a visual scene. It works under the assumption that where the individual is looking indicates where their attention is focused and thus what they are processing at any given moment.

In a seminal paper, Cooper (1974) reported that when people are simultaneously presented with spoken language and a visual scene, they naturally direct their eye gaze towards entities in the visual scene that are most semantically related to the meaning of the language currently being heard. For example, when participants heard phrases such as 'suddenly I noticed a hungry lion' they fixated on a lion present in the visual scene. And not only did participants fixate on critical objects, but the majority of these fixations occurred either while the corresponding word was spoken or within 200 ms after its offset. Cooper proposed that the relationship between spoken language and eye gaze fixations could be viewed as an active online process, such that spoken language gets interpreted incrementally in the context of the visual field, with eye movements being closely time-locked to the language. Furthermore, because participants often began fixating on targets prior to their mention, he argued that such a process may be anticipatory, in that participants appeared to be sensitive to contextual cues to upcoming words. Cooper suggested that the existence of this relationship could be used as a tool for investigating online perceptual and cognitive processes, such as online language comprehension. Adding to this argument, Hoffman and Subramaniam (1995) found that people are unable to orient their attention to one location while simultaneously executing a saccade (i.e., a rapid eye movement) to a different location, suggesting that attention guides saccadic eye movements. In other words, eye gaze is not directed towards related entities haphazardly, rather it is because that specific information is being processed that eye gaze is directed towards those entities. Taken together, these two studies suggest that where an individual is looking is an indication of what they are processing, such that one's attention is focused on the information they are processing, and that same attentional mechanism guides their saccadic eye movements.

In order to gain a better understanding of how the visual scene influences the rapid mental processes that accompany spoken language comprehension, Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) recorded participants' eye movements as they followed spoken instructions and manipulated real objects that were visible in front of them. They observed eye movements on a millisecond (ms) time scale and found that they were closely time-locked to the linguistic input, such that participants looked at objects almost immediately after hearing the relevant words. For example when participants heard instructions such as 'Touch the starred yellow square', they fixated on the correct target ~250 ms after hearing the word that uniquely identified that target. That is, if there was only one starred object, this fixation occurred ~250 ms after hearing 'starred', but if there were two starred yellow objects, this fixation occurred ~250 ms after hearing 'square'; the same was the case for more complex instructions. They also found that when the visual context contained objects with similar onsets, such as 'candy' and 'candle', it took participants 230 ms to fixate on the correct object, however, when all items had unique onsets it only took 145 ms. In the latter case they concluded that participants were able to identify the correct target before hearing the entire word, given that it takes ~200 ms to plan an eye movements (Matin, Shao, & Boff, 1993), further showing how closely eye movements to visual referents are time-locked with the linguistic input and that this mapping can be reflective of anticipatory processes. The findings from this study were in line with what Cooper had reported 20 years prior. Both studies were instrumental in the development of the visual world eye tracking paradigm as a tool for investigating online language processing.

Two primary linking hypotheses have been proposed to explain how eye movement patterns reflect the underlying cognitive processes involved in spoken language comprehension. The *joint representation of linguistic meaning and visual information* account (Altmann & Mirkovic, 2009) posits that the anticipated linguistic meaning and information from the visual scene interact with each other in such a way that they are indistinguishable from each other, such that updating occurs on the joint representation of the two. They note that in most VWP studies the visual scene is available prior to the onset of the spoken stimuli;re thus the speech input activates features of the visual scene that are already in the hearer's mind, which results in early eye movements to relevant objects either in a predictive or integrative manner (Altmann & Kamide, 2007, 2009). In contrast the *The Coordinated Interplay Account* (CIA: Knoeferle & Crocker, 2006; 2007) suggests a much more active influence of the visual scene on linguistic processing. This account posits that individuals search for objects and events within the visual scene in order to identify possible visual referents that are likely to be referred to. The visual scene then influences the comprehension process by confirming or altering the interpretation incrementally.

Although there are different accounts of how eye movement patterns reflect underlying cognitive processes, the relationship between the spoken language input and the visual scene likely depends on the task at hand. As such, the CIA hypothesis may better account for eye movement behavior in certain types of tasks, while the *joint representation of linguistic meaning and visual information* hypothesis may better account for eye movement behavior in different types of tasks. Thus, for tasks like those described above in Tanenhaus et al. (1995), where listeners have to manipulate the scene according to spoken instructions (i.e., action based VWP), the CIA account may provide a better explanation of the relationship between the language input and eye movement behavior, such that in order to successfully carry out the instructions, listeners must correctly identify objects and move them accordingly and thus, the active search for objects makes perfect sense. However, there are other language comprehension tasks that do not involve such an explicit 'find this object and do this thing' component. For example, when watching a television show it is unlikely that an individual actively searches for referents that are likely to be referred to, at least not to the same extent as they would if they were asked to move around different objects. In these more passive tasks, the *joint representation of linguistic meaning and visual information* account may provide a better explanation of eye movement behavior. Although evaluating the different linking hypotheses is not the aim of the current thesis, it is important to understand that eye movement patterns are assumed to be representative of the underlying cognitive processes involved in spoken language comprehension (based on the previous research described above) and that there are different accounts of how this mapping works. Below the key findings from previous VWP studies that relate to the purposes of the current thesis will be discussed.

1.2. VWP Garden-Path Processing

For a long time, the dominant view was that language processing is modular, involving an initial syntactic analysis stage, in which individuals only parsed the syntax. The mechanisms involved in the initial stage were thought to be separate from other linguistic, cognitive and perceptual systems, including those involved in semantic interpretation (Fodor, 1983). Primary evidence for this view came from reading studies showing that when people read sentences with temporary syntactic ambiguities, such as ‘Put the apple on the towel in the box’, their reading times slowed down when they came to ‘on the box’, suggesting that they had initially taken 'on the towel' to be the destination (or goal) of the verb *put*, rather than the (correct) noun phrase (NP) modifier interpretation, and thus they needed to revise their initial interpretation (for review, see Frazer, 1987). This is known as the garden-path effect, which has been used to support models that assumed that language comprehension proceeds from syntactic analysis (first) to semantic interpretation (second).

Tanenhaus and colleagues (1995) investigated whether information from the visual context immediately affects sentence processing. They used the same types of temporarily ambiguous sentences as those used in the reading studies (e.g., ‘Put the apple on the towel in the box’). Eye movements revealed that when only a single apple was present in the visual context (1-referent condition), participants temporarily considered the prepositional phrase (PP) ‘on the towel’ as the destination of the verb *put* and only adopted the correct interpretation upon hearing the subsequent PP ‘in the box’, just as in the reading studies. However, when a second apple (located on a towel) was present in the scene (2-referent condition), the garden-path effect disappeared, demonstrating that participants took the visual scene into account during the early stages of sentence processing in order to restrict the possible syntactic interpretations. These findings are in line with constraint-satisfaction theories of sentence processing (e.g., MacDonald, Perlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Trueswell, Tanenhaus, & Garnsey, 1994), which assume that listeners simultaneously take various sources of information into account to immediately constrain their structural analyses during comprehension. Not only do we take various sources of linguistic information into account, such as contextual information, subcategory frequency, semantics and prosodic cues (e.g., Altman, Garnham, & Dennis, 1992; Altman, Clifton, & Mitchell, 1998; Carlson, Clifton, & Fraizer, 2001; Clifton, Traxler, Williams, Mohammed, Morris, & Rayner, 2003; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Kjelgaard & Speer, 1999; Trueswell et al., 1994), but also information from other domains, such as the visual context. In contrast, these findings do not align with modular perspectives (e.g., Frazier, 1979, Frazier & Rayner, 1982; Rayner, Carlson, & Frazier, 1983), which argue that (even) when faced with ambiguity, listeners will build a single syntactic representation, adopting the simplest structure first, and only consider additional sources of information, during later stages of processing. Therefore, modular accounts argue that when processing sentences such as ‘Put the apple on the towel in the box’, listeners/readers will always first interpret 'on the towel' as the destination, regardless of the situational context. However, the findings from Tanenhaus et al. (1995) clearly demonstrate that listeners are able to use information from the visual scene, in order to avoid adopting the more simple (and incorrect) structure.

In addition to using the VWP to show that eye movements are closely time-locked to linguistic input (i.e., 'Touch the starred yellow square'), Tanenhaus and colleagues (1995) were also able to use this time-sensitive mapping to show that listeners take various sources into account during sentence processing, which in turn challenged more prevalent models of sentence structure processing. These findings demonstrate how VWP eye-tracking can provide novel insights into the nature (and time course) of information integration during language processing. In sum, not only can the VWP be used to investigate time-sensitive language processing but there can also be theoretical implications of the findings.

1.3. Further Applications of the VWP

It is clear that visual scene information can influence language processing. However, the time-locking of eye movements to visual referents with the auditory linguistic input can also be used to inspect the relative time course of various linguistic factors that influence online language processing. For example Altmann and Kamide (1999) investigated whether listeners are able to use verb (semantic) information to predict upcoming arguments. Participants listened to sentences such as 'the boy will *eat* the cake' or 'the boy will *move* the cake', while viewing a scene containing a cake and various other objects (which were inedible). The eye gaze data revealed that participants fixated on the target (i.e., the cake) prior to its onset in the *eat* condition; however, in the *move* condition, saccadic eye movements did not occur until the spoken onset of 'cake'. These findings were taken to suggest that participants used verb information (and real world knowledge of things that get eaten) to restrict the number of possible grammatical objects that could be referred to. This study used an adapted 'listen-and-look' version of the VWP, where linguistic input is interpreted in the presence of a visual scene, rather than the visual scene specifically being used to influence parsing choices (see section 1.4 for a more detailed discussion).

The VWP has been used to investigate language processing at various levels, including spoken word recognition (e.g., Allopenna, Magnuson & Tanenhaus, 1998; Huettig & McQueen, 2007; McMurray, Aslin, & Tanenhaus, 2008), reference processing (e.g., Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Järvikivi, Van Gompel, Hyönä, & Bertram, 2005; Kaiser, Runner, Sussman, & Tanenhaus, 2009), and sentence structure processing (e.g., Chambers, Tanenhaus, & Magnuson, 2004; Snedeker & Trueswell, 2004; Tanenhaus et al., 1995; Trueswell, Sekerina, Hill, & Logrip, 1999). It has also be used to investigate how a number of different linguistic factors can affect language processing, including acoustic properties of the speech signal (e.g., Dahan, Tanenhaus, & Chambers, 2002; Porretta, Tucker, & Järvikivi, 2016; Weber, Braun, & Crocker, 2006), semantics (i.e., verb properties) (e.g., Altmann and Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Pyykkönen, Matthews, & Järvikivi, 2010), as well as contextual and pragmatic properties (Chambers, Tanenhaus, Eberhard, Filip, & Carlson 2002; Engelhardt, Bailey & Ferreira, 2006; Hanna & Tanenhaus, 2004). It should also be noted that the VWP can also be used to investigate online language production (e.g., Gleitman, January, Nappa, & Trueswell, 2007; Griffin & Bock, 2000; Meyer, Sleiderink, & Levelt, 1998), however, for the purposes of the current thesis we will focus on comprehension.

In addition to the high temporal resolution offered by the VWP, there are also other practical benefits to adopting this methodology. Many offline measures require participants to make some sort of meta-linguistic judgement, which can encourage participants to use strategies that make use of explicit knowledge, and as such may not be representative of everyday language comprehension (Berends, Sprenger, & Brouwer, 2015). The VWP on the other hand, can be used to investigate language processing under relatively realistic conditions, and as such may do a better job at tapping into processes involved in more naturalistic language comprehension. Furthermore, the VWP does not require participants to read or carry out demanding tasks, nor does it require the comparison of performance on linguistically 'correct'/'plausible' stimuli to linguistically incorrect/implausible ones, as is often the case with other online methods, such as self-paced listening or reading. As such, the VWP can be used to investigate online language processing in children, which is potentially its greatest benefit, as it allows for the direct comparison of patterns displayed by adults versus children without the potential confounds introduced by response requirements. Using the VWP we can gain insights into when and how become ‘adultlike’ in their language processing.

In a seminal study, Trueswell, Sekerina, Hill and Logrip (1999) presented 5-year-old children with the same task as in Tanenhaus et al. (2015; i.e., ‘Put the apple on the towel in the box’) and found that they were equally as likely to look at the incorrect destination (i.e., the towel), in both the 1- and 2-referent conditions. This result suggested that children did not take the visual context into account while activating different potential structural analyses, unlike adults. The contrast in findings between children and adults highlights the importance of having methodologies that can be used with both groups. The VWP has the potential to provide a more valid comparison of child and adult language processing. One phenomenon that the VWP has been extensively used to investigate, with both children and adults, is the online processing of reference, especially pronoun resolution. The findings from these studies will be presented below.

1.4. VWP Pronoun Resolution

In order to identify different people, places and things we use different types of *referring expressions*, including both full noun phrases (NPs) (e.g., *Sarah*, *the tree*, *a bear*) and pronouns (e.g., *she*, *him*, *it*, *they*). Not only must we appropriately choose referring expressions when speaking or writing, but we also must be able to determine who or what a referring expression is identifying when we come across one during language comprehension. Understanding these referential relations is crucial for successful communication and involves being able to rapidly map pronouns and other referring expressions onto their antecedents (i.e., the preceding entity that a pronoun co-refers with). However, determining who or what a pronoun refers to is not always straightforward since the meaning of a pronoun is completely dependent on the antecedent. For example:

1) *He wanted to go see it.*

In this example there is no way of knowing who *he* is or what *it* is, because there are no possible antecedents. Now compare the sentence in 1) with example below:

2) *Noah heard that the new Marvel movie was now playing in theatres. He wanted to go see it.*

It is now clear that he is referring to *Noah* and it is referring to *the new Marvel movie*, such that *Noah* and *the new marvel movie* are the antecedents. However, in real language use the relationship between a pronoun and its antecedent can often be ambiguous. Compare the sentence in 2) with the example below:

3) *James told Noah that the new Marvel movie was now playing in theatres. He wanted to go see it.*

In this example *he* could technically refer to either *James* or *Noah*. Despite this ambiguity, we know from previous research that people are more likely to assume the pronoun is referring to *James*. This is because there is a preference for the subject of the preceding clause to be the antecedent for an ambiguous third person pronoun, which is known as the *subject bias* (sometimes also referred to as first-mention bias, because in English the two are most often confounded; e.g., Crawley, Stevenson, & Kleinman, 1990; Frederiksen, 1981; Garnham, Traxler, Oakhill, & Gernsbacher, 1996; Gernsbacher, 1989; McDonald & MacWhinney, 1995). For the purposes of the current thesis, we are specifically concerned with the third person singular pronoun *he*. Thus, when we use the term ‘pronoun’, it is in reference to this particular type.

Arnold, Eisenband, Brown-Schmidt, and Trueswell (2000) were the first to use the VWP to investigate how order of mention influences online pronoun resolution. Employing a ‘listen-and-look’ task, participants listened to scenarios containing temporarily ambiguous pronouns, while viewing scenes with two characters (e.g., Donald and Mickey). For example:

4) *Donald is bringing some mail to Mickey while a violent storm is beginning. He's carrying an umbrella, and it looks like they're both going to need it.*

This VWP task differs from those previously described, in that participants are not required to carry out a motor task based on the linguistic input. Rather participants were instructed to listen to the scenarios and determine whether or not they matched the pictures (i.e., a picture verification task). The eye gaze data revealed that participants were more likely to fixate on the first-mentioned character (i.e., the grammatical subject) when the scene depicted the first-mentioned character carrying out the action of the second clause (i.e., Donald carrying an umbrella). However, when the scene depicted the second-mentioned character (i.e., the grammatical object) carrying out the action (i.e., Mickey carrying an umbrella) participants were equally as likely to fixate on the two characters, suggesting that they expected the pronoun to co-refer with the first-mentioned character. Thus, the inconsistency between participants' expectations and the visual scene is reflected by their eye gaze data. These findings are consistent with previous studies (e.g., Garnham et al., 1996; Gernsbacher, 1989; McDonald & MacWhinney, 1995) and provide further evidence that individuals prefer an ambiguous pronoun to co-refer with the (first-mentioned) subject of the preceding clause. However, the results of the picture-verification task revealed that, in the end, participants responded correctly even when the second-mentioned character performed the action. This highlights the importance of using online measures, such as VWP eye-tracking, when interested in the time course of language processing, because offline measure often only reflect the 'end state'.

There are different accounts as to why there is a bias to the subject/first-mention referent. *First-mention* accounts (e.g., Carreiras, Gernsbacher, & Villa, 1995; Gernsbacher & Hargreaves, 1988; Gernsbacher, Hargreaves, & Beeman, 1989) argue that the preferred antecedent for an ambiguous pronoun is the first-mentioned entity in the preceding clause (independent of its grammatical function), as it forms the foundation for which all other information is mapped. This foundational mapping is a general cognitive principle, rather than language specific one. In contrast, *subject-preference* accounts argue that the preferred antecedent for the ambiguous pronoun is the actual grammatical subject of the preceding clause, independent of its position in the sentence (e.g., Crawley et al., 1990; Frederiksen, 1981). There are also alternative accounts, such as the structural parallelism account (e.g., Chambers & Smyth, 1998; Sheldon, 1974; Smyth, 1994), which argues that the prefered antecedent for an ambiguous pronoun is the noun phrase in the preceding clause that has the same grammatical role (and/or syntactic position); thus the preferred antecedent for an object pronoun is the preceding object and the prefered antecedent for a subject pronoun is the preceding subject. Because English has a relatively fixed subject-verb-object (SVO) word order, the grammatical subject is almost always the first-mentioned referent; thus it is difficult to contrast the different accounts based on English alone (but see e.g., Chambers & Smyth, 1998; Fukumura & van Gompel, 2015).

Järvikivi, Van Gompel, Hyönä, and Bertram (2005) investigated the influence of order-of-mention and grammatical role on pronoun resolution in Finnish. Finnish is a language with flexible word order, such that grammatical role is marked morphosyntactically; thus it is possible for both grammatical subjects and objects to be mentioned first. Using a similar design to that of Arnold et al. (2000), participants listened to sentences in which word order was manipulated, appearing in either SVO or OVS. The eye gaze data revealed that both order-of-mention and grammatical role information influenced pronoun resolution, such that each had an independent effect. This findings suggest that neither a subject preference nor first-mention account can fully account for the existence of a first-mention/subject bias.

Although there is ample evidence that adults expect (personal) pronouns to co-refer with the subject of the preceding clause (e.g., Arnold et al., 2000; Gordon, Grosz & Gilliom, 1993; Järvikivi, et al., 2005; Kaiser. & Trueswell, 2008; Yang, Gordon, Hendrick & Hue, 2003), when exactly this bias develops has been a topic of interest in more recent research. VWP studies have reported that children as young as 2.5 to 4-years old appear to be sensitive to some of the same contextual information as adults, showing a preference for the subject/first-mentioned referent as the antecedent for an ambiguous pronouns (e.g., Hartshorne, Nappa, & Snedeker, 2015, for an overview; Järvikivi, Pyykkönen-Klauck, Schimke, & Hemforth, 2014; Pyykkönen, Matthews, & Järvikivi 2010; Song & Fisher 2005; 2007). However, there is evidence that such a bias does not show up until relatively late in the eye movement record, approximately 1200 ms after the pronoun onset (Pyykkönen et al., 2010; Hartshorne, Nappa, & Snedeker, 2015, for an overview). This suggests that although children may be sensitive to some of the same contextual information as adults, they may be slower at processing. Furthermore, these findings demonstrate the effectiveness of VWP in picking up on time sensitive differences in language processing between adults and children.

1.4.1. Limitations of Previous VWP Pronoun Studies

It is assumed that pronouns serve as a shortcut means to refer to highly salient entities, that is, entities that are highly active in a listener's discourse representation and are thus easily accessible to be picked up as the antecedent (e.g., Foraker & McElree, 2007; Gernsbacher, 1990; Gernsbacher & Hargreaves, 1988; Gundel, Hedberg, & Zacharski, 1993). Although the previously mentioned studies provide evidence that order-of-mention and subjecthood influence a referent's salience (otherwise individuals would be equally as likely to look the the second-mentioned object), one should be aware of the types of experimental items used in these studies. Most of these items were only two to four sentences long, in which two referents get introduced, one performs an action and then immediately following there is an ambiguous pronoun. For example:

5) *Here is a panda and a parrot. The panda hit the parrot by the lake. He wanted to take a nap*

With these types of items, salience essentially becomes synonymous with subjectivity, such that besides the fact that one referent inevitably has to be mentioned first, neither are truly that 'salient'. The reason being, is that many additional factors also influence a referent's linguistic salience (see Arnold, 1998 for an overview), such as agentivity, information structure (given vs. new; topic vs. focus), discourse status (i.e. whether or not the referent is the current topic of the discourse), discourse connectives and verb semantics. Even with these types of VWP studies, the subject bias can be overridden by manipulating factors, such as properties of the verb (e.g., Järvikivi, Van Gompel, & Hyönä, 2017; Pyykkönen, et al., 2010; Schumacher, Roberts, & Järvikivi, 2017). For example, Schumacher and colleagues (2017) investigated the influence of grammatical role (subject vs. object), thematic role (agent vs. patient, as signalled by the verb) and the information status of potential referents (canonical vs. non- canonical word order), on pronoun resolution in German, which has a flexible word order. They found that semantic role was a better predictor of pronoun resolution than both subjecthood and word order. This demonstrates that manipulating factors within these types of isolated items can drastically change how individuals interpret the pronouns, as reflected by the eye movement patterns. Rather than it being a subject bias, the preference could actually be for the ambiguous pronoun to co-refer with the most salient referent of the preceding clause, and as such semantic role may have a stronger influence on salience than does grammatical role. Furthermore, there are also different discourse properties that can influence referent salience, such as a referent's discourse status and different discourse connectives, which cannot be captured with isolated items.

In addition to the lack of larger discourse context, in many of these previous studies the visual display remained on the screen for multiple seconds after the onset of the ambiguous pronoun. This means that participants were able to process the pronoun without interference from new incoming information, which is not representative of real language processing. It is unclear whether eye movements during spoken discourse processing could serve as indicators of comprehension in the the same way they do during more contrived language processing contexts. For example, when children took 1200 ms to fixate on the subject after pronoun onset, there was no additional incoming information, meaning there was sufficient amount of time for this evidence to surface. However, it is not clear whether this same evidence (i.e., fixation to the subject) would surface if children also had to process new incoming information, as would be the case when listening to a longer discourse.

Finally, in almost all of these studies there were only two referents in the visual scene. This may have led to a further overestimation of a subject bias, such that the proportion of looks to the subject may have been lower if there had simply been more to look at on the screen. There is evidence that increasing the number of referents in the visual scene influences eye gaze patterns. For example, Ferreira, Foucart & Engelhardt (2013) investigated whether the complexity of the visual scene influences online sentence processing. They first replicated the findings of the original Tanenhaus et al. (1995) study, by having participants listen to instructions such as, 'Put the book on the chair in the bucket' while viewing a traditional 4-object array (i.e., an array with (1) a chair, (2) a bucket, (3) a book on a chair, and (4) either a separate book: 2-referent condition or unrelated object: 1-referent condition). Just as in the original study, participants displayed a garden path effect in the 1-referent condition, but not in the 2-referent condition. However, when participants listened to the same instructions while viewing a 12-object array, they were actually more likely to look at the distractor than the target during the first time window in the 2-referent condition, suggesting that they were still searching for the target. In fact, looks to target did not increase until much later on for the 12-object display compared to the traditional 4-object one; thus it was concluded that participants' inferencing is influenced by the complexity of the visual array. This suggests that certain VWP findings are only replicable under specific conditions. Therefore, it is unclear whether there would still be evidence for a subject bias if there were more characters in the visual scene. Furthermore, in the real world there are often numerous entities in our visual field, and it is not as if we fixate on these entities everytime they are mentioned. Thus, investigating eye movements when there are only 2-4 things to look at is not reflective of everyday language processing.

Although the previous studies were able to use the VWP to provide insight into the time course of online pronoun resolution, they lacked the naturalistic essence of real language use. Spivey and Huette (2016) argued that our experimental designs should consist of ecologically valid tasks that situate language users in realistic language environments. It is hard to argue that the presentation of numerous items in isolation, with a lack of context, is representative of a realistic language environment. Spivey and Huette (2016) further argued that if we "continue to focus on one *"process-in-question"* and one contextual manipulation, while brutishly eliminating all other contextual variables from the stimulus environment, then our research field risks producing results that do not generalize to natural situations" (p. 5). In other words, we need to take steps to study language processing in naturalistic language settings, where numerous sources of information are available and thus, relevant to processing.

As such, it needs to be made clear that the findings of the previously mentioned studies may only apply to the specific contexts in which they were investigated, meaning that they might not generalize in more naturalistic language settings. It is unclear whether there truly is a subject bias or if it is just an artifact the a simplistic set-up of the previous studies. Additionally, it is unclear to what extent the VWP can be used to capture what is going on during real-time language processing under more naturalistic conditions. For example, say underlyingly there truly is a subject bias, it is unclear whether evidence of such would surface in the VWP eye movement record when processing longer discourses with multiple characters in the visual scene. The point is that just because the eye movement record does not provide evidence for a certain phenomenon, does not mean that phenomenon does not exist. One way to begin answering some of these questions is to actually investigate pronoun resolution within the realm of continuous discourse processing. Natural communicative language often consists of series of utterances, where each utterance typically has some relation to what came before it and what will come after it, and as such we need to investigate language processing in the same type of contexts.

1.5. Reference Processing within a Continuous Discourse

In order to understand a narrative (or any piece of linguistic discourse) one must construct some sort of situation model, which is a mental representation of the events described by the text (or dialogue) that continuously gets updated as new information is received (Johnson-Laird, 1983; van Dijk & Kintsch, 1983). Referential coherence is a key aspect of discourse representation, as situation models are thought to be primarily built around the characters (or protagonists) of the discourse, as it is their actions that drive the events that take place (Zwaan & Radvansky, 1998). Upon receiving new information a listener or reader must determine to who that information refers to and relate it to what they already know about said character (Morrow, 1985). In the most basic terms, a listener or reader must keep track of who does what to who and when.

To date, only a single study has used the VWP to investigate referential processing during continuous discourse comprehension. Engelen, Bouwmeester, de Brain and Zwaan (2014) had children listen to a 7-minute long story while viewing a display with black and white line drawings of four animals. They analyzed eye gaze data for both names (e.g., rabbit) and pronouns (e.g., he). The reason for such, is that names and pronouns are assumed to be used for different purposes, and therefore different mechanisms are uniquely involved in their processing. Names can be directly mapped onto discourse entities, and are normally used to introduce a new entity or shift attentional focus (or topic) from one entity to another. Pronouns on the other hand, require inferential processing, such that their interpretation depends on the immediate (linguistic or conceptual) context. Generally, pronouns signal referential continuity, such that they cue the maintenance of the most highly activated entity in an individual’s mental representation of the discourse representation (or in working memory). The eye gaze data in Engelen et al. (2014) revealed that the probability of fixating on the target increased during the 2 seconds after the onset of the referring expressions, however, it increased more for names than pronouns (i.e., pronouns had a shallower slope). In line with this, the probability of fixating on the target at onset (signaling anticipation) was higher for pronouns than names. Engelen et al. (2014) interpreted this finding as being consistent with the fact that pronouns are generally used to refer to highly activated entities, and thus they are easier to predict, as reflected by higher proportion of fixations at the onset. Furthermore, they found that as the story progressed over time, children were less likely to fixate on the target picture. They suggested that the visual scene may be particularly useful when building a mental representation of the discourse, such that listeners will search for the appropriate referent in the visual scene, as reflected by the eye movements. However, once a mental representation is well established, the supporting role of the visual scene becomes less obvious since it does not provide any additional information. Additionally, when , the authors examined good and poor comprehenders separately, they found that the probability of fixating on the target (for both names and pronouns) was overall higher for the good comprehenders. However, both good and poor comprehenders adjusted their eye gaze similarly in response to hearing a referring expression (i.e., the slopes for gaze to the referent over the 2 s time window were the same, but higher overall for the good comprehenders). They also found that good comprehenders were more likely to make anticipatory fixations, especially for pronouns. They suggested the reason for this may be a combination of good comprehenders being better at monitoring and applying forward-looking cues in order to anticipate the referents of a pronoun, and poor comprehenders struggling to make the appropriate inferences necessary for identifying a pronoun’s referent. Lastly, good comprehenders were even less likely to fixate on the targets as the story unfolded over time, suggesting that they may have established a sufficient mental representation of the discourse sooner than poor comprehenders.

Although Engelen and colleagues (2014) attempted to combat some of the issues associated with more traditional VWP studies by investigating the processing of reference within a continuous discourse, the study was not without limitations. First, children viewed the same visual display for the entire 7-minute duration of the story. This was a simple array of 4 black and white animal line drawings. As such, it is not surprising that looks to the targets decreased as the story unfolded. Not only did the visual scene not aid comprehension, but it is also likely that the children quickly became bored with it. Furthermore, 8 out of the 14 pronouns that were analyzed all came from the same paragraph, 6 of which referred to the same referent, as can be seen below (Engelen et al., 2014; p.71):

> *6) “Hello squirrel!’’ the mouse said in a loud voice. ‘‘Have you seen rabbit?’’ the squirrel asked. ‘‘Rabbit?’’ the mouse replied. ‘‘No, we thought he’d be here with you. Where did **he** go?’’ ‘‘Err . . . ,’’ the squirrel stammered. ‘‘Err . . . **He** just left for the giant rock. To collect something.’’ ‘‘Anyway, we didn’t see him,’’ the mouse said. ‘‘But **he** left through the same bushes you just came from. **He** wouldn’t be lost, would he?’’ the squirrel asked. ‘‘Well, I’m sorry,’’ the mouse said. ‘‘Perhaps **he** got stuck in a blackberry bush.’’ ‘‘Oh no!’’ the squirrel cried. ‘‘But was **he** in a hurry? Otherwise he’d be more careful, right?’’ the mouse said. ‘‘Err ...,’’ the squirrel hesitated. ‘‘Yes, he was in a hurry, I guess.’’ And then **he** decided to confess the prank he had pulled. **He** was very sorry and started to weep quietly.*

Finally, none of the analyzed pronouns immediately followed a clause or sentence with both a subject and object and as such, nothing can be said about the influence of grammatical role or order of mention on the online processing of pronouns within a continuous discourse. Despite the promising nature of this study, the results should be interpreted with these limitations in mind.

1.6. Individual Differences

Another question that arises is what makes an individual a ‘good’ versus a ‘poor’ comprehender? When investigating reference processing within the realm of discourse, it is important to consider the underlying cognitive mechanisms that are thought to influence one’s ability to appropriately understand a piece of discourse. As previously mentioned, understanding discourse involves constructing a mental representation of the discourse, and as it unfolds and new information is received, the mental representation must be continuously updated. As such, working memory is assumed to play an important role in language comprehension/discourse processing, as a listener must be able to keep track of who does what to whom and when. Miyake and Shah (1999) defined working memory (WM) as “the mechanisms or processes that are involved in the control, regulation, and active maintenance of task-relevant information in the service of complex cognition” (p. 450). Thus, WM is necessary for understanding anything that unfolds in real time, as it allows for us to integrate various sources of information (including real world knowledge and perceptual information), as those sources become available to us. A number of studies have investigated the relationship between WM and and different types language processing, including reading comprehension (e.g., Cain, Oakhill & Bryant, 2004; Carretti, Cornoldi, De Ben & Romanò, 2005; Seigneuric, Ehrlich, Oakhill & Yuill, 2000), tests of verbal analogies (van der Sluis, de Jong & van der Leij, 2007), sentence comprehension (e.g., Engel de Abreu, Gathercole & Martin, 2011; Kidd, 2013; Kidd, Lieven, & Tomasello, 2006; Montgomery, Magimairaj & O'Malley, 2008) and the comprehension of temporal relations (e.g., Blything & Cain, 2016; Blything, Davies, & Cain, 2015).

Daneman and Carpenter (1980) were among the first one to investigate whether WM capacity is related to one's ability to successfully interpret pronominal reference. In their study participants completed two different tasks. The first was a complex span task designed to assess WM capacity, in which participants listened to a series of sentences increasing in length (i.e., 1 sentence, 2 sentences, etc.), then judged whether the sentences were true or false, and were then asked to recall the final word of each sentence. The second task was a listening comprehension task, in which participants listened to passages ranging from two to seven sentences and were then asked pronominal reference questions that required them to appropriately identify the antecedent of a pronoun. The findings revealed that performance on the complex span task was strongly correlated with performance on the comprehension questions. Furthermore, individuals with higher scores on the span task were better at accurately identifying the antecedent of a pronoun over greater distances. These findings were taken as evidence that individuals with greater WM capacity are better at appropriately interpreting pronominal reference. More recently Whiteley and Colozzo (2013) investigated the relationship between WM (and more specifically the updating component of WM) and children's ability to appropriately choose referring expressions during narrative production. Children completed a series of tasks designed to assess the updating component of WM, and also produced two narrative stories. Overall, they found that the ability to adequately choose referring expressions was related to WM capacity, especially in the case of maintaining reference to the same character over multiple utterances (i.e., effectively using pronouns). Also using a production task, Kuijper, Hartman and Hendriks (2015) found that when reintroducing a referent that is not the current topic of the discourse, children with poorer WM were more likely to use an ambiguous pronoun rather than a more appropriate NP. Although there is evidence that WM capacity is related to both the comprehension and production of pronominal reference within longer discourses, very little is known about WM's relation to the actual time-course of online pronoun resolution. In other words, the previously mentioned studies looked at the 'end state' of reference comprehension and production (i.e., the final decision about who a pronoun was referring to, or the final decision of which referring expression use). WM capacity likely influences the online processing of reference within a discourse (especially in the case of pronouns), and may be particularly important for integrating linguistic and visual information, as is the case in the VWP.

In addition to domain-general cognitive mechanisms, such as working memory, fairly robust language measures have also been shown to predict discourse and reference processing abilities. For example, receptive vocabulary predicts eighth-grader's ability to make appropriate discourse inferences (e.g. Karasinski & Weismer, 2010) and productive vocabulary predicts 2-year old's ability to anticipate upcoming spoken linguistic input (e.g., Mani & Huettig, 2014). More recent studies, have also shown that vocabulary knowledge predicts online pronoun processing (e.g., Arnold, Strangmann, Hwang, Zerkle, & Nappa, 2018; Järvikivi, Porretta, Paradis, Govindarajan & Day, 2017), however, these studies investigated online pronoun resolution using isolated items, rather than a continuous discourse.

Chapter 2: Present Study

The present study took the first steps to assess the application of the visual world eye tracking paradigm (VWP) as a tool for investigating online language processing in a more naturalistic setting. Thirty-seven children and 21 adults listened to a 5-minute long storybook while wearing eye tracking glasses. We were specifically interested in language mediated eye movements with respect to the onset of referring expressions that occurred throughout the story. As such, eye gaze patterns were analyzed for both names and pronouns, as different mechanisms are thought to be uniquely involved in their processing. Furthermore, we collected measures of working memory and vocabulary to explore whether working memory capacity and/or vocabulary knowledge influence(s) children's online processing of reference throughout a continuous discourse. To date no study has used the visual world eye tracking paradigm to investigate the online processing of reference in a context where both the language input and visual scene are more closely representative of a natural language environment. In accordance with investigating reference processing in as naturalistic of setting as possible, we opted to use eye tracking glasses as opposed to a more traditional eye tracker. The eye tracking glasses are akin to normal reading glasses and allow for the participant to move more freely throughout the duration of the experiment, which may particularly beneficial when working with children, as it is often difficult for them to sit still for long periods of time. It is unknown whether language mediated eye movements patterns are the same in longer and more natural discourses, as they are are in more traditional VWP studies, and as such, the following research questions were addressed:

1. Are the eye gaze patterns in a naturalistic discourse similar to the eye gaze patterns reported in more traditional VWP studies? More specifically,

1.a. Upon hearing a referring expression, is there an increase in the proportion of looks to entity that is being referred to?

1.b. In the case of ambiguous pronouns, is there a higher proportion of looks to the subject of the preceding clause compared to the object?

2. Are eye movement patterns uniform across the entire discourse?

2.a. Are we more or less likely to fixate on the intended referents at the beginning of the discourse as compared to the end?

3. Do measures of working memory and vocabulary predict children's eye movement patterns?

Chapter 3: Method

3.1. Participants

Thirty-eight children recruited from preschools and daycares in Edmonton participated in the study. Written parental consent was obtained prior to participation and the children received stickers and a t-shirt in exchange for their participation. Two children were excluded from the analysis due to the fact they were significantly older than the rest of the children (> 8 years old). This is because the original aim was to at collect data on a wider age range, however, due to practical concerns, almost all of the children ended up being between 4 and 6 years old. An additional 3 children were excluded due to the fact that they were non-native speakers of English, based on parental report. Furthermore, 17 children were excluded from the analysis due to issues with the eye-tracking glasses (discussed in the Analysis in section 4). Thus, a total of 16 children (7 female; M_{age}= 4.9 years; range = 4.2-6.8) were included in the final analysis. All children had normal vision and hearing based on parental report.

Twenty-one adults also participated to serve as a control group. All adults were undergraduate students at the University of Alberta and received partial course credit in exchange for their participation. Five adults were excluded from the analysis because they were non-native speakers, based on self report. An additional 4 adults were excluded due to issues with the eye-tracking glasses. Thus, a total of 12 adults (10 female; M_{age}= 20.0 years; range = 18.2-22.0) were included in the final analysis. All adults had normal vision and normal hearing based on self-report.

3.2. Materials

3.2.1. Storybook

An 22-page electronic storybook was constructed to be similar in style to a typical everyday storybook that would be read to children. The story was about a group of animal friends helping a duckling find his father. It contained multiple referring expressions, both full NPs and pronouns, and increased in complexity as the story unfolded. Complexity was defined as a combination of discourse length and the number of animal characters on the page. As such, it began with a single character, *Bear*, and after every 3-5 pages a new character was introduced (*Fox, Duckling, Frog,* and *Daddy Duck,* respectively). Furthermore, different characters were the topic of the discourse at different times, which should influence the respective salience of the different character at different points in the story. All of the characters described had the same gender (male), to ensure the ambiguity of the pronouns.

All of the illustrations were created in GIMP Photo Editor (GIMP 2.8.10, www.gimp.org), using images from freepik.com. Animal size was not controlled for, as we wanted the storybook to be as naturalistic as possible. However, each animal appeared in several different locations in the illustrations throughout the story. The storybook audio was recorded by a female native speaker of English in a sound-attenuated booth. The illustrations and audio were pieced together in Keynote (Apple Inc. Keynote 7.3.1), utilizing the *Page-Flip* transition in order to mimic the page turning that takes place during typical storybook reading. The Keynote presentation was then converted to an .m4v video. In total, the storybook was 5 minutes and 26 seconds long and consisted of 22 unique pages. The only written text that appeared in the storybook was on the first and last page, reading "*Bear, Friends and the Lost Duckling*" and "*The End*", respectively. Neither the first nor the last page contained any of the animal characters, and thus were not included in any analysis. The storybook illustrations and dialogue can be found in Appendix 1.

3.2.1.1. *NPs: Proper names.* Thirty-seven proper nouns embedded throughout the story were selected as full NP experimental items; all items were names of the animal characters. These items were selected with the criteria that they did not overlap in time with other referring expressions, such that there was not another referring expression occurring within the ~1415 ms window used for analysis. The first critical name occured 11.5 seconds into the story and the last critical name occured 5 minutes and 17 seconds into the story ($M_{\text{DistanceApart}}$ = 8.49 s; range = 1.97-26.61; *SD* = 5.96). Ten of these items were 'Bear', 9 were 'Fox', 8 were 'Duckling', 7 were 'Frog', and 3 were 'Daddy Duck'. The first 5 critical names occurred when there was one character on the page, the next 10 occurred when there were two characters on the page, the next

9 occurred when there were three characters on the page, the next 7 occured when there were four characters on the page, and the final 6 occurred when there were five characters on the page. The unequal distribution comes from the fact the that different characters were present in the story for different proportions of time.

3.2.1.2. *Pronouns.* Ten ambiguous pronouns embedded throughout the storybook were selected as pronoun experimental items. These items were selected with the criteria that they did not overlap with other referring expressions and that they immediately followed a sentence or clause with both a subject and an object. For example:

9) *Fox thanked Bear. <u>He</u> wanted to play a different game.*

Because we did not first develop the experiment items and then construct a story around them, there was variation in the exact structure of the items. This is because the aim was to investigate the processing of reference in as naturalistic of a context as possible and using highly controlled experimental items would be counterintuitive to this goal. For example, as seen in 9) above, the pronoun immediately followed the mention of the object in the preceding sentence, however, for other items a relative clause occurred between the mention of the object and the pronoun (see example 10 below). For 4 of the experimental items the pronoun immediately the object, and for the other 6 experimental items a relative clause occured between the object and the pronoun. All items can be bound in the Appendix.

10) Bear told Fox to go and hide. <u>He</u> started counting to five.

The first critical pronoun occured 55.4 seconds into the story and the last critical pronoun occurred 4 minutes and 54.8 seconds into the story ($M_{\text{DistanceApart}}$ = 26.59 s; range = 4.77-93.60; *SD* = 29.78). Bear was the subject of the preceding clause for 5 of the critical pronoun items, Fox was the preceding subject for 3 of the items and Duckling and Daddy Duck were each the preceding subject for 1 item. The first 4 critical pronoun items occurred when there were two characters on the page, the next 4 occurred when there were three characters on the page and the last 2 occurred when there were five characters on the page. The unequal distribution again comes from the fact that different characters were present in the story for different proportions of time

3.2.2. Vocabulary Knowledge

In order to assess children's vocabulary knowledge, the *Peabody Picture Vocabulary Test* (4th Edition; Dunn & Dunn, 2007) was used. The PPVT-4 is a standardized measure of receptive vocabulary for individuals aged 2:6 to 90^{+}. The assessment was administered on an iPad using the Q-interactive platform. Children were shown a 4-picture array on the iPad, and asked to touch the picture that corresponded with the word provided by the experimenter. For each word a new array was displayed. The task ended when children responded incorrectly to 8 items within a set. Standard scores were automatically calculated in Q-interactive (M_{StdScore}= 113.5; range = 102-137, *SD* = 8.0).

3.2.3. Working Memory

Working memory (WM) capacity was assessed using two different measures.

3.2.3.1 *The Nebraska Barnyard Task* (*NBT*). The *NBT* (Wiebe, Sheffield, Nelson, Clark, Chevalier, & Epsy, 2011) is a computerized complex span task adapted from the Noisy Book working memory task (Hughes, Dunn, & White, 1998). In this task children must remember a sequence of animal names and press corresponding buttons on a touchscreen laptop. During the initial training phase children learn the location/colour of the buttons corresponding to six different animals. Each animal is associated with a different coloured button (green button~frog, orange button~cat, red button~chicken, black button~horse, pink button~pig, blue button~sheep) and the location of the buttons remains the same throughout the task. During the test trials the animal images are removed from the buttons, such that six blank coloured buttons remain on the computer screen. The experimenter tells the child an animal sequence and the child must press the corresponding buttons in the correct order. Sequence length increases incrementally, where initially the experimenter names a 2 animal sequences, and after every three trials the sequence length increases by 1 animal, working all the way up to 7 animal sequences. If the first two trials for a span are correct, the third trial is skipped, and if all trials for a span are incorrect, the task is discontinued. Based on previous studies that have used this task (e.g., Chevalier, James, & Wiebe, 2014; Chevalier, Sheffield, & Nelson, 2012, Orchinik, Taylor, Epsy, & Minich, 2011; Wiebe et al., 2011) the dependent variable selected for analysis was a summary score, calculated by first dividing the number of correct button presses by the total button presses for each span length and then summing these scores across all administered span lengths (*M* = 1.86; range = 1-3.33; *SD* = 0.77). Data for two children were missing, these children were specifically excluded from the analyses that included WM.

3.2.3.2. *Listening Recall Task* (adapted from Gathercole & Pickering, 2000). In the listening recall task children are asked to listen to a series of sentences and judge whether they are correct (i.e., true or 'silly'), after which, they are asked to recall the final word of each sentence. During the first block children listen to a single sentence and are then probed to recall the final word. If children correctly recall the final word of at least three out of the four items in the block, they continue onto a second block. In the second block children listen to and judge the correctness of two sentences, and are then asked to recall the final word of each sentence. For example:

Sample trial:

1) Apples have hands (silly)
2) Tomatoes are red (true)

Recall probe: "hands", "red"

Again, if children correctly recall the final words of at least three out of the four items within the block they continue to the next block; otherwise the task is terminated. This continues all the way up until blocks of four sentences. However, none of the children in the current study made it past the block of three sentences. The dependent variable used for our analysis was the total raw score (*M* = 3.0; range = 0.0-7.0; *SD* = 2.07). This was calculated by tallying up the correct responses for all the items. Each item was scored as correct (1) or incorrect (0), regardless of how many sentences the item was. For example, in the sample trial above, the child would have to recall both "hands" and "red' (in that order), for the item to be scored as correct. Items where the child was only able to recall one of the last words, resulted in that item being scored as incorrect. Data for two children were missing, these children were specifically excluded from the analyses that included WM.

3.3. Procedure

All children were tested individually at their preschool or daycare during two separate visits. During the first visit children sat approximately 50 cm in front of a Lenovo laptop, with the experimenter sitting either to their right or left (alternated between visits). Using PowerPoint presentation, children were first familiarized to the animal characters. They were shown each animal individually and asked to name the animal. In the event that the child misnamed the animal they were corrected. Most of the times this was accidentally naming the duckling a *duck*. Once the children completed the animal familiarization, they were then told they would listen to a short story about the animals while wearing special eye tracking glasses (ETG). The eye tracking glasses were then placed on the child's head and secured using an adjustable strap. Before listening to the story, children completed a 3-point calibration. This was done in PowerPoint presentation, where children were shown a sun at three different locations on the computer screen (appearing once at the top-center and then once in each bottom-corner). Children were told to follow the moving sun with just their just eyes. After the calibration, the experimenter played the storybook in QuickTime Player.

The eye gaze data were collected using SensoMotoric Instruments (SMI) ETG wireless 2 eye tracking glasses, which are operated using a Samsung Galaxy Note 4 smartphone. Registration was binocular with a sampling rate of 60 Hz ($16.\overline{6}$ ms/frame). The ETG are akin to normal reading glasses and include a built-in high-definition scene camera.

To ensure that children had paid attention to the story, they were asked a series of five comprehension questions. In order to be included in the analysis, they had to answer at least four correctly. The only children that did not meet this criteria were non-native speakers, and as such were already excluded from the analysis of the current thesis. The comprehension questions can be found in the Appendix. Following the eye tracking and comprehension questions, children were administered the PPVT.

During the second visit children completed an executive function (EF) battery designed to assess working memory, inhibition and cognitive flexibility (set-shifting). However, the current thesis will only consider the measures of working memory, namely the Nebraska Barnyard and the Listening Recall tasks. The primary reason for this is that the (analyzable) sample size ended up being considerably smaller than anticipated and running complex statistical models requires a sufficient amount of data. Therefore, we would not have been able to include multiple measures of EF into the same model. Working memory was selected over the other two EFs based on the amount of supporting literature.

In order to provide an interpretive context for the children's gaze data, given this was a novel adaptation of the VWP, eye tracking data were also collected for a group of adult participants. All adults were tested at the *Centre for Comparative Psycholinguistics* at the University of Alberta. The adults did not complete the measures of vocabulary or executive functioning because we were only interested in these measures for the children.

Chapter 4: Analysis

4.1. Eye Gaze Data Exportation

The eye tracking data were imported into SMI's BeGaze analysis software from the smartphone SIM card. Unique scan path videos (in .avi format) were then exported for each participant. These were continuous audio and video recordings of everything the participants were hearing and seeing along with the location of their eye gaze overlayed on the visual image (in the form of a small colored circle). Still images of the scan path videos can be found in the Appendix.

4.2. Coding

4.2.1. Eye Gaze Coding

All scan path videos were hand coded frame by frame by the author using Noldus Observer XT, Version 11.0 (Noldus Information Technology, 2012). The gaze was coded as either being on Bear, Fox, Ducking, Frog, Daddy Duck or Elsewhere (i.e. in the sky, on a flower, etc.). Cases where there was no record of eye gaze were coded as NA. The coding was done quite conservatively, meaning that eye gaze had to clearly be within the bounds of the interest areas. In order to limit any biases of the coder, the coding was done without any accompanying audio (audio at such a slow speed is unintelligible, and thus was muted). During the coding phase it became apparent that there was no record of eye gaze for majority of the story for a large number of children. This was due to the fact that the glasses were not specifically designed for children and therefore, they were too big. It should be noted that at the time the current study was being designed, SMI was supposed to be releasing a child-adapted version of the ETG2 Wireless eye tracking glasses. However, this version was never released due to the fact that SMI was acquired by Apple Inc. Nevertheless, in all cases where track loss occurred for more than 50% of the story, we did not continue coding and excluded the children from the analysis. This resulted in

17 children being excluded from the analysis. The gaze data for all of the participants who were retained (16 children and 12 adults) were exported into a single spreadsheet, where each row represented a single frame, with Subject ID and Gaze Location as columns.

4.2.2. Critical Item Coding

Using the original storybook audio recording, all critical item onsets were found, for both NPs and pronouns. This was done in Praat (version 6.0.42; Boersma & Weenink, 2018), utilizing the spectrogram in order to determine the precise moment when the first segment of each item occurred. The unique onset time of the first item was then found for each participant. This is because each participant's scan path video was a different length, as it depended on when the eye tracking glasses began recording. This was done by uploading the audio of the scan path videos into Praat. Once the unique onset time was known for each participant, all other onsets were adjusted accordingly. All onset times were then converted from ms to frames by dividing by ~16.6 ms (based on a the 60 Hz sampling rate).

4.3. Data Preparation

The gaze, item and subject data were merged together in R version 3.1.1 (R-Core-Team, 2016), where again each row represented a single frame. The Gaze Location column was used to create separate interest area (IA) columns. For the NPs these were: IA_Target (gaze located on the mentioned character), IA_Elsewhere (gaze located anywhere else on the screen, including any of the other characters) and IA_NA (loss of gaze). For the pronouns these were: IA_Subject (gaze located on the subject of the preceding clause), IA_Object (gaze located on the object of the preceding clause), IA_Elsewhere, and IA_NA. All IA columns had either a value of 0 (gaze within interest area) or 1 (gaze outside interest area), meaning that only a single IA column could have a value of 1 for any given frame.

4.3.1. Binning and Empirical Logit Transformation

The data were binned into 5-frame bins, where each bin represented ~83.3 ms. This means that 5 frames of data were collapsed into a single row, so instead of an IA value being 0 or 1, it could be any value between 0 and 5. A value of 0 means that the eye gaze did not fall within the bounds of the IA for any of the 5 frames and a value of 5 means that the eye gaze fell within the bounds of the IA for all 5 frames. An empirical logit transformation was then performed on each IA column. This transforms the data in the bins into continuous measures that are unbounded. Furthermore, it allows for us to account for autocorrelation, unlike with binomial models. The empirical logits were calculated following Barr (2012)'s '*Walkthrough of an "empirical logit" analysis in R*' equation:

Empirical logit: $elog{=}log(y{+}.5/N{-}y{+}.5)$

The y variable is the number of frames the eye gaze fell within the bounds of the IA (i.e., the value from 0 to 5) and N is the number of frames in each bin (i.e., 5). Weights were also calculated using the formula below, as the variance depends on the mean. The weights estimate the variance in each time bin and are included in the statistical models to inform the models of the variance.

Weights: $(1/y{+}.5)\ {+}(1/N{-}y{+}.5)$

4.3.1.1. *Analysis windows.* For the NPs the window analyzed was 2 bins before the onset and 15 bins after the onset, for a total of 17 bins (85 frames, ~1415 ms). For the pronouns the window analyzed was 5 bins before the onset and 30 bins after the onset, for a total of 35 bins (175 frames, ~2915 ms). There are two reasons why different time windows were chosen. First, in order to be consistent, we wanted to use a time window similar to previous VWP pronoun resolution studies. However, a longer analysis window would result in more items being excluded from the analysis due to overlap. Therefore, we opted for a smaller time window for the NPs in order to maximize the amount of experimental items that were included in the analysis. We further reasoned that the shorter analysis window would be sufficient for the full NPs because of the one to one mapping between the linguistic input and the corresponding entity in the visual scene (unlike with the pronouns, which need to be 'resolved').

4.4. Statistical Methods

The data were analyzed in R version 3.1.1 (R-Core-Team, 2016) using Generalized Additive Mixed Models (GAMM, Wood 2006, *mgcv* R-package), which are specifically designed to model nonlinear data. Like most time series data, visual world eye-tracking data is typically nonlinear, and as such a GAMM analysis is more optimal than more standardized linear modeling. The nonlinear relationship between the dependent variable and the predictors is modeled as a smooth function, which is a weighted sum of a set of base functions that each have a different shape. In the most basic terms, GAMMs work by looking for the nonlinear function that underlies the data by adding together multiple nonlinear forms that have been weighted according to how well they fit the data. Additionally, GAMMs are particularly well suited for dealing with interactions between continuous variables, which are included throughout the analyses. A GAMM output consists of two components: a parametric component and a non-parametric component. The parametric component is the linear part and provides information about the intercept and slope, which can be interpreted just like standard linear models, where the intercept is the expected mean value of the response variable when the predictor variables are zero. The non-parametric component contains information about nonlinearity (wiggly curves and wiggly (hyper)surfaces), interactions and random effects, which are specified using 'smooth' terms. The significance of the smooth terms are evaluated through the effective degrees of freedom (edf). An edf of 1 indicates that the effect of the predictor is linear, and the higher the edf value the more 'wiggly', non-linear, the relationship is. The p-value of the smooth term indicates whether the wiggliness is warranted. The non-parametric component does not specify the direction or magnitude of the effects, just whether or not they are significant; as such the effects must be estimated through visualization. Nonlinear terms in the model are interpreted by plotting the partial effect of the smooth together with a 95% confidence interval. For the current thesis this was done using the *itsadug* R-package (van Rij, Wieling, Baayen, & van Rijn, 2017). GAMMs have been used to model various types of linguistic data including reaction times (e.g., Baayen, 2010a; Porretta et al., 2016), production latencies (e.g., Lõo, Tucker, Järvikivi, Tomaschek & Baayen, 2018), event-related potentials (e.g., Kryuchkova, Tucker, Wurm, & Baayen, 2012; Meulman, Wieling, Sprenger, Stowe, & Schmid, 2015) and visual world-eye tracking (e.g., van Rij, Hollebrandse, & Hendriks, 2016).

Chapter 5: Results

5.1. NPs: Proper names

5.1.1. Variables

The input variables to the model were as follows: The primary variable of interest was (time) *Bin* (where bin 0 is the onset of the referring expression). In addition to this, *Group* (adult vs. child), and *Story Position* (how far into the story the referring expression occurred measured in seconds) were included. The response variable of the model was the empirical logit *Target Looks*.

5.1.2. Model fitting and evaluation

The input variables above were fit to the response variable (elog Target Looks). Interactions between Bin and Group, and Story Position and Group were included in the model as nonlinear smooths (Wood, 2006). A 3-way interaction between Bin, Story Position and Group was included in the model using a tensor product smooth (Baayen, 2010b). Random slopes and intercepts for Event (a combination of subject and item) were also included. To account for autocorrelation, an AR1 model was included by specifying the rho parameter and starting point for each time series. Finally the weights of the response variable were included, to inform the model of the variance within the time bins.

The model was fit using a backward-fitting model comparison procedure (Zuur, Ieno, Walker, Saveliev, & Smith, 2009), starting with the full model where all predictors and interactions were included and method was set to ML (Maximum Likelihood). The predictors' contributions to the model were evaluated based on the estimated p-values of the smoothing parameters and parametric components, which indicate whether or not the functional form of the predictor is different from zero. Predictors with p-values greater than the conventional alpha level of 0.05 were considered for removal, by comparing the AIC (Akaike information criterion; Akaike, 1998) value of the model without the predictor to that of the model with the predictor, using the compareML() function from the *itsadug* package (van Rij et al., 2017). The use of AIC is an information-theoretic approach that supplies information on the strength of evidence for a particular model given the data when a particular smoothing parameter is removed. Lower AIC values indicate increased evidence for that model. Predictors that did not significantly contribute to the model, based on an alpha value greater than 0.05 in the compareML summary, were removed. The main effect of Group was nonsignificant, but remained in the final model because it entered into significant interactions with Bin and Story Position, as well as a 3-way interaction with bin and story position. The final model accounted for 65.6% of the deviance explained and the full model summary is presented in Table 5.1.

Table 5.1
Generalized additive mixed model for the NP Target Looks, reporting parametric coefficients (Part A) and the estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), F- and p- Values for smooth terms (Part B).

A. Parametric coefficients	Estimate	Std. Error	t-Value	p-Value
Intercept	-1.22440	0.07736	-15.827	<2e-16 ***
GroupChild	0.04355	0.10235	0.426	0.67
B. Smooth terms	**EDF**	**Ref.df**	**F-Value**	**p-Value**
s(Bin):GroupAdult	4.931	6.214	3.118	0.004648 **
s(Bin):GroupChild	3.989	5.102	7.431	4.61e-07 ***
s(StoryPos):GroupAdult	1.004	1.005	16.031	5.96e-05 ***
s(StoryPos):GroupChild	1.004	1.005	28.509	8.96e-08 ***
ti(Bin,StoryPos):GroupAdult	11.158	13.427	2.133	0.006632 **
ti(Bin,StoryPos):GroupChild	4.213	5.158	4.466	<0.000396 ***

s(Bin,Event)	789.031	1032.000	40.984	< 2e-16 ***
s(Event)	873.272	1032.000	46.857	< 2e-16 ***

The GAMM fitted to empirical logit Target Looks revealed a nonlinear relationship between Target Looks and Bin for both adults and children. The partial effect of Bin (i.e., the isolated effect when all other variables are held constant) for both adults and children can be visualized in the upper panels of Figure 5.1. Target Looks increased as Bin progressed for both adults and children up until approximately bin 10 (~ 830 ms), after which Target Looks started decreasing. It can also been seen that, the relationship was more wiggly for adults. The summed effect of Bin (i.e., the effect when all partial effects are summed) for children and adults can be visualized in the lower left panel of Figure 5.1. The difference between the summed effects of Bin for adults and children can be visualized in the lower right panel of Figure 5.1. The dashed red lines indicate the difference between children and adults was significant between approximately bins 5-10 (~ 415-830 ms). There was a significant 3-way interaction between Bin, Story Position and Group, which can be visualized in Figure 5.2. The contour plots show how the relationship between Bin and Target Looks changed as the story progressed (i.e., Story Position) for both adults and children. The contour plots can be read like a topographic map with peaks and valleys, where dark blue indicates fewer Target Looks and yellow indicates more Target Looks. The contour lines that are spaced closely together represent steeper slopes and the contour lines that are spaced further apart represent shallower slopes (i.e., slower changes to Target Looks). Both adults' and children's Target Looks increased as Bin progressed, however, this likelihood decreased in a nonlinear fashion as the story progressed over time.

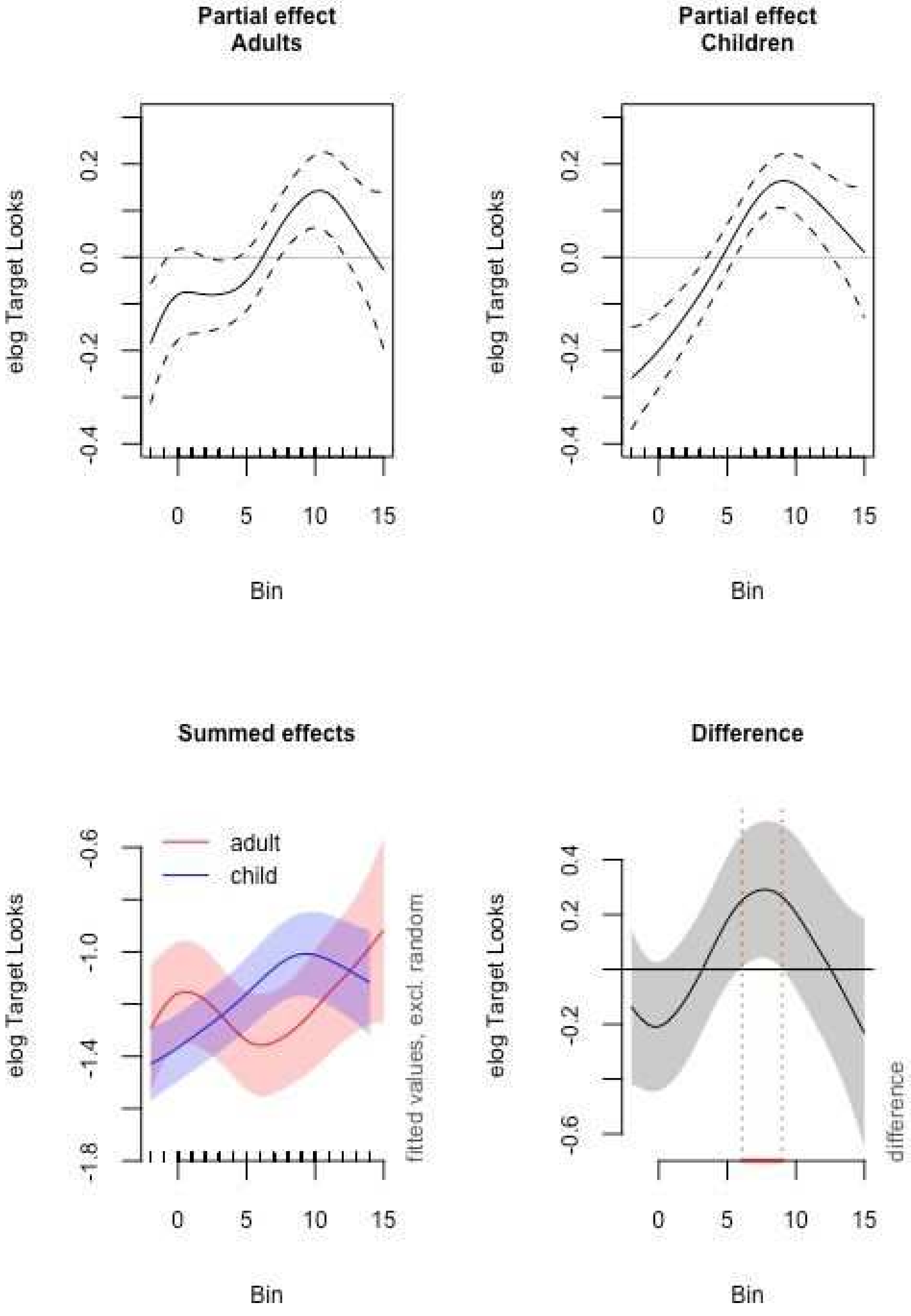


Figure 5.1: *Upper panels.* Partial effect of Bin in Target Looks for adults (left) and children (right). The solid horizontal line represents the zero effect and dashed lines represent the 95% confidence bands of the regression line. *Lower left panel*. Summed effect of Bin for adults (red) and children (blue). *Lower right panel*. Difference between the summed effect of Bin for children and adults. The dashed lines indicate where the difference is significant.

For example, in Figure 5.2 it can be seen that ~50 seconds into the story Target Looks increased as Bin number increased, as reflected by the colour changing from green to yellow as you move from left to right. However, ~250 seconds into the story there was almost no effect of Bin, as reflected by the solid blue colour as you move from left to right. The peak was also steeper for the children, indicating more Target Looks as compared to adults, and the effect was present for a longer amount of time as the story progressed.

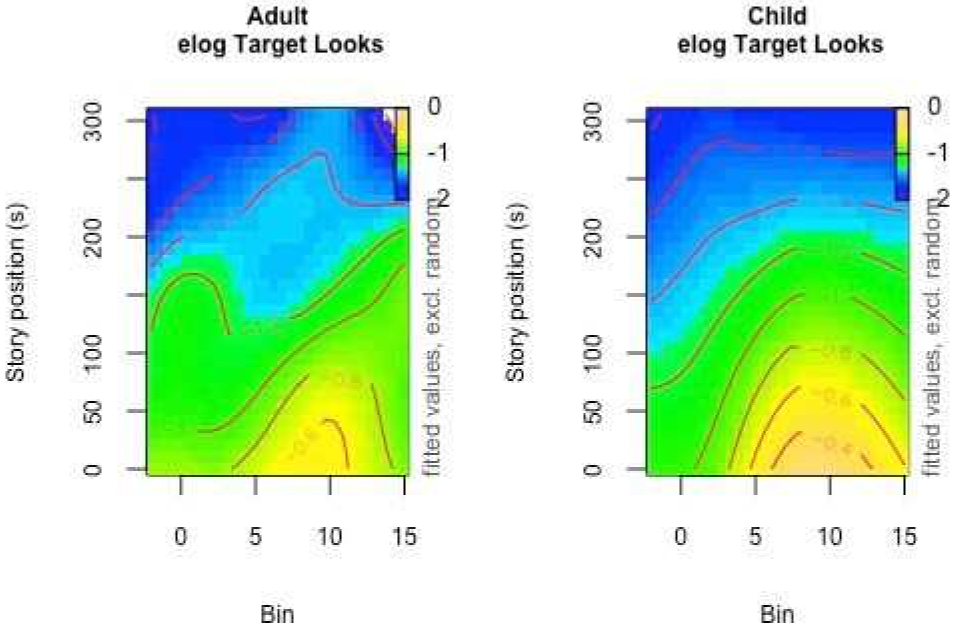


Figure 5.3: Contour plots of the interaction between Bin and Story Position by Group in Target Looks. Dark blue indicates lower (elog) Target Looks and yellow indicates higher (elog) Target Looks.

To summarize, upon hearing a full NP at the beginning of the story, both adults' and children's looks to the target increased as (time) Bin progressed. However, as the story progressed this likelihood decreased, such that by the end of the story the mention of a full NP did not influence looks towards the target. Although the pattern was similar, this happened sooner in the story for adults.

5.2. Pronouns

5.2.1. Variables

The input variables to the model were as follows. The primary variable of interest was (time) *Bin* (where Bin 0 is the onset of the referring expression). In addition to this, *Group* (adult vs. child), and *Story Position* (how far into the story the referring expression occurred measured in seconds) were included. The response variable of the model was *Subject Preference Looks*, which was calculated by taking the difference between empirical logit looks to the subject of the preceding clause and empirical logit looks to the object of the preceding clause. Positive values indicate a subject preference and negative values indicate an object preference.

5.2.2. Model fitting and evaluation.

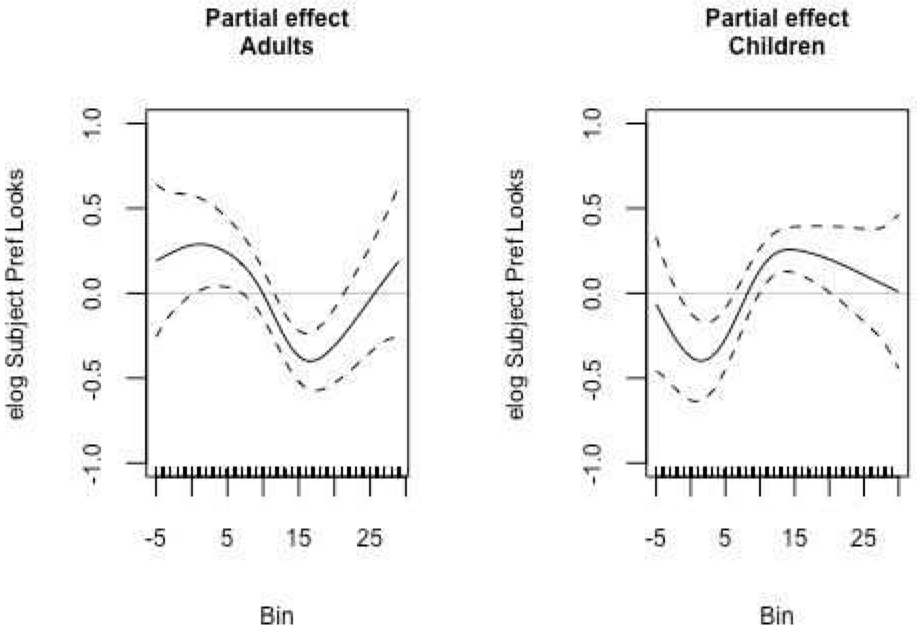
The input variables above were fit to the response variable (Subject Preference Looks). Interactions between Bin and Group, and Story Position and Group were included in the model as nonlinear smooths (Wood, 2006). A 3-way interaction between Bin, Story Position and Group was included in the model using a tensor product smooth (Baayen, 2010b). Random slopes and intercepts for Event (a combination of subject and item) were also included. To account for autocorrelation an AR1 model was included by specifying the rho parameter and starting point for each time series. Because the response variable was the difference in looks between two IAs, weights were not included, as it is not clear how they should be calculated.

Here, the model fitting procedure was the same as in the previous model. During the fitting process, the interaction between Story Position and Group was removed, as it did not significantly contribute to the model. Although nonsignificant, the main effects of Group and Story Position were retained in the final model, as they entered into significant interactions with Bin. The final model accounted for 50.7% of the deviance explained and the full model is presented in Table 5.2.

Table 5.2
Generalized additive mixed model for Subject Preference Looks, reporting parametric coefficients (Part A) and the estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), F- and p- Values for smooth terms (Part B) for the Pronoun Subject Preference model

A. Parametric coefficients	Estimate	Std. Error	t-Value	p-Value
Intercept	0.160973	0.403900	0.399	0.690
GroupChild	0.177604	0.329798	0.539	0.590
StoryPos	-0.001845	0.002087	-0.884	0.377
B. Smooth terms	EDF	Ref.df	F-Value	p-Value
s(Bin):GroupAdult	4.650	5.697	6.151	4.26e-06 ***
s(Bin):GroupChild	5.023	6.119	5.376	1.31e-05 ***
ti(Bin,StoryPos):GroupAdult	12.786	14.424	4.932	1.13e-09 ***
ti(Bin,StoryPos):GroupChild	9.457	11.735	3.575	2.70e-05 ***
s(Bin,Event)	246.884	277.000	389.217	< 2e-16 ***
s(Event)	258.123	277.000	413.129	< 2e-16 ***

The GAMM fitted to Subject Preference Looks revealed a nonlinear relationship between Subject Preference Looks and Bin for both adults and children. The partial effect of Bin (i.e., the isolated effect of the interaction when all other variables are held constant) for both adults and children can be visualized in the upper panels of Figure 5.3. For adults, Subject Preference Looks decreased nonlinearly with Bin up until approximately bin 15 (~1250 ms), after which Subject Preference Looks started increasing. The opposite was true for children, such that Subject Preference Looks increased nonlinearly with Bin up until approximately bin 15 (~1250 ms), after which Subject Preference Looks started decreasing. However, the model is less confident about the relationship between Bin and Subject Preference Looks at the extreme ends of Bin, as can be seen by the widening confidence intervals. The summed effect of Bin (i.e., the effect when all the partial effects are summed) for adults and children be be visualized in the lower left panel of Figure 5.3. The difference between the summed effects of Bin for children and adults can be visualized in the lower right panel of Figure 5.3. The dashed red lines indicate that the difference between children and adults was significant between approximately bins 13-17 (~ 1080-1415 ms).



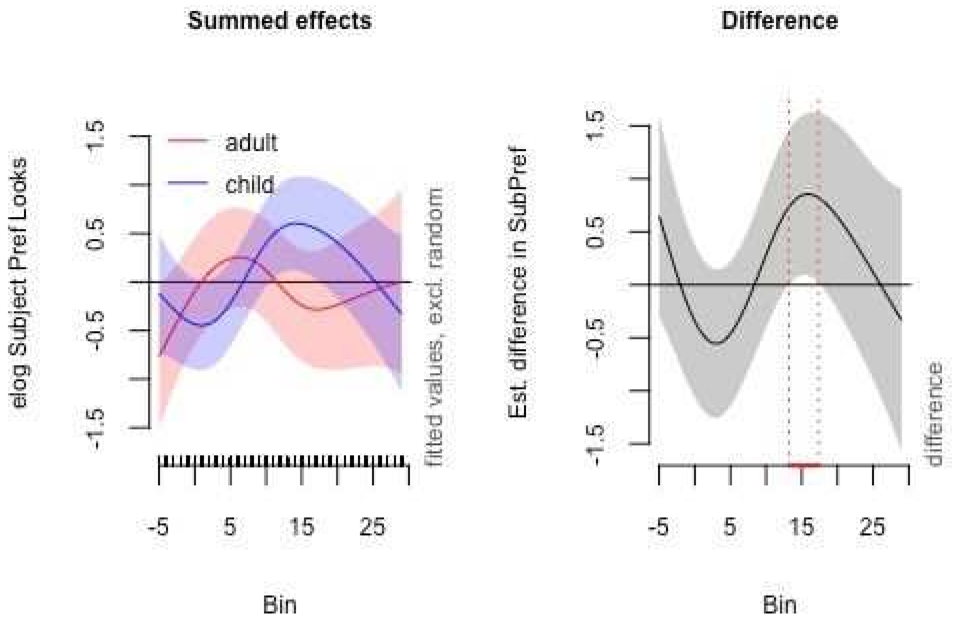


Figure 5.3: *Upper panels.* Partial effect of Bin in Subject Preference Looks for adults (left) and children (right). The solid horizontal line represents the zero effect and dashed lines represent 95% confidence bands of the regression line. *Lower left panel*. Summed effect of Bin for adults (red) and children (blue). *Lower right panel*. Difference between the summed effect of Bin for children and adults. The dashed lines indicate where the difference is significant.

There was a significant 3-way interaction between Bin, Story Position and Group, which can be visualized in Figure 5.4. The contour plots show how the relationship between Bin and Subject Preference Looks changed as the story progressed (i.e., Story Position) for both adults and children. The contour plots can be read like a topographic map with peaks and valleys, where this time dark blue indicates an object preference and yellow indicates a subject preference. Adults' Subject Preference Looks decreased as Bin progressed (meaning the strongest subject preference occurred at the pronoun onset), however, this likelihood decreased in a nonlinear fashion as the story unfolded over time (Story Position). For example, in the left panel of Figure

5.4 it can be seen that ~100 seconds into the story Subject Preference Looks decreased as Bin number increased, as reflected by the colour changing from yellow to blue as you move from left to right. However, ~150 seconds into the story and onwards, there was no clear relationship between Subject Preference Looks and Bin. Children's Subject Preference Looks increased as Bin progressed, however, the time course was not uniform throughout the story. For example, in the right panel of Figure 5.4 it can be seen that Subject Preference Looks increased as Bin progressed, as reflected by the colour changing from blue/green to yellow as you move left to right. However, that transition occurs much more quickly in the first half of the story (between bins 5 and 10; i.e, ~ 415-830 ms) compared to the second half of the story (between bins 15 and 20; i.e., ~ 1250-1660 ms).

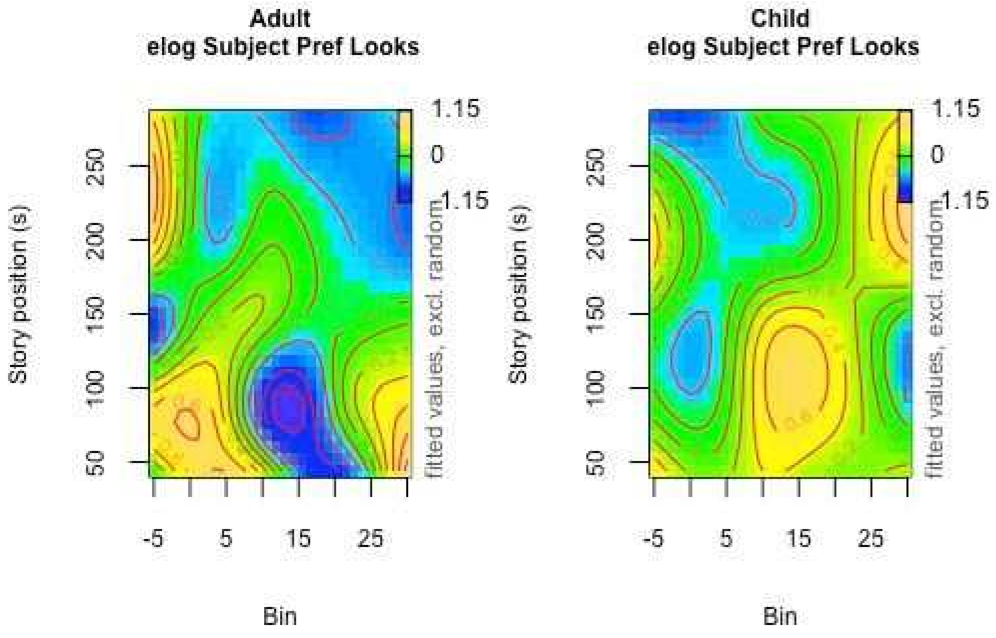


Figure 5.4: Contour plots of the interaction between Bin and Story Position by Group in Subject Preference Looks. Dark blue indicates an object preference and yellow indicates a subject preference.

To summarize, upon hearing a pronoun at the beginning of the story, adults' looks to the subject of the preceding clause decreased as (time) Bin progressed, meaning that adults' subject preference was strongest at pronoun onset. However, as the story progressed this likelihood decreased, such that there was not a clear relationship between the mention of a pronoun and adults' eye gaze patterns. Conversely, upon upon hearing a pronoun at the beginning of the story, children's looks to the subject of the preceding clause increased as (time) Bin progressed. However, as the story progressed this increase in looks to the subject took longer.

5.3. Individual Differences

5.3.1. Working memory

In order to explore the relationship between working memory (WM) capacity and the online processing of reference as it occurs within a continuous discourse, we ran models similar to those reported above but also included the measures of WM capacity, namely the Listening Recall task and the Nebraska Barnyard task. As can be seen in Figure 5.5, there was a significant correlation between the two measures of WM capacity, and thus, separate models were ran for each.

Overall we found that for both the full NPs and ambiguous pronouns there was a relationship between children's eye gaze patterns and both measures of WM. Given that the models ended up being nearly the same between the two different WM measures, we will only report the models that included Listening Recall scores. Furthermore, because we were primarily interested in the eye gaze patterns with respect to mention of ambiguous pronouns, the full NP model can be found in the Appendix. The Listening Recall pronoun model is reported below.

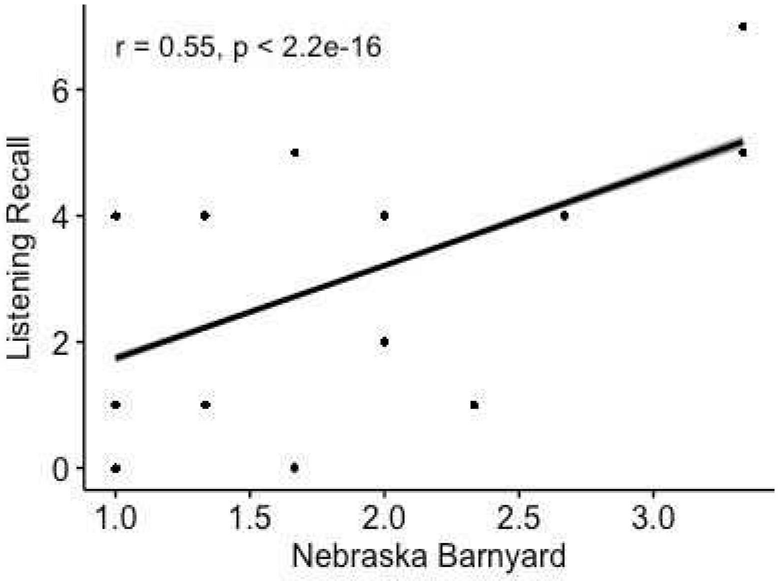


Figure 5.5: The correlation between Nebraska Barnyard Task and Listening Recall Task. Both of which are designed to assess children's working memory capacity.

5.3.1.1. *Variables*. The input variables to the model were as follows: The primary variable of interest was (time) *Bin* (where Bin 0 is the onset of the referring expression). In addition to this, *Working Memory (WM;* i.e., raw *Listening Recall* score) and *Referring Expression Position* (how far into the story the referring expression occurred, this time as a factor, i.e., first half vs second half of the story) were included. The response variable of the model was Subject Preference, which was calculated by taking the difference between empirical logit looks to the subject of the preceding clause and empirical logit looks to the object of the preceding clause. Positive values indicate a subject preference and negative values indicate an object preference.

5.3.1.2. *Model-fitting and evaluation*. The input variables above were fit to the response variable (Subject Preference). Interactions between Bin and WM, and Referring Expression Position and

WM were included in the model as nonlinear smooths (Wood, 2006). A 3-way interaction between Bin, WM and Referring Expression Position was included in the model using a tensor product smooth (Baayen, 2010b). Random slopes and intercepts for Event (a combination of subject and item) were also included. To account for autocorrelation an AR1 model was included by specifying the rho parameter and starting point for each time series. Because the response variable was the difference in looks between two IAs, weights were not included, as it is not clear how they should be calculated.

Here, the model fitting procedure was the same as in the previous models. During the fitting process, the interaction between WM and Referring Expression Position was removed, as it did not contribute significantly to the model. Although nonsignificant, the main effects of WM and Referring Expression Position remained in the final model, as they entered into significant interactions with Bin. The final model accounted for 51.4% of the deviance explained and the full model is presented in Table 5.3.

Table 5.3
Generalized additive mixed model for WM and Subject Preference Looks, reporting parametric coefficients (Part A) and the estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), F- and p- Values for smooth terms (Part B).

A. Parametric coefficients	Estimate	Std. Error	t-Value	p-Value
Intercept	-0.1571	0.4392	-0.358	0.721
RefPosSecond	-0.4131	0.4619	-0.894	0.371
WM	0.1320	0.1039	1.207	0.227
B. Smooth terms	**EDF**	**Ref.df**	**F-Value**	**p-Value**
s(Bin):RefPosFirst	6.136	7.294	8.279	2.55e-10 ***
s(Bin):RefPosSecond	1.031	1.062	3.850	0.0448 *
ti(Bin,WM):RefPosFirst	13.567	16.993	4.617	8.83e-10 ***

ti(Bin,WM):RefPosSecond	15.889	19.528	3.998	5.58e-09 ***
s(Bin,Event)	121.158	137.000	379.109	< 2e-16 ***
s(Event)	126.822	137.000	397.303	< 2e-16 ***

The GAMM fitted to Subject Preference Looks revealed a nonlinear relationship between Subject Preference Looks and Bin for both items in first and second half of the story. The partial effect of Bin (i.e., the isolated effect of the interaction when all other variables are held constant) for both first and second half items can be visualized in the upper panels of Figure 5.6. Subject For items in the first half, Subject Preference Looks increased from pronoun onset until approximately bin 15 (~1250 ms), after which Subject Preference Looks started decreasing. For items in the second half, Subject Preference Looks appeared to increase across the entire time window. However, as can be seen by the large confidence intervals in the right panel of Figure 5.6, the model was not confident about the relationship between Bin and Subject Preference Looks for items in the second half of the story. The summed effect of Bin (i.e., the effect when all the partial effects are summed) for items in the first and second half of the story can be visualized in the lower left panel of Figure 5.6. The difference between the summed effect of Bin for items in the first and second half of the story can be visualized in the lower right panel of Figure 5.6. The dashed red lines indicate that the difference between items in the first and second half of the story was significant between approximately bins 13-17 (~ 1080-1415 ms). There was a significant 3-way interaction between Bin, WM, and Referring Expression Position, which can be visualized in Figure 5.7. The contour plots show how the relationship between Bin and Subject Preference Looks changed as WM capacity increased, for both items in the first and second half of the story.

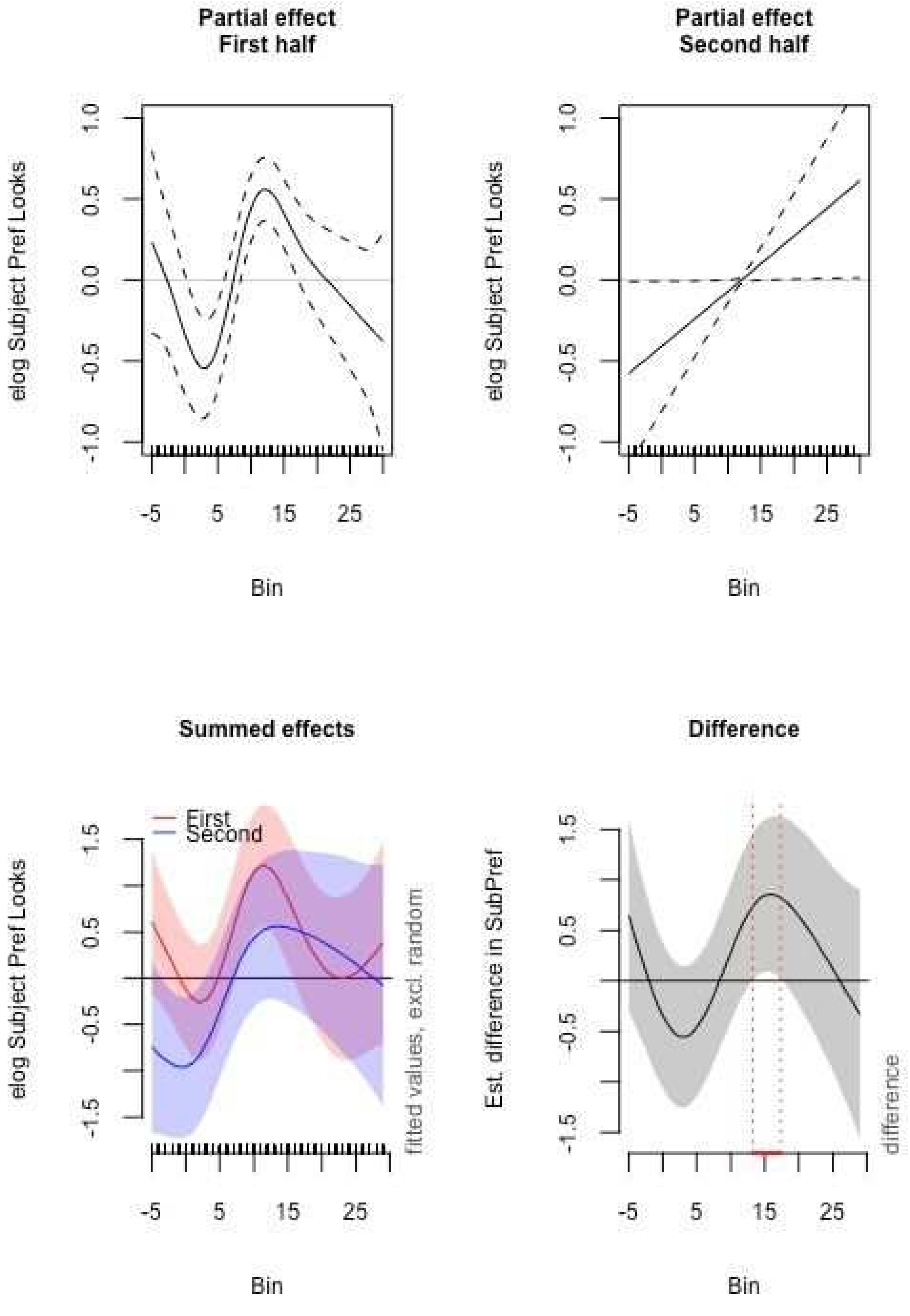


Figure 5.6: *Upper panels*. Partial effect of Bin in Subject Preference Looks for items in the first (left) and second (right) half of the story. The solid horizontal line represents the zero effect and dashed lines represent 95% confidence bands of the regression line. *Lower left panel*. Summed effect of Bin for items in the first (red) and second (blue) half of the story. *Lower right panel*. Difference between the summed effect of Bin for items in the first and second half of the story. The dashed lines represent where the difference is significant

The contour plots can be read like a topographic map with peaks and valleys, where dark blue indicates an object preference and yellow indicates a subject preference. For items in the first half, Subject Preference Looks increased as Bin progressed, for children with moderate to high WM capacity, such that they displayed a subject preference between approximately bins 5-20 (~ 415-1660 ms). However, for children with low WM there was no relationship between Bin and Subject Preference Looks. For example, in the left panel of Figure 5.7 it can be seen that Subject Preference Looks increased as Bin progressed for children who scored at least a 3 on the Listening Recall task, as reflected by the colour changing from green to yellow in the upper half of the plot. However, for children who scored lower than 3 on the Listening Recall task there was no indication of increasing Subject Preference Looks, as reflected by the relatively solid green colour in the lower half of the plot. For items in the second half, Subject Preference Looks decreased as Bin progressed for children with high WM capacity (meaning that they had the strongest subject preference at the pronoun onset). For children with low to moderate WM capacity, Subject Preference Looks increased starting approximately between bins 5 and 17 (~415-1415 ms) and evidence of a subject preference did not appear until approximately bin 25 (~ 2000 ms). For example, in the right panel of Figure 5.7 it can be seen that Subject Preference Looks decreased as Bin progressed for children who scored higher than a 5 on the Listening Recall task, as reflected by the colour changing from yellow to blue in the upper half of the plot. However, for children who scored lower than 5 on the Listening Recall, Subject Preference Looks increased as Bin progressed, as reflected by the colour primarily changing from blue to green.

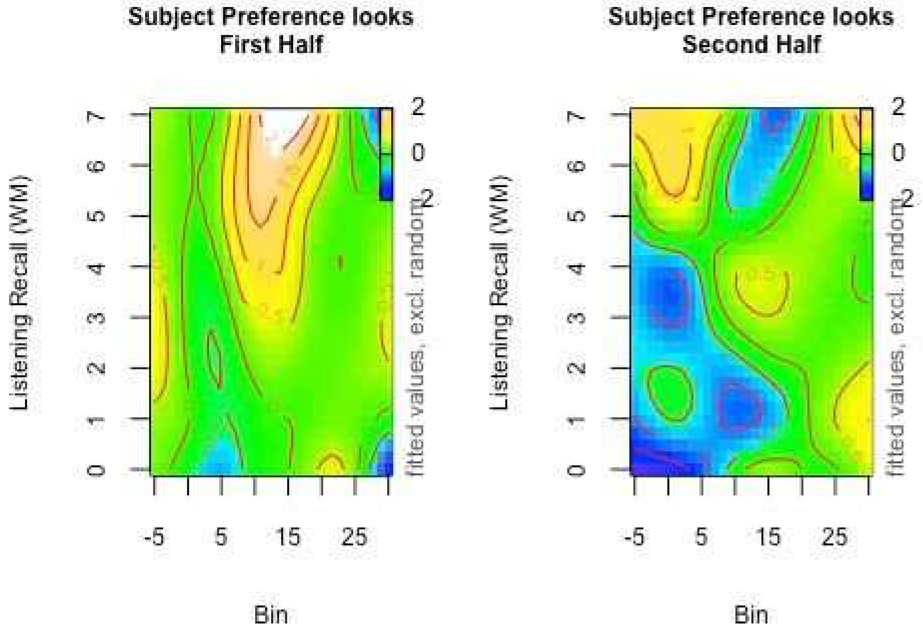


Figure 5.7: Contour plots of the interaction between Bin and Working Memory by Referring Expression Position in Subject Preference Looks. Dark blue indicates an object preference and yellow indicates a subject preference.

To summarize, upon hearing a pronoun in the first half of the story, children with moderate to high WM capacity displayed an increase in looks to the subject of the preceding clause as (time) Bin progressed. Children with low WM capacity showed no evidence of a relationship between looks to the subject of the preceding clause and (time) Bin. Upon hearing a pronoun in the second half of the story, children with high WM capacity displayed a decrease in looks to the subject of the preceding clause as (time) Bin progressed, meaning that they had the strongest subject preference at pronoun onset. Children with low to moderate WM capacity displayed an increase in looks to the subject of the preceding clause as (time) Bin progressed.

5.3.2. Vocabulary Knowledge

In order to explore the relationship between vocabulary knowledge and the online processing of reference as it occurs within a continuous discourse, we ran models that included children's scores on the PPVT-4 (a standardized vocabulary assessment). Overall we found that for both the full NPs and ambiguous pronouns there was a relationship between vocabulary knowledge and children's eye gaze patterns. However, given that working memory capacity appeared to better predict children's eye gaze patterns, the PPVT modes can be found in the Appendix (for sake of brevity).

Chapter 6: Discussion

The current thesis assessed the application of the visual world eye tracking paradigm (VWP) as a tool for investigating online language comprehension in a naturalistic setting. Previous applications of the VWP have demonstrated that eye movement patterns reflect the underlying cognitive processes involved in real-time spoken language comprehension, when the linguistic input is related to the visual scene. However, the majority of these studies focused on relatively isolated items, thus, it is unclear whether these same comprehension processes would be captured by VWP eye movement patterns in a continuous discourse setting. We analyzed children and adults' eye gaze patterns with respect to the onset of referring expressions, both full noun phrases (NPs) and pronouns, that occurred throughout a five minute long storybook. Overall, we found that eye movement patterns differed between NPs and pronouns, as well as between children and adults. Furthermore, we found that eye movement patterns differed between the beginning of the discourse compared to the end of the story. In other words, the eye gaze pattern that occured after the mention of a referring expression differed depending on when in the story that referring expression occurred. Although this was the case for both NPs and pronouns, as well as for both adults and children, different conclusions can be made with respect to each unique combination (i.e., NPs x Adults; NPs x Children; Pronouns x Adults; Pronouns x Children).

Since the seminal work of Cooper (1974) it has been assumed that people naturally direct their eye gaze towards entities in the visual scene that are (semantically) related to the meaning of the language being heard, and that this mapping reflects active online processes during language comprehension. The processing of full NPs reflects a one to one mapping between the linguistic input and corresponding entities in the visual scene. Thus, it is assumed that when an individual hears the name of a character, he or she will look at the corresponding character in the visual scene. We found that during the first half of the story, when children and adults heard a character name, their looks towards the corresponding character in the visual scene increased, as expected. However, during the second half of the story, children and adults' eye movement patterns were essentially unaffected by the mention of a character name. These findings demonstrate that even within a single task and under the most basic assumptions, there is not a uniform relationship between linguistic input and eye movements within the visual scene. Therefore, these initial assumptions may largely oversimplify such a relationship. There are a few possible explanations as to why eye gaze patterns differed at the beginning of story compared to at the end of the story. First, it is possible that the underlying processes involved in comprehension of full NPs at the beginning of a discourse are different than the underlying processes involved in comprehension of full NPs at the end of a discourse, which gets reflected in the eye gaze patterns. Another possibility is that the underlying processes involved are the same, but the relationship between the linguistic input and the visual scene is different at the beginning of the discourse compared to the end of the discourse. However, before we can compare these distinct explanations, we first must consider how we think eye movement patterns reflect the underlying cognitive processes involved in spoken language comprehension (i.e., the linking hypothesis). We will return to this discussion later on.

Our findings are in line with those of Engelen and colleagues (2014), who also investigated referential processing during continuous discourse comprehension and found that children were less likely to fixate on target referents as the story progressed. They suggested that the visual scene may be particularly useful when first building a mental representation of the discourse, such that listeners search for appropriate referents in the visual scene. However, once a mental representation is well established, the visual scene does not provide any additional information. They also found that good comprehenders were even less likely to fixate on the targets as the story unfolded over time. Although we found that for both children and adults the likelihood of fixating on the intended referent decreased as the story progressed, this happened sooner for adults. Thus, the good comprehenders in Engelen et al. (2014), may be behaving more 'adultlike', in that they are quicker at establishing and better at maintaining a sufficient mental representation of the discourse.

We also explored whether the VWP can be used to investigate inferential processes involved in discourse comprehension, namely ambiguous pronoun resolution. Unlike with NPs, there is not a direct one to one mapping between the linguistic input and entities in the visual scene when it comes to pronouns. In other words there is no referent named 'he', rather the listener must make an inference about who 'he' is referring to (i.e., who the antecedent is).

Previous VWP pronoun resolution studies have found that both children and adults have a preference for an ambiguous pronoun to co-refer with the subject of the preceding clause, as reflected by an increased proportion of eye movements to the subject as compared to the object (e.g., Arnold et al., 2000; Järvikivi, et al., 2005; Hartshorne, et al., 2010; Pyykkönen, et al., 2010). However, all of these previous studies used isolated experimental items (i.e., items were only 2-4 sentences long), with only two characters present in the visual scene. Thus, it was largely unknown how VWP online pronoun resolution works within a continuous discourse containing multiple characters both in the linguistic input and visual scene.

As with the NPs, eye movement patterns differed depending on where within the story the pronoun occured. Although this was the case for both children and adults, the patterns were largely different. At the beginning of the story, adults had the strongest subject preference right at the pronoun onset, suggesting that they had predicted that the subject was going to be referred to (under the assumption that it takes ~200 ms to plan a saccadic eye movement). To already be looking at the subject when the pronoun is mentioned means that adults must have had the expectation that that referent would be referred to before it actually was. Then the mention of the pronoun confirmed that expectation and adults were able to move on to process the next bit of information. This sort of prediction can be seen as compatible with the findings of Altmann and Kamide (1999), who showed that when participants listened to sentences such as ‘the boy will *eat* the cake’, they fixated on the ‘cake’ in the visual scene upon hearing the verb *eat.* However, one difference in the case of pronouns is that the prediction is not that there will be an upcoming ‘he’, rather the prediction is that the referent will be referred to, which could be in the form of an NP or pronoun. Furthermore, this finding is also consistent with the findings of Engelen et al.

(2013), who found that good comprehenders were more likely to make anticipatory fixations than poor comprehenders. They suggested that this may be because good comprehenders are better at monitoring and applying forward-looking cues in order to anticipate the referents, which is again in line with the notion that the good comprehenders display more 'adultlike' processing. Studies have shown that children's productive vocabularies and word reading skills predict their ability to anticipate upcoming spoken linguistic input (e.g., Mani & Huettig, 2012; 2014), which may help explain why good comprehenders in Engelen et al. (2014) were more likely to anticipate upcoming referents (if one presumes that good comprehenders also have larger productive vocabularies and better word reading skills than poor comprehenders).

The previous VWP pronoun resolution studies were not designed to investigate the ability to anticipate the upcoming mention of the referent (i.e., the subject), and therefore it is unclear whether the ability to do so is unique to continuous discourse processing. However, one could speculate that the ability to appropriately predict upcoming referents may be greater in a continuous discourse, given that additional discourse cues are available (see Arnold, 1998 for review).

For pronouns that occurred during the second half, there was not a clear relationship between pronoun onset and adults' eye gaze patterns, which again suggests that there is not a uniform relationship between the linguistic input and eye movements within the visual scene. If the visual scene is particularly useful when first building a mental representation of the discourse, but not as useful once that mental representation is already built, it makes sense that eye movement patterns would differ throughout an unfolding discourse. So in the case of the previous VWP pronoun resolution studies, every item would require the listener to build a new mental representation (since the items do not relate to one another), and thus the eye gaze patterns of each item would be assumed to reflect the same underlying processes. Conversely, for our items and those in Engelen et al. (2014), only a single mental representation (or situation model) would be required; and although it would still need to be updated, it crucially would not need to be built from 'scratch' for each item. Thus, the eye gaze patterns may be reflecting the fact that different processes are involved in building versus maintaining a mental representation of the discourse, or at least that the relationship between the linguistic input and eye movements within the visual scene is different when building/updating versus maintaining a mental representation. In the current study, eye movement patterns that were observed for pronouns that occured at the beginning of the story, may be more similar to the eye movement patterns (and corresponding underlying processes) reported in previous VWP pronoun resolution studies. This does not necessarily mean that eye gaze patterns for items that occured later in the discourse are arbitrary (or not meaningful), but just as with the NPs, may reflect different underlying processes that are not as well understood.

In contrast to adults, children did not show evidence of prediction. During the first half of the story children's eye movement record first showed evidence of a subject preference ~ 400 ms after the pronoun onset. This is quite early compared to previous VWP pronoun resolution studies, where evidence of a subject bias sometimes did not show up in children's eye movement record until 1200 ms after the pronoun onset. It is likely that additional discourse cues in the present study influenced the salience of the subject, and children were able to use these cues to fixate on the subject sooner than in previous VWP studies. During the second half of the story, children's looks towards the subject did not start increasing until ~1250 ms after the pronoun onset and did not show evidence of a subject preference until almost 2000 ms after the pronoun onset. This shows that although children may be slower at resolving the ambiguity (i.e., determining the antecedent of the pronoun) during the second half of the discourse, they are still sensitive to the mention of the pronoun. This is interesting given the fact that adults did not appear to be sensitive to the mention of the pronoun during the second half of the story, or at least their eye gaze patterns did not suggest so. Again, just because the eye movement record does not provide evidence for a certain phenomenon, does not mean that phenomenon does not exist. In other words, adults may also be (and likely are) sensitive to the mention of the pronoun during the second half of the discourse, but it is just not reflected in their eye gaze patterns. Nevertheless, the fact that such a sensitivity showed up in the eye movement record of children and not adults, suggests that children and adults differ their processing of ambiguous pronouns in a continuous discourse. There are a few explanations as to why evidence of a subject bias surfaced much sooner for pronouns that occured in the first half of the story than for pronouns that occurred later on. One possibility is that solving (ambiguous) referential relations during the later parts of a discourse, requires greater working memory (WM) capacity, given that more has taken place and therefore, there has been more to keep track of. This may especially be the case in the current study, given that the number of characters present in the story increased as the story progressed, hence, adding to the complexity. Furthermore, we know that increasing the number of referents in the visual scene influences VWP eye gaze patterns (e.g., Ferreira et al., 2013), so perhaps increasing the number of potential referents in the visual scene made it more difficult for children to fixate on the appropriate referent.

We explored whether there was a relationship between children's WM capacity and their eye gaze patterns throughout a continuous discourse. More specifically we were interested in whether WM capacity could partially explain children's subject preference looks after an ambiguous pronoun. We found that for pronouns that occurred in the first half of the story, children with moderate to high WM capacity displayed a subject preference ~ 400 ms after the pronoun onset. Children with low WM capacity (i.e., a listening recall raw score of < 3) did not show any evidence of a subject preference in the first half of the story. For pronouns that occurred in the second half of the story, children who scored at the highest end of WM capacity (i.e., a listening recall raw score of > 5) actually appeared to be predicting the mention of the subject, such their subject preference was strongest at pronoun onset (similar to adults with pronouns in the first half). For children with low to moderate WM capacity, looks to subject slowly increased starting between ~400-1250 ms after the pronoun onset, however there was no reliable evidence of a subject preference until almost 2000 ms after the pronoun onset. Overall, there is evidence that there is a relationship between WM capacity and ambiguous pronoun resolution within a continuous discourse. It was slightly surprising that children with high WM appeared to predict the mention of the subject in the second half of the story. However, as previously alluded to, it is possible that the more cues there are available, the easier it is to predict upcoming referents. Children with with high WM may have been better able to take advantage of additional cues (or perhaps the strength of certain cues increases), in order to predict the mention of the subject. For example, the longer a referent is the topic of the discourse (i.e., discourse status), the more salient that referent becomes; if a listener is able to use this information and integrate it with additional information, he or she may be more likely to predict that that referent will be referred to again. Unlike children with higher WM capacity, children with lower WM capacity may not have been able to do so. This does not explain why adults showed prediction for pronouns in the first half of the story but not in the second half. However, it is still unclear whether adults' eye movement patterns are reflective of their language processing during later parts of the discourse, which will be discussed further below. It should be noted that the WM analysis was based off of 14 children and 10 pronouns, thus we must err of the side of caution when interpreting the results. Collecting additional data would allow for greater statistical power, but nonetheless the findings are promising.

To summarize thus far, it is clear that eye movement patterns differed for items at the beginning of the discourse compared to items at the end of the discourse, and as such we can be fairly certain that there is not a uniform relationship between the linguistic input and eye movements within the visual scene. If we restrict our focus to the first half of the story, we can conclude that eye movement patterns differed between names and pronouns. This is expected given that names and pronouns are used for different purposes, and thus it is assumed that different mechanisms are uniquely involved in the processing of the two. Furthermore, we can conclude that in the case of the pronouns, children and adults' eye gaze patterns differed. This is also expected given that the processing of pronouns requires inference, and adults are likely better at making inferences than children.

Finally, why were eye movement patterns different in the second half of the story? One possibility is that the underlying processes involved in spoken language comprehension at the beginning of a discourse are different from underlying the processes involved in spoken language comprehension at the end of the discourse, and this gets reflected in the eye gaze patterns. Another possibility is that underlying cognitive processes involved in spoken language comprehension are the same, but there is not a uniform relationship between the linguistic input and eye movements within the visual scene across the entirety of the discourse. Whether we believe it to be the former versus latter explanation partially depends on how we think eye movement patterns reflect the underlying processes involved in spoken language comprehension. For example, Altmann & Mirkovic (2009)'s *joint representation of linguistic meaning and visual information* account argues the anticipated linguistic meaning and information from the visual scene interact with each other in such a way that they are indistinguishable from each other, such that updating occurs on the joint representation of the two. If the anticipated linguistic meaning is 'indistinguishable' from visual scene information, then a difference in eye movement patterns would suggest a difference in processing. However, there is reason to believe that this is not the case. At this point it should be noted that the results from the current thesis cannot say for certain, but rather can only be used to speculate. Nonetheless, let us consider the case of the NPs. During the first half of the story, upon hearing an NP, both children and adults looked towards the corresponding character in the visual scene. However, during the second half of the story, the mention of an NP had no influence on adults' or children's eye gaze patterns. It is hard to believe that comprehending 'bear' at the end of the discourse would involve some underlyingly different processes than comprehending 'bear' at the beginning of the story. Thus, the more likely explanation may be that the supporting role of visual scene decreases once a sufficient mental representation has been established (as suggested in Engelen et al., 2014).

To build upon this idea, let us consider why the visual scene is particularly useful when first building a mental representation of the discourse. It is likely the case that the visual scene serves as a scaffold for building a mental representation, and therefore eye movements are closely time-locked to the linguistic input as we try to ‘set the scene’ so to speak. However, once this mental representation has been established, we no longer need the scaffolding. Compare a picture book to an audiobook. In the case of an audiobook (or reading a novel) the listener must create his or her own mental imagery of the discourse and although there are processing costs associated with using mental imagery, it actually results in better comprehension (Dennis, 1982). In the case of a picture book the imagery is already provided (in the form of an illustration), and the listener can use the provided visual scene to save the processing costs associated with having to create one mentally. So why do we use the visual scene more at the beginning? Still images can only provide so much information, meaning that each illustration in a storybook only depicts a single ‘state’ or event, however, this does not align with the linguistic input, in which multiple events takes place. Therefore, there are many points throughout the discourse where the visual scene actually contrasts with the linguistic input. For example when you hear ‘Daddy Duck looked up and saw Ducking. He started swimming across the pond’, while viewing a scene that depicts Daddy Duck standing in some grass, the visual scene is probably not all that helpful. Assuming that we still use mental imagery even when a visual scene is provided, there are many cases where our mental image would be more aligned with the linguistic input than the provided visual scene. For example, if upon hearing ‘Daddy Duck started swimming across the pond’ we construct a mental image of this event happening, then that mental image would be more aligned with the linguistic input than the provided visual scene in of Daddy Duck standing in some grass. Thus, the visual scene helps us construct a mental representation of the discourse, but at some point our mental imagery becomes more useful than the visual scene, because unlike the visual scene it is dynamic. This could explain why adults' eye gaze patterns during the second half of the story did not appear to have a relationship with the linguistic input (for both the NPs and the pronouns), such that adults are likely better at using mental imagery than children and therefore were less reliant on the visual scene.

Thus, VWP eye movements that are time-locked with the linguistic input may not actually reflect the underlying cognitive processes involved in spoken language comprehension, rather they may represent a unique aspect of online language comprehension where listeners can use the visual information to aid processing. This would also explain why children's eye gaze patterns appeared to have a relationship with the linguistic input for a longer period of time compared to adults. In other words, eye movements within the visual scene are only closely time-locked with the linguistic input when the visual scene aids language processing; and these eye movements do not necessarily reflect the underlying cognitive processes. Therefore, it is not that the visual scene is never used once a mental representation is well-established, it is just that the eye movements are not time-locked with the linguistic input. Although at this point it is only speculation, should it be the case that eye movements do not necessarily reflect the underlying cognitive processes, one would have to question the validity of using the visual world paradigm as a tool for investigating real-time spoken language processing.

6.1. Challenges and Future Directions

Given that the present study used a novel adaptation of the visual world eye tracking paradigm, it was not without its challenges. Although the study was designed with hopes of using child adapted eye tracking glasses, we ended up having to use adult-sized eye tracking glasses with the children. This resulted in a substantial amount of ‘track loss’ for the children and ultimately led to more than 50% of the children we collected data on being excluded from the analyses. Furthermore, because we did not create experimental items prior to writing the story, we ended up with a lot fewer experimental items than we had hoped, especially in the case of the pronouns. The reason we did not want to create experimental items and then write a story around those items is because we wanted to investigate the processing of reference in as naturalistic of a setting as possible. Taken together, these two issues resulted in us having a much smaller dataset than we had anticipated. Had we had a larger dataset, we would have had the statistical power to run more complex models (i.e., include multiple measures of executive functioning in a single model) and also be more confident in the interpretation of results. Future research is needed in order to gain a better understanding of the relationship between eye movements and spoken language input, especially within a continuous discourse. With a larger sample size and more experimental items, future studies could investigate eye gaze patterns using a storybook, but instead compare eye gaze patterns between items where the visual scene is actually depicting the unfolding event to items the visual scene has does not depict the unfolding event (but still has relevant characters). Future studies could also compare eye gaze patterns between dynamic versus still visual scenes, using the same linguistic input. For example if the visual scene were a dynamic cartoon depicting the events described by the linguistic input, we would expect eye movements to be closely time-locked to the visual scene throughout the entire discourse, unlike in the current study that used still visual scenes.

Nonetheless, these findings demonstrate the importance of investigating language processing under naturalistic conditions, as the findings of highly controlled studies are not always generalizable. Future research is needed to determine whether or not the VWP should be used as a tool for investigate ‘real’ language processing under naturalistic settings.

References

Allopenna P., Magnuson J.S., & Tanenhaus M.K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38 (4)*, 419-439.

Altman, G. T. M., Clifton, C., Jr., & Mitchell, D. C. (1998). Lexical Guidance in sentence processing. *Psychonomic Bulletin & Review,* 5, 265-270.

Altman, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language*, 31, 685-712.

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

Altmann, G. M., & Kamide, Y. (2007). The Real-Time Mediation of Visual Attention by Language and World Knowledge: Linking Anticipatory (and Other) Eye Movements to Linguistic Processing. *Journal Of Memory And Language*, *57*(4), 502-18. doi:10.1016/j.jml.2006.12.004

Altmann, G. T., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, *111*55-71. doi:10.1016/j.cognition.2008.12.005

Altmann, G. M., & Mirković, J. (2009). Incrementality and Prediction in Human Sentence Processing. *Cognitive Science*, *33*(4), 583-609. doi:10.1111/j.1551-6709.2009.01022.x

Arnold, J. (1998). Reference form and discourse patterns. PhD dissertation, Stanford University, Palo Alto, CA.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The immediate use of gender information: Eye-tracking evidence of the time-course of pronoun resolution. Cognition, 76, B13–B26

Arnold, J. E., Strangmann, I. M., Hwang, H., Zerkle, S., & Nappa, R. (2018). Linguistic experience affects pronoun interpretation. *Journal Of Memory And Language*, *102*41-54. doi:10.1016/j.jml.2018.05.002

Barr, D. (2012). Walkthrough of an "empirical logit" analysis in R. [Website]. http://talklab.psy.gla.ac.uk/tvw/elogit-wt.html.

Baayen, R. H. (2010a). Demythologizing the word frequency effect: a discriminative learning perspective. The Mental Lexicon, 5(3), 436–561.

Baayen, R. H. (2010b). The directed compound graph of English. An exploration of lexical connectivity and its processing consequences. In S. Olson (Ed.), New impulses in word-formation (Linguistische Berichte Sonderheft 17) (pp. 383–402). Hamburg: Buske.

Berends, S., Brouwer, S., & Sprenger, S. (2015). Eye-Tracking and the Visual World Paradigm. 55-78. 10.1007/978-3-319-11529-0_5.

Blything, L. P., & Cain, K. (2016). Children's processing and comprehension of complex sentences containing temporal connectives: The influence of memory on the time course of accurate responses. *Developmental psychology*, *52*(*10*), 1517.

Blything, L. P., & Cain, K. (2016). Children's processing and comprehension of complex sentences containing temporal connectives: The influence of memory on the time course of accurate responses. *Developmental psychology, 52*(*10*), 1517.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.42.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's Reading Comprehension Ability: Concurrent Prediction by Working Memory, Verbal Ability, and Component Skills. *Journal of Educational Psychology, 96*(1), 31-42. http://dx.doi.org/10.1037/0022-0663.96.1.31

Carlson, K., Clifton Jr., C., & Fraizer, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, *16*, 58-81.

Carreiras, M., Gernsbacher, M. A., & Villa, V. (1995). The Advantage of First Mention in Spanish. *Psychonomic Bulletin & Review: A Journal Of The Psychonomic Society, Inc*, *2*(1), 124-29. doi:10.3758/BF03214418

Carretti, B., Cornoldi, C., De Beni, R., & Romanò, M. (2005). Up- dating in working memory: A comparison of good and poor comprehenders. Journal of Experimental Child Psychology, 91, 45–66. doi:10.1016/j.jecp.2005.01.005

Chambers, C. G., & Smyth, R. (1998). Structural parallelism and discourse coherence: A test of centering theory. Journal of Memory and Language, 39(4), 593-608.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains in real-time sentence comprehension. Journal of Memory and Language, 47, 30−49.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Action-based affordances and syntactic ambiguity resolution. Journal of Experimental Psychology: Learning, Memory & Cognition, 30, 687−696.

Chevalier, N., James, T., Nelson, J., Espy, K., & Wiebe, S. (2014). Contribution of reactive and proactive control to children's working memory performance: Insight from item recall durations in response sequence planning. *Developmental Psychology*, *50*(7), 1999-2008. doi:10.1037/a0036644

Chevalier, N., Sheffield, T. D., Nelson, J. M., Clark, C. C., Wiebe, S. A., & Espy, K. A. (2012). Underpinnings of the Costs of Flexibility in Preschool Children: The Roles of Inhibition and Working Memory. *Developmental Neuropsychology*, *37*(2), 99-118. doi:10.1080/87565641.2011.632458

Clifton, C., Traxler, M. J., Williams, R., Mohammed, M., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, *49*, 317-334.

Cooper, R. M. (1974). The Control of Eye Fixation by the Meaning of Spoken Language: A New Methodology for the Real-Time Investigation of Speech Perception, Memory, and Language Processing. *Cognitive Psychology*, *6*84-107.

Crawley, R., Stevenson, R., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. Journal of Psycholinguistic Research, 4, 245–264.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. Journal of Memory and Language, 47, 292−314.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19, 450–466. doi:101016/s0022-5371(80)90312-6

Dunn, L. M., Dunn, D. M. (2007) *PPVT-4 :Peabody picture vocabulary test* Minneapolis, MN. : Pearson Assessments.

Engel de Abreu, P. M. J., Gathercole, S. E., & Martin, R. (2011). Disentangling the relationship between working memory and language: The roles of short-term storage and cognitive control. Learning and Individual Differences, 21, 569–574.

Engelen, J. A., Bouwmeester, S., de Bruin, A. B., & Zwaan, R. A. (2014). Eye movements reveal differences in children's referential processing during narrative comprehension. *Journal Of Experimental Child Psychology*, *118*57-77. doi:10.1016/j.jecp.2013.09.005

Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? Journal of Memory and Language, 54, 554−573.

Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal Of Memory And Language (Print)*, (3), 165.

Fodor, J. A. (1983). *Modularity of mind : an essay on faculty psychology.* Cambridge, Mass.: MIT Press.

Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representations. Journal of Memory and Language, 56, 357-383.

Frazier, L. (1979). On Comprehending Sentences: Syntactic Parsing Strategies. *Dissertation Abstracts: Section A. Humanities And Social Science*, *40*

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 559-586). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Frazier, L., & Rayner, K. (1982). Making and Correcting Errors during Sentence Comprehension: Eye Movements in the Analysis of Structurally Ambiguous Sentences. *Cognitive Psychology*, *14*(2), 178-210. doi:10.1016/0010-0285(82)90008-1

Frederiksen, J. (1981). Understanding anaphora: Rules used by readers in assigning pronominal referents. Discourse Processes, 4, 323–347.

Fukumura, K., & van Gompel, R. P. (2015). Effects of order of mention and grammatical role on anaphor resolution. Journal of Experimental Psychology: Learning, Memory, and Cognition, 41(2), 501.

Garnham, A., Traxler, M., Oakhill, J., & Gernsbacher, M. A. (1996). Regular Article: The Locus of Implicit Causality Effects in Comprehension. *Journal Of Memory And Language*, *35*517-543. doi:10.1006/jmla.1996.0028

Garnsey, S. M., Pearlmutter, N. J., Myers, E., Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*, 58-93.

Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal Of Educational Psychology*, *70*(2), 177-194.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*, *32*(2), 99–156.

Gernsbacher, M. A. (1990). Language comprehension as structure building. Hillsdale, NJ: Lawrence Erlbaum.

Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing Sentence Participants: The Advantage of First Mention. *Journal of Memory and Language*, *27*(6), 699–717. http://doi.org/10.1016/0749-596X(88)90016-2

Gleitman, L., 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.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. Cognitive science, 17(3), 311-347.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. Language, 69(2), 274-307.

Griffin, Z. M., & Bock, K. (2000). What the Eyes Say about Speaking. *Psychological Science*, (4), 274.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. Cognitive Science, 28, 105−115.

Hartshorne, J. K., Nappa, R., & Snedeker, J. (2010). Ambiguous pronoun processing development: Probably not u-shaped. In N. Danis, K. Mesh, & H. Sung (Eds.), BUCLD 35: Proceedings of the 35th annual Boston University Conference on Language Development (pp. 272-282). Somerville, MA: Cascadilla Press.

Hartshorne, J. K., Nappa, R., & Snedeker, J. (2015). Development of the first-mention bias. Journal of child language, 42(2), 423-446.

Hoffman, J. E., & Subramaniam, B. (1995). The Role of Visual Attention in Saccadic Eye Movements. *Perception & Psychophysics*, *57*(6), 787-95. doi:10.3758/BF03206794

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. Journal of Memory and Language, 54, 460−482.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(Visual search and visual world: Interactions among visual attention, language, and working memory), 151-171. doi:10.1016/j.actpsy.2010.11.003

Hughes, C., Dunn, J., & White, A. (1998). Trick or treat? Uneven understanding of mind and emotion and executive dysfunction in "hard-to-manage" preschoolers. Journal of Child Psychology and Psychiatry, 39, 981-994.

Järvikivi, J., Porretta, V., Paradis, J., Govindarajan, K. & Day, K. Language knowledge predicts 3-6 year-old mono- and bilingual children's pronoun processing.*14th International Congress for the Study of Child Language* (*IASCL*), Lyon, France, July 17-21, 2017.

Järvikivi, J., Pyykkönen-Klauck, P., Schimke, S., Colonna, S., & Hemforth, B. (2014). Information structure cues for 4-year-olds and adults: Tracking eye movements to visually presented anaphoric referents. *Language, Cognition And Neuroscience*, *29*(7), 877-892.

Järvikivi, J., van Gompel, R. G., & Hyönä, J. (2017). The Interplay of Implicit Causality, Structural Heuristics, and Anaphor Type in Ambiguous Pronoun Resolution. *Journal Of Psycholinguistic Research*, *46*(3), 525-550.

Järvikivi, J., van Gompel, R. P. G., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. Psychological Science, 16(4), 260-264.

Johnson-Laird, P. N. (1983). Mental models: Towards a cognitive science of language, inference, and consciousness. Cambridge, MA: Harvard University Press.

Kaiser, E. (2016). Discourse Level Processing. In P. Knoeferle, P. Pyykkönen-Klauck, M. W. Crocker (Eds.) , *Visually Situated Language Comprehension* (pp. 151-184). Amsterdam, Netherlands: Benjamins. doi:10.1075/aicr.93.06kai

Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*, (1), 55.

Kaiser, E., & Trueswell, J. C. (2008). Interpreting Pronouns and Demonstratives in Finnish: Evidence for a Form-Specific Approach to Reference Resolution. *Language And Cognitive Processes*, *23*(5), 709-748.

Karasinski, C., & Ellis Weismer, S. (2010). Comprehension of Inferences in Discourse Processing by Adolescents with and without Language Impairment. *Journal Of Speech, Language, And Hearing Research*, *53*(5), 1268-1279.

Kidd, E., Lieven, E., & Tomasello, M. (2006). Examining the role of lexical frequency in the acquisition and processing of sentential complements. *Cognitive Development*, *21*93-107. doi:10.1016/j.cogdev.2006.01.006

Knoeferle, P., & Crocker, M. W. (2006). The Coordinated Interplay of Scene, Utterance, and World Knowledge: Evidence from Eye Tracking. *Cognitive Science: A Multidisciplinary*

Journal Of Artificial Intelligence, Linguistics, Neuroscience, Philosophy, Psychology, *30*(3), 481-529. doi:10.1207/s15516709cog0000_65

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*, 153-194.

Knoeferle, P., & Crocker, M. W. (2007). The Influence of Recent Scene Events on Spoken Comprehension: Evidence from Eye Movements. *Journal Of Memory And Language*, *57*(4), 519-43. doi:10.1016/j.jml.2007.01.003

Kryuchkova, T., Tucker, B. V., Wurm, L. H., & Baayen, R. H. (2012). Danger and usefulness are detected early in auditory lexical processing: evidence from electroencephalography. Brain and Language, 122(2), 81–91, http://dx.doi.org/10.1016/j.bandl.2012.05.005.

Kuijper, S.J., Hartman, C.A., & Hendriks, P. (2015) Who Is He? Children with ASD and ADHD Take the Listener into Account in Their Production of Ambiguous Pronouns. PLoS ONE 10(7): e0132408. https://doi.org/10.1371/journal.pone.0132408

Kryuchkova, T., Tucker, B. V., Wurm, L. H., & Baayen, R. H. (2012). Danger and usefulness are detected early in auditory lexical processing: evidence from electroencephalography. Brain and Language, 122(2), 81–91, http://dx.doi.org/10.1016/j.bandl.2012.05.005.

Lõo, K., Tucker, B., Järvikivi, J., Tomaschek, F., & Baayen, R.. (2018). Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology*, *28*(1), 71-97. doi:10.1007/s11525-017-9318-7

MacDonald, M. C., Perlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. Psychological Review, 101, 676–703.

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., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. Journal of experimental child psychology, 126, 264-279.

Matin, E., Shao, K., & Boff, K. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372-380. doi:10.3758/BF03206780

McDonald, J. L., & MacWhinney, B. (1995). The Time Course of Anaphor Resolution: Effects of Implicit Verb Causality and Gender. *Journal Of Memory And Language*, *34*(4), 543-66. doi:10.1006/jmla.1995.1025

McMurray, B., Aslin, R., Tanenhaus, M., Spivey, M., & Subik, D. (2008). Gradient sensitivity to within-category variation in speech: Implications for categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, *34(6)*, 1609-1631.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the Influence of Thematic Fit (and Other Constraints) in On-Line Sentence Comprehension. *Journal Of Memory And Language*, *38*(3), 283-312. doi:10.1006/jmla.1997.2543

Meulman, N., Wieling, M., Sprenger, S. A., Stowe, L. A., & Schmid, M. S. (2015). Age effects in L2 grammar processing as revealed by ERPs and how (not) to study them. PLoS One, 10 (12), e0143328, http://dx.doi.org/10.1371/journal.pone.0143328.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. Cognition, 66(2), B25−B33.

Miyake, A., & Shah, P. (1999). *Models of working memory : mechanisms of active maintenance and executive control*. Cambridge : Cambridge University Press, 1999.

Montgomery, J. W., Magimairaj, B., & O'Malley, M. (2008). The role of working memory in typically developing children's complex sentence comprehension. Journal of Psycholinguistic Research, 37, 331–354.

Morrow, D. G. (1985). Prominent characters and events organize narrative understanding. Journal of Memory and Language, 24, 390–404.

Noldus Information Technology. (2012). The Observer XT reference manual 11.0. Wageningen, the Netherlands: Author.

Orchinik, L., Taylor, H., Espy, K., Minich, N., Klein, N., Sheffield, T., & Hack, M. (2011). Cognitive Outcomes for Extremely Preterm/Extremely Low Birth Weight Children in Kindergarten. *Journal of the International Neuropsychological Society, 17*(6), 1067-1079. doi:10.1017/S135561771100107X

Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). Research Article: The influence of gradient foreign accentedness and listener experience on word recognition. *Journal Of Phonetics*, *58*1-21. doi:10.1016/j.wocn.2016.05.006

Pyykkönen, P., Matthews, D., & Järvikivi, J. (2010).Three-year-olds are sensitive to semantic prominence during online language comprehension: A visual world study of pronoun resolution. Language and Cognitive Processes, 25, 115-129.

R Development Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rayner, K., Carlson, M., & Frazier, L. (1983). The Interaction of Syntax and Semantics during Sentence Processing: Eye Movements in the Analysis of Semantically Biased Sentences. *Journal Of Verbal Learning And Verbal Behavior*, *22*(3), 358-374. doi:10.1016/S0022-5371(83)90236-0

Schumacher, P. B., Roberts, L., & Järvikivi, J. (2017). Agentivity drives real-time pronoun resolution: Evidence from German er and der. Lingua, 185, 25-41.

Seigneuric, A., Ehrlich, M., Oakhill, J. V., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Reading & Writing*, *13*(1/2), 81-103.

Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. Journal of Verbal Learning and Verbal Behavior, 13, 272–281.

Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. Journal of Psycholinguistic Research, 23, 197–229.

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.

Song, H., & Fisher, C. (2005). Who's ‘‘she’’? Discourse prominence influences preschoolers' comprehension of pronouns. Journal of Memory & Language, 52(1), 29-57.

Song, H., & Fisher, C. (2007). Discourse prominence effects on 2.5- year-old children's interpretation of pronouns. Lingua, 117, 1959- 1987.

Spivey, M. J., & Huette, S. (2016). Toward a Situated View of Language. In P. Knoeferle, P. Pyykkönen-Klauck, M. W. Crocker (Eds.) , *Visually Situated Language Comprehension* (pp. 1-30). Amsterdam, Netherlands: Benjamins. doi:10.1075/aicr.93.01spi

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context Effects in Syntactic Ambiguity Resolution: Discourse and Semantic Influences in Parsing Reduced Relative Clauses. *Canadian Journal Of Experimental Psychology*, (2), 276. doi:10.1037/h0078826

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science*, (5217), 1632.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. Cognition, 73, 89−134.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic Influences on Parsing: Use of Thematic Role Information in Syntactic Ambiguity Resolution. *Journal Of Memory And Language*, *33*(3), 285-318. doi:10.1006/jmla.1994.1014

van der Sluis, S., de Jong, P., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. Intelligence, 35, 427–449. doi:10.1016/j.intell. 2006.09.001

van Dijk, T. A., & Kintsch, W. (1983). Strategies in discourse comprehension. New York: Academic Press.

van Rij, J., Hollebrandse, B., & Hendriks, P. (2016). Children's Eye Gaze Reveals Their Use of Discourse Context in Object Pronoun Resolution. In A. Holler, K. Suckow (Eds.) , *Empirical Perspectives on Anaphora Resolution* (pp. 267-294). Berlin, Germany: de Gruyter Mouton.

van Rij, J., Wieling, M., Baayen, R.H., van Rijn, H.: itsadug: interpreting time series and autocorrelated data using GAMMs (2015)

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. Language and Speech, 49, 367−392.

Whitely, C., & Colozzo, P. (2013). Who's Who? Memory Updating and Character Reference in Children's Narratives. Journal Of Speech, Language & Hearing Research, 56(5), 1625-1636. doi:1092-4388(2013/12-0176)

Wiebe, S.A., Sheffield, T., Nelson, J.M., Clark, C.A.C., Chevalier, N., & Espy, K.A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, *108*, 436–452.

Wood, S. N. (2006). Generalized additive models: an introduction with R (Vol. 66. Boca Raton: Chapman & Hall/CRC Press.

Yang, C. L., Gordon, P. C., Hendrick, R., & Hue, C. W. (2003). Constraining The Comprehension of Pronominal Expressions in Chinese. *Cognition*, (3), 283-315.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. (2009). Mixed effects models and extensions in ecology with R. New York: Springer.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. Psychological Bulletin, 123, 162–185.

Appendix

Storybook Dialogue

Page 1:

It was a beautiful spring day in the meadow. The sun was shining, the flowers were blooming, and **Bear** was enjoying the fresh air. He loved spending time in the meadow. All of a sudden **Bear** heard his tummy growl, so he headed into the forest to look for some berries.

Page 2:

When **Bear** got to the forest, there were so many kinds of berries that he didn’t know which to choose. There were red raspberries, small blueberries, and juicy strawberries. He decided to try a handful of each.

Page 3:

Bear sat down on a log and ate the berries. He was so full from all the eating, that he started to feel sleepy. So he headed back to the meadow to take a nap in the warm sun.

Page 4:

When **Bear** got back to the meadow, his friend **Fox** was there flying a kite. Maybe **he** wasn’t so sleepy after all! Fox ran towards **Bear**, with his kite soaring high in the sky.

Page 5:

But before he could make it to **Bear**, a big gust of wind pulled him off his feet. Oh no! What was he going to do!? “Help!” he shouted.

Page 6:

Bear ran over to help Fox. **He** jumped up, grabbed **Fox**’s paw and pulled him down to safety. Fox thanked Bear. **He** wanted to play a different game.

Page 7:

Bear suggested they play hide-and-seek. He pulled a shiny coin out of his pocket and said if it landed on heads he would hide, but if it landed on tails then Fox would hide. He threw the coin up in the air.

Page 8:

When the coin hit the ground, both Bear and **Fox** leaped to look at it. “Heads!” shouted **Bear.** “That means I get to hide first!” **Fox** started counting to five, “1...2...3...4…”

Page 9:

“5!...Ready or not here I come!” he yelled. First **Fox** looked in a patch of tall grass... but no Bear. Then he looked in a bed of flowers… but still no Bear. Finally, he looked inside a hollow log and...

Page 10:

...it was BEAR!!! Bear told Fox to go and hide. He started counting to 5, when all of a sudden he heard a small cry and looked up. “Do you hear that?” he asked. “Yeah, it sounds like someone is crying” **Fox** replied.

Page 11:

Bear told **Fox** to check the flowers, while **he** looked in the bushes. Just then, Fox spotted a **duckling** at the top of hill. **He** looked like he was crying.

Page 12:

Bear asked the duckling why he was crying. **Duckling** opened his eyes and stared at Bear and **Fox**. After a long pause he finally told them that he was lost and couldn't find his dad. "Well do you remember where you last saw your dad?" asked Fox.

Page 13:

Duckling couldn't remember. He told Bear and Fox that he was suppose to stay close but he wandered off. Fox patted Duckling's head. **He** knew they would be able to help.

Page 14:

"Do you remember where you last were?" Bear asked Duckling. **He** knew they had to take it one step at a time. **Ducking** thought for a minute. Then he remembered smelling the pretty flowers. Bear told Fox and Duckling to follow along. He knew where the flowers were.

Page 15:

When the three friends got to the flowers, Fox asked Duckling if he remembered where he was before the flowers. He hoped they were getting closer. **Duckling** thought for minute. Then he remembered playing with his friend **Frog** at the playground. "Follow me" he said.

Page 16:

The three friends made it to the playground. But before Bear and Fox could ask **Duckling** any questions, he was already too swinging on swings. Luckily, he swung so high that he spotted his friend **Frog** at the top of the slide. He jumped off the swings and headed towards **Frog**.

Page 17:

Duckling told Frog that Bear and **Fox** were helping him find Daddy Duck. He asked if **Frog** could help too. Frog told Bear, Fox and **Duckling** that before the playground they were at the pond. He told them he would take them there under one condition…they all had to hop like frogs!

Page 18:

The four friends hopped over one brown log, two golden dandelions, and three small rocks. When they got to the pond, Bear told Duckling, Fox and **Frog** to keep an eye out for Daddy Duck. He knew they had to be close.

Page 19:

But before anyone could start looking, Duckling spotted **Daddy Duck** across the pond! **He** flapped his wings with excitement. **Daddy Duck** looked up and saw Duckling. **He** sighed with relief and started swimming across the pond.

Page 20:

Duckling told Daddy Duck, that Bear, Fox, and **Frog** helped him find his way back.
Daddy Duck thanked the friends. He had been so worried. "It was no problem, we just took it one step at a time" said **Bear**. "You mean one hop at a time!" said **Frog**. And they all laughed.

Example Storybook Illustrations

Listening Comprehension Questions

1. Can you name three animals in the story?
2. Why was Duckling crying?
3. One place that the friends went, was in the meadow. Do you remember another place that they were?
4. Where did the friends find Daddy Duck?
5. How did Duckling feel when he saw Daddy Duck across the pond?

Child Participant Descriptive Statistics

Table A.1
Child Participant Descriptive Statistics

Dependent measure	n	M	SD	range
Age (years)	16	4.9	0.6	4.2-6.8
PPVT_Std	16	113.5	8.0	102-137
Nebraska Barnyard	14	1.86	0.77	1-3.33
Listening Recall	14	3.0	2.07	0-7.0

Working Memory Full NP Model

Table A.2
Generalized additive mixed model for WM and Target Looks, reporting parametric coefficients (Part A) and the estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), F- and p- Values for smooth terms (Part B).

A. Parametric coefficients	Estimate	Std. Error	t-Value	p-Value
Intercept	-0.9919	0.1173	-8.456	< 2e-16 ***
RefPosSecond	-0.5766	0.1595	-3.614	0.000304 ***
B. Smooth terms	EDF	Ref.df	F-Value	p-Value
s(Bin):RefPosFirst	4.664	5.710	13.220	4.52e-14 ***

s(Bin):RefPosSecond	1.884	2.351	1.999	0.1811
s(WM)	1.001	1.001	3.859	0.0495 *
ti(Bin, WM):RefPosFirst	12.168	15.721	3.401	6.46e-06 ***
ti(Bin, WM):RefPosSecond	12.893	16.322	2.904	8.24e-05 ***
s(Bin,Event)	457.020	515.000	393.041	< 2e-16 ***
s(Event)	481.913	515.000	422.724	< 2e-16 ***

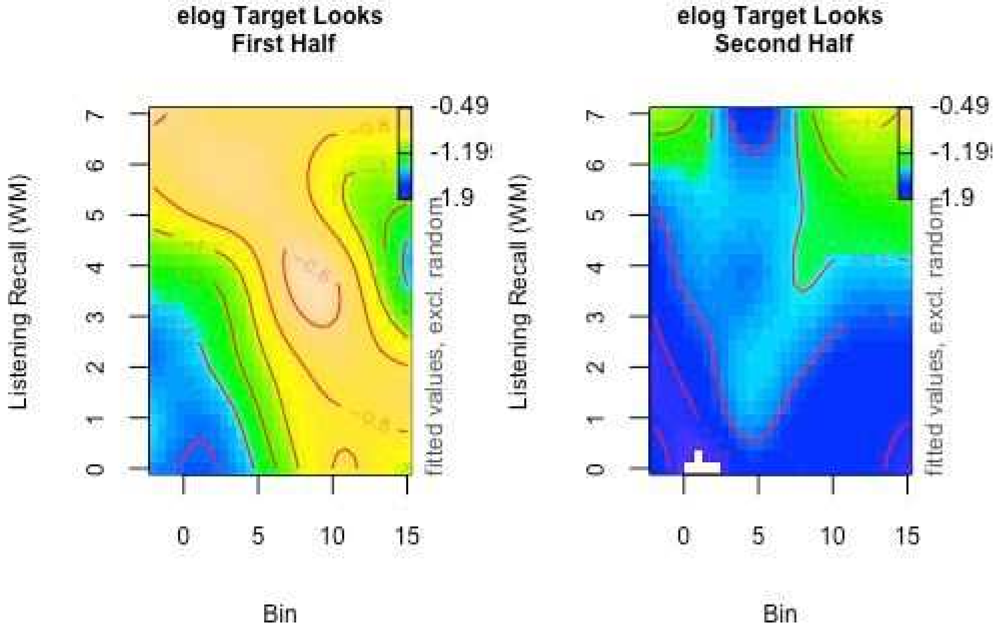


Figure A.1: Contour plots of the interaction between Bin and Vocabulary by Referring Expression Position in (elog) Target Looks. Dark blue indicates lower (elog) Target Looks and yellow indicates higher (elog) Target Looks.

Vocabulary Models

Pronouns

Table A.3
Generalized additive mixed model for PPVT and Subject Preference Looks, reporting parametric coefficients (Part A) and the estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), F- and p- Values for smooth terms (Part B).

A. Parametric coefficients	Estimate	Std. Error	t-Value	p-Value
Intercept	-1.11795	3.14296	-0.356	0.722
RefPosSecond	-0.35175	0.43525	-0.808	0.419
PPVT_Std	0.01160	0.02758	0.420	0.674
B. Smooth terms	**EDF**	**Ref.df**	**F-Value**	**p-Value**
s(Bin):RefPosFirst	5.424	6.560	7.857	6.02e-09***
s(Bin):RefPosSecond	2.015	2.515	1.999	0.0944.
ti(Bin, PPVT_Std):RefPosFirst	10.311	12.574	2.539	0.0041**
ti(Bin, PPVT_Std):RefPosSecond	11.123	13.257	4.198	5.09e-07 ***
s(Bin,Event)	138.814	157.000	372.234	< 2e-16 ***
s(Event)	145.593	157.000	393.245	< 2e-16 ***

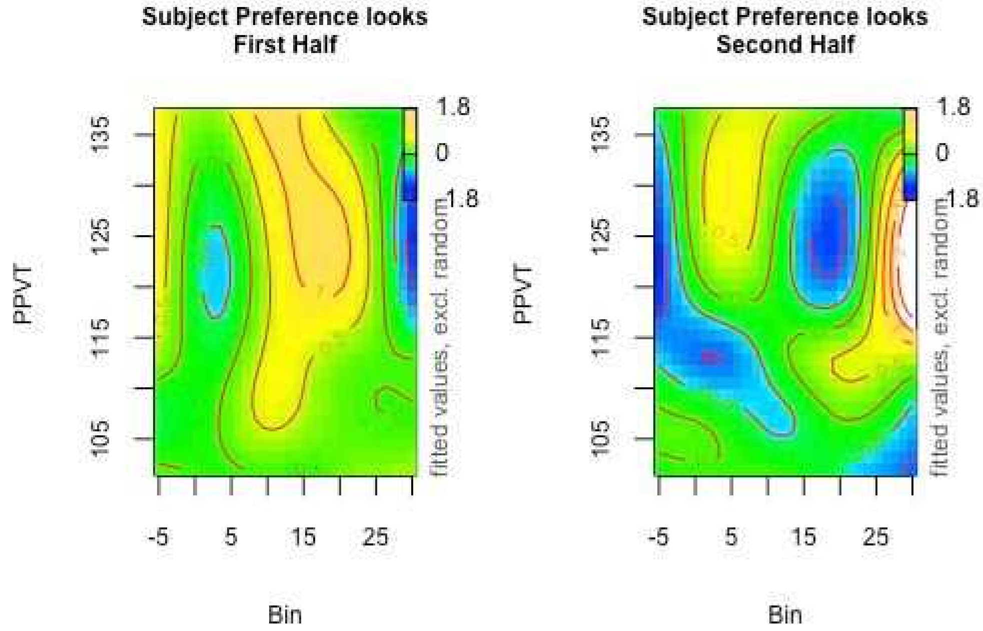


Figure A.2: Contour plots of the interaction between Bin and Vocabulary by Referring Expression Position in Subject Preference Looks. Dark blue indicates an object preference and yellow indicates a subject preference

Full NPs: names

Table A.4

Generalized additive mixed model for PPVT and Target Looks, reporting parametric coefficients (Part A) and the estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), F- and p- Values for smooth terms (Part B).

A. Parametric coefficients	Estimate	Std. Error	t-Value	p-Value
Intercept	-2.394143	1.118101	-2.141	0.032280 *
RefPosSecond	-0.586732	0.152033	-3.859	0.000115 ***
PPVT_Std	0.012911	0.009802	1.317	0.187802
B. Smooth terms	EDF	Ref.df	F-Value	p-Value
s(Bin):RefPosFirst	5.007	6.099	15.442	< 2e-16 ***
s(Bin):RefPosSecond	2.082	2.596	2.829	0.0382 *

ti(Bin, PPVT_Std):RefPosFirst	13.646	17.166	3.464	1.6e-06 ***
ti(Bin, PPVT_Std):RefPosSecond	1.038	1.074	1.967	0.1587
s(Bin,Event)	523.844	589.000	385.798	< 2e-16 ***
s(Event)	550.948	589.000	415.002	< 2e-16 ***

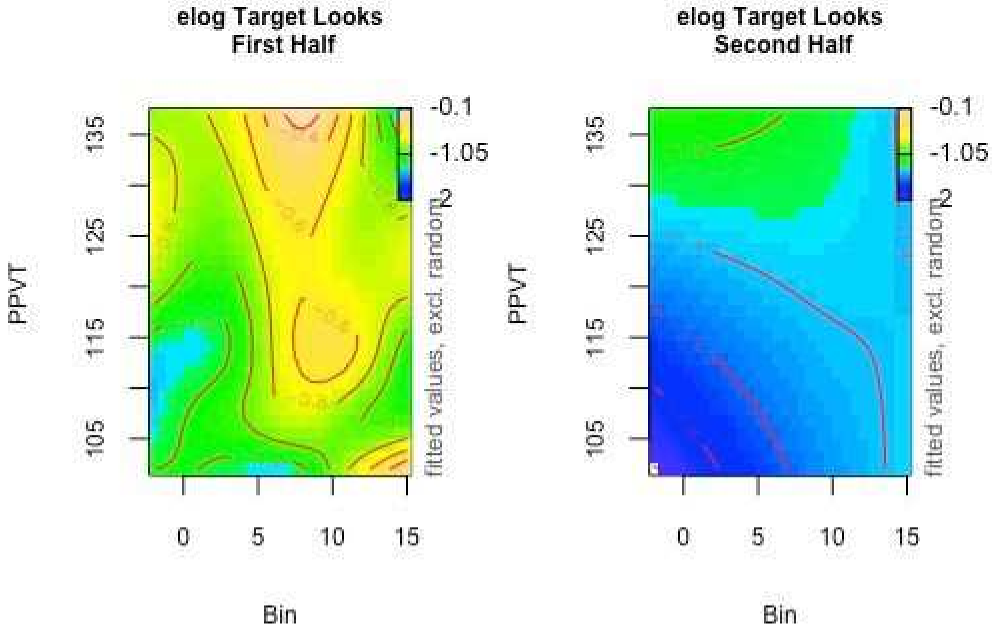


Figure A.3: Contour plots of the interaction between Bin and Vocabulary by Referring Expression Position in (elog) Target Looks. Dark blue indicates lower (elog) Target Looks and yellow indicates higher (elog) Target Looks.