Developing incrementality in filler-gap dependency processing

Emily Atkinson[1], Matthew W. Wagers[2], Jeffrey Lidz[3], Colin Phillips[3], & Akira Omaki[4]

[1]University of Michigan, [2]University of California Santa Cruz, [3]University of Maryland, [4]University of Washington

Corresponding author:

Emily Atkinson

University of Michigan, Department of Linguistics

611 Tappan Street

440 Lorch Hall

Ann Arbor, MI 48109

Email: eeatkins@umich.edu

Author Notes

This work was supported in part by an NSF grant awarded to Akira Omaki, NSF BCS-1423117 and BCS-1737736.

Abstract

Much work has demonstrated that children are able to use bottom-up linguistic cues to incrementally interpret sentences, but there is little understanding of the extent to which children's comprehension mechanisms are guided by top-down linguistic information that can be learned from distributional regularities in the input. Using a visual world eye tracking experiment and a corpus analysis, the current study investigates whether 5- and 6-year-old children incrementally assign interpretations to temporarily ambiguous *wh*-questions like *What was Emily eating the cake with __?* In the visual world eye-tracking experiment, adults demonstrated evidence for active dependency formation at the earliest region (i.e., the verb region), while 6-year-old children demonstrated a spill-over effect of this bias in the subsequent NP region. No evidence for this bias was found in 5-year-olds, although the speed of arrival at the ultimately correct instrument interpretation appears to be modulated by the vocabulary size. These results suggest that adult-like active formation of filler-gap dependencies begins to emerge around age 6. The corpus analysis of filler-gap dependency structures in adult corpora and child corpora demonstrate that the distributional regularities in either corpora are equally in favor of early, incremental completion of filler-gap dependencies, suggesting that the distributional information in the input is either not relevant to this incremental bias, or that 5-year-old children are somehow unable to recruit this information in real-time comprehension. Taken together, these findings shed light on the origin of incremental processing bias in filler-gap dependency processing, as well as on the role of language experience and cognitive constraints in the development of incremental sentence processing mechanisms.

Key words: Sentence processing, visual world, filler-gap dependency, prediction, child-directed speech

1. Introduction

How do sentence comprehension mechanisms develop over time? This developmental question has recently drawn much attention in the field of language development, as well as in sentence processing research. For language development research, the main reason for investigating parser development is two-fold. First, parsing the input is a key sub-process of language acquisition, as children must assign linguistic representations to the input first in order to infer the linguistic knowledge that allowed the speakers to formulate the input utterances (Frazier & de Villiers, 1990; Omaki & Lidz, 2015; Valian, 1990). Second, parser development provides an important testing ground for the question of nature and nurture in language development. The adult sentence processing literature has shown that comprehenders incrementally assign syntactic and semantic representations as the language input unfolds (e.g., Altmann & Kamide, 1999; Frazier & Rayner, 1982; Marslen-Wilson, 1973; Staub & Clifton Jr., 2006; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Trueswell, Tanenhaus, & Garnsey, 1994, among others), but such mental operations are not directly observable for language learners. This problem has led researchers to propose that basic architectural constraints on parsing, such as incrementality, are innately given to language learners, and that incremental processing biases will emerge as relevant knowledge of linguistic cues is acquired (e.g., Fodor, 1998; Pinker, 1996). While much work on child sentence processing has documented early presence of adult-like incremental processing mechanisms (for recent reviews, see Omaki & Lidz, 2015; Snedeker & Huang, 2016), these findings only indicate that incremental biases develop by a certain age, and do not address whether the development of biases was independent of language experience.

On the other hand, adult sentence processing research has seen a surge of interest in the relation between language experience and incremental comprehension mechanisms. For example, influential models of sentence processing such as surprisal (Hale, 2001; Levy, 2008) or entropy reduction (Hale, 2003, 2006) share the critical assumption that incremental parsing is guided by probabilistic expectations of syntactic structures in the upcoming linguistic input, and that these parse probabilities are derived from the distribution of syntactic structures in language input (Chang, Dell, & Bock, 2006; Jurafsky, 1996; MacDonald, Pearlmutter, & Seidenberg, 1994; Mitchell, Cuetos, Corley, & Brysbaert, 1995; Pickering, Traxler, & Crocker, 2000; Tanenhaus & Trueswell, 1995). In other words, cumulative experience of syntactic structures throughout life is assumed to play a fundamental role in shaping the core properties of the parser.

Despite the strong emphasis on the role of language experience in parser development, little work has used developmental data to explore the relation between language experience and parser development. Empirical tests of these models have mostly explored the relation between parsing biases in adults and structure distributions in text corpora (Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Gennari & MacDonald, 2009; Levy, Fedorenko, Breen, & Gibson, 2012; Levy, Fedorenko, & Gibson, 2013; Levy & Keller, 2013; Linzen & Jaeger, 2015; Trueswell, Tanenhaus, & Kello, 1993, among others), or how adults' syntactic expectations adapt to the distributional manipulation within an experimental session (e.g., Fine & Jaeger, 2013; Fine, Jaeger, Farmer, & Qian, 2013; Jaeger & Snider, 2013; Myslín & Levy, 2016; Thothathiri & Snedeker, 2008; Tooley & Bock, 2014; Traxler, 2008; cf. Pozzan & Trueswell, 2015; Wonnacott, Newport, & Tanenhaus, 2008). However, these approaches only provide a snapshot of the relation between parsing mechanisms and language experience during adulthood, and largely leave open the question of whether parsing behaviors in adults were actually shaped through experience over time.

The present paper reports a novel experimental investigation that uncovers developing incrementality in children's processing of filler-gap dependencies in *wh*-questions. Filler-gap dependencies are long-distance syntactic dependencies that have received relatively little attention in the developmental psycholinguistics literature, but as the review below indicates, these structures provide an ideal testing ground for incremental structure building processes. Our visual world eye-tracking experiment probes the comprehension time course for filler-gap dependencies in children and adults, and documents the developmental trajectory for incremental processing of this structure. Specifically, we show that an adult-like active dependency formation mechanism is not available in 5-year-old children, but it starts to emerge around age 6. Moreover, our distributional analyses of filler-gap dependency structures in adult and child speech corpora show that there is little difference in the potential distributional cues for incremental processing, suggesting that the child parser may not be sensitive to distributional information on filler-gap dependencies before age 6. We argue that these developmental findings shed light on linguistic and cognitive factors that form the basis of incremental filler-gap dependency formation bias, and as such, provide theoretical implications for probabilistic parsing models in adult sentence processing.

1.1. Incrementality in the developing parser

There is a growing body of evidence for adult-like incremental sentence comprehension in children, but time course evidence for incrementality has been mostly limited to an incremental use of lexical information in anticipation of upcoming nouns, or in local structural ambiguity resolution that arises from optionality in verb argument structure. For example, a visual world eye-tracking study by Borovsky et al. (2012) examined 3- to 10-year-old children's comprehension of sentences like *The pirate hides the treasure,* and showed that children incrementally use the verb semantics to predict a plausible argument before the noun phrase (NP) was presented (for related findings, see also Gambi, Pickering, & Rabagliati, 2016; Lew-Williams & Fernald, 2007; Mani & Huettig, 2012; Nation, Marshall, & Altmann, 2003). The anticipatory fixations to a plausible object image were also modulated by the vocabulary size in children, suggesting that children's cumulative integration of lexical information is largely adult-like, but the execution of this incremental processing critically relies on how efficiently children are able to uptake the word input and access their lexicon during real-time sentence processing.

Much evidence for children's incremental resolution of syntactic ambiguities comes from research on prepositional phrase (PP) attachment ambiguities and the impact of verb information (cf. Huang, Zheng, Meng, & Snedeker, 2013). For example, Trueswell, Sekerina, Hill, and Logrip (1999) investigated processing of temporarily ambiguous sentences like *Put the frog on the napkin in the box* in adults and 5-year-old children. In this sentence, the prepositional phrase *on the napkin* could potentially be analyzed as either an argument PP that indicates the destination, or a locative modifier that specifies the location of the preceding NP referent (meaning *the frog that is on the napkin*). Eye movement data indicated that both adults and children incrementally adopted the destination interpretation, but this interpretation often perseverated in 5-year-old children (for related findings, see Anderson, Farmer, Goldstein, Schwade, & Spivey, 2011; Choi & Trueswell, 2010; Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000; Kidd, Stewart, & Serratrice, 2011; Weighall, 2008; Woodard, Pozzan, & Trueswell, 2016). Snedeker and Trueswell (2004) further demonstrated that the incremental resolution of PP attachment ambiguity is immediately constrained by verb information. For example, when adults and 5-year-old children were presented with sentences like *Tickle/Choose the frog with the feather,* both groups of participants incrementally associated the PP with the verb *tickle* and adopted an instrument interpretation, but when the verb was *choose,* they demonstrated an opposite bias to analyze the PP as a modifier of the preceding NP. The nature of the verb bias remains unclear, as it may reflect a probabilistic influence that results from distributional regularities of verb-specific structural attachment patterns in the input, or the semantic plausibility of resulting interpretations (e.g., choosing an object with an instrument is not very plausible; see Kidd et al., 2011). Under either interpretation, however, these studies indicate that children learn to incrementally use verb information to resolve PP attachment ambiguities by age 5. Given that PP attachment bias can be encoded as part of the verb lexicon (Spivey-Knowlton & Sedivy, 1995; Tanenhaus & Trueswell, 1995), these findings suggest that children are at least able to use bottom-up, lexical information for the purpose of incremental sentence comprehension.

On the other hand, children's structural ambiguity resolution may be less sensitive to cues that require use of top-down information that goes beyond lexical information (for discussions, see Snedeker, 2013). For example, one notable difference between adults and children is that 5-year-old children fail to incrementally use referential information for PP attachment ambiguity resolution. When the scene contains two referents for the object NP (e.g., a frog on a napkin vs. a frog on a towel), adults can immediately use this referential information to analyze the following PP as the NP modifier, as the definite description for the object NP pragmatically requires a unique referent. Children, on the other hand, show a very strong bias towards the destination interpretation in both 1-referent and 2-referent contexts (Choi & Trueswell, 2010; Trueswell et al., 1999). Some evidence suggests that their ability to use visual contexts may start to emerge around age 5, but it is not robust enough to reliably guide their comprehension through an entire experiment (Snedeker & Trueswell, 2004).

This type of child-adult contrast in processing behaviors presents an opportunity for investigating how parsing biases develop in children. One plausible explanation is that children have not experienced a sufficient number of communication situations in which visual contexts help to resolve syntactic ambiguities, and this lack of experience has prevented children from learning the critical dependency between the number of referents and likelihood of NP modifier analysis. It is challenging to empirically assess this claim, however, as it is not feasible to estimate how often children encounter utterance situations like this. In fact, even if this frequency information is available, it is difficult to infer how children mentally represented the scene information in such contexts.

In sum, developmental work has shown that children can access lexical information to incrementally assign interpretations to the input, but may be less sensitive to non-lexical, top-down cues, such as referential information. As we illustrate below, incremental filler-gap dependency resolution requires a use of top-down syntactic knowledge of structural candidates, and for this reason, its development may plausibly be delayed in children. On the other hand, unlike the availability of referential information, it is feasible to estimate its distributional regularities in language experience based on available corpora (see Section 2). This corpus information can be used to assess the role of language experience in the development of incrementality in filler-gap dependency processing.

1.2. Filler-gap dependency processing in adults

Filler-gap dependencies are seen in constructions like *wh*-questions (e.g., *What did Emily eat the cake with __ ?*) or relative clauses (e.g., *The book that the author wrote __ was on sale*) where a constituent (such as *what* or *the book*) is fronted to a non-canonical structural position. During real-time comprehension of filler-gap dependencies, the parser must hold the fronted constituent (called *filler*) in memory, and retrieve the filler after identifying the correct thematic position (called *gap,* which is used here as a descriptive term for the sentence position at which the filler is interpreted, with no commitment to the nature of the representation). However, this gap search process is not a trivial task. First, there is no clear bottom-up (e.g. auditory) signal for the silent gap position (Straub, Wilson, McCollum, & Badecker, 2001). Second, the filler-gap dependency can span over multiple phrases (e.g., *What did the girl with red hair eat the cake with __?*) or even clauses (e.g., *What did the girl believe that the boy said that John ate the cake with __ ?*). Due to this flexibility in the structural position of a gap, filler-gap dependencies create a temporary ambiguity in the gap position, and the parser must maintain the filler in memory until this ambiguity is resolved with late-arriving information (Fodor, 1978).

Psycholinguistic research with adults has provided much evidence for active dependency formation bias, i.e., incremental completion of the dependency at the earliest possible position (Aoshima, Phillips, & Weinberg, 2004; Chacón et al., 2016; Crain & Fodor, 1985; Frazier, 1987; Frazier & Clifton, 1989; Frazier & Flores D'Arcais, 1989; Garnsey, Tanenhaus, & Chapman, 1989; Johnson, Fiorentino, & Gabriele, 2016; McElree & Griffith, 1998; Omaki et al., 2015; Omaki & Schulz, 2011; Parker, 2017; Pickering & Traxler, 2003; Staub, 2010; Stowe, 1986; Traxler & Pickering, 1996; Wagers, Borja, & Chung, 2015; Wagers & Pendleton, 2016). For example, an eye-tracking during reading study by Traxler and Pickering (1996) manipulated the semantic fit of the filler and the verb (e.g., ... *the city/book that the author wrote about…*), and found reading time increase at the verb region when the filler was an implausible object of the verb (*city-wrote*) compared to when it was a plausible object of the verb (*book-wrote*). This finding suggests that the parser actively associated the filler with the verb, even though the correct gap position appeared later in the sentence (e.g., after a preposition *about*; for related findings on plausibility mismatch effects, see Chow, Smith, Lau, & Phillips, 2015; Garnsey et al., 1989; Omaki & Schulz, 2011; Staub, 2007; Wagers & Phillips, 2014).

Converging evidence comes from Sussman and Sedivy (2003), who used a visual world eye-tracking experiment to investigate the time course of filler-gap dependency processing in adults. Participants were presented a story (e.g., Jody was eating breakfast, saw a spider, and squashed it with her shoe) with a visual display of four related pictures (e.g., Jody, a spider, a shoe, and breakfast). After the story, participants were presented either a *wh-* or *yes-no* question about the story (e.g., *What did Jody squash the spider with?* vs. *Did Jody squash the spider with her shoe?*). Eye movement data during the target question showed that during the verb region, there was significantly greater proportion of fixations on the patient image (*spider*) in the *wh-*question condition than in the *yes-no* question condition. This finding indicated that adult listeners actively associated the *wh*-phrase with the verb (i.e., *What did Jody squash* __ *?*), and directed their gaze towards the patient image which (temporarily) constituted the answer to the *wh*-question (for related findings, see also Omaki, 2010).

There is much consensus in this literature that the active dependency formation mechanism in adults differs in one critical way from other syntactic ambiguity resolution (e.g., PP attachment ambiguity), which is that filler-gap dependency resolution pays little regard to bottom-up, lexical cues that are encoded in the potential gap host, such as verb subcategorization information. For example, Pickering and Traxler (2003) as well as Staub (2007) showed that active dependency formation is observed even for optionally transitive verbs that are more frequently used as an intransitive verb. Moreover, Omaki et al. (2015) provided evidence that the parser attempts to associate the filler even with strict intransitive verbs. These findings highlight the fact that the adult-like active dependency formation mechanism relies on top-down syntatic knowledge about which structural position is the earliest or most likely gap position, rather than bottom-up lexical information about whether e.g., a verb can plausibly host the filler. Given the observations discussed above about children's immature use of top-down information, it is plausible that children may not actively form filler-gap dependencies in an adult-like fashion.

Two major classes of explanations for the active dependency formation bias also suggest that factors such as cognitive constraints on working memory and statistical learning may play a role in the development of the bias. The first class of explanation suggests that active formation of dependencies reflects constraints on working memory that favor minimization of syntactic dependency itself. We will refer to this position as memory constraint account of active dependency formation. The longer distance between the filler and the gap may incur a large memory cost due to active maintenance and retrieval of decaying information (Chen, Gibson, & Wolf, 2005; Fiebach, Schlesewsky, & Friederici, 2002; Gibson, 1998; Grodner & Gibson, 2005; Grodner, Gibson, & Tunstall, 2002), and it may also increase the chances of retrieval interference due to the increase of intervening words (Lewis & Vasishth, 2005; van Dyke, 2007; van Dyke & McElree, 2006). Under this view, populations with limited memory resources like children may have a stronger bias than adults to complete filler-gap dependencies early in the structure in order to reduce memory costs or interference.

The second major class of explanation attributes the active dependency formation to a general comprehension bias to pursue and select more probable syntactic structures (Levy, 2008; Wagers & Pendleton, 2016). Let us call this probabilistic prediction account of active dependency formation. As will be shown below in Section 2, filler-gap dependencies with direct object gaps (e.g., *What did Emily eat __ with the cake?*) are more frequent and probable in English than those with prepositional object gaps (e.g., *What did Emily eat the cake with __ ?*). This account predicts that if this distributional information is accessible to adults and children during real-time comprehension, a direct object gap should be expected.

It is important to note that these accounts are not mutually exclusive. Both constraints on working memory and distributional information could contribute to the active dependency formation bias. Critically, under both explanations, the identification of the closest or most probable gap position would require top-down, syntactic knowledge of what gap positions are available for a given filler, rather than bottom-up, lexical information from the verb. It is unknown how working memory constraints and language experience condition the availability of top-down linguistic information in children's real-time processing. Because children differ from adults in their memory capacity and potentially differ in their experience with language, developmental investigations on filler-gap dependency processing could shed light on how memory constraints, language experience, or a combination of these two factors influence this incremental parsing bias. Section 2 will provide corpus analyses on the distribution of *wh*-questions and gap positions in child-directed speech and adult conversations in order to compare the nature of distributional cues in the input to adults and children.

While these two accounts provide insights on the origin of filler-gap dependency processing biases, it is also possible that neither the memory constraint account nor the probabilistic prediciton account fully captures the entire mechanistic details of what constitutes the active dependency formation bias. Given that both accounts were primarily based on the sentence processing behaviors of competent adult readers, these accounts may have missed opportunities to shed light on important factors that condition the execution of active dependency formation. For example, it is possible that children have a weaker ability to encode lexical and structural information in working memory, which may cause a failure to trigger the bias to shorten the filler-gap dependency. As for the probabilistic prediction account, it is still unknown whether distributional information can be accessed and used in the same way by different populations. For example, suppose that there are two competing structures that have probabilities of 0.6 versus 0.4, and that the same distributional information is available for adults and children in their respective input. This distributional cue may give rise to the same processing bias in adults and children, but it is also possible that children apply a different (e.g., higher) threshold for making syntactic commitments, such that children are willing to make commitments only when a sharper probability contrast is present (e.g., 0.9 versus 0.1). In this sense, if children demonstrate behaviors that are not predicted by either the memory constraint or probabilistic prediction accounts in their current formulations, such developmental data could provide novel insights on the core foundation of the filler-gap dependency mechanism that was not evident in the existing adult sentence processing research.

1.3. Past research on filler-gap dependency processing in children

There are a few studies on children's processing bias in comprehension of filler-gap dependencies, but the existing studies do not provide strong time course evidence for active formation of filler-gap dependency. For example, Love (2007) used a cross-modal picture priming study with filler-gap dependency stimuli like *the zebra that the hippo had kissed __ on the nose ran away*, and found that 4- to 6-year-olds made an edibility judgment (able to be eaten vs. not able to be eaten) more quickly when the presented picture was of the fronted direct object NP, e.g., *zebra*, than when it was of an unrelated animal, e.g., *camel*, or of the relative clause subject NP, e.g., *hippo* (for a related study, see Roberts, Marinis, Felser, & Clahsen, 2007). Love (2007) interpreted this data to indicate that children incrementally reactivated the filler noun phrase at the verb in anticipation of the direct object gap, but it is possible that the facilitation effect may have resulted from interpreting the picture as the continuation of the sentence fragment (e.g., *the hippo had kissed…*). Integrating *zebra* as a new object produces a syntactically and semantically congruent sentence (*the hippo had kissed the zebra*), while integrating *hippo* as an object is less natural (*the hippo had kissed the hippo*). An unrelated animal (e.g. *camel*) would probably take long simply because it has not been introduced in the context.

Omaki et al. (2014) examined cross-linguistic variations in offline interpretation preference to argue that 5-year-old children have active dependency formation bias. Using ambiguous bi-clausal *wh*-questions in English (e.g., *Where did Lizzie tell someone that she was gonna catch butterflies?*) and also in a verb-final language like Japanese, Omaki and colleagues showed that adults as well as children have an interpretation bias to associate the *wh*-phrase with the first verb in the sentence: English-speaking adults and children preferred to answer the location for the main clause event (e.g., telling), whereas Japanese adults and children answered the location for the embedded clause event (e.g., catching butterflies), which corresponds to the first predicate in Japanese due to its verb finality. Furthermore, Japanese children's first verb association bias persisted even when the sentence explicitly mentioned the location of the embedded clause event (e.g., answering ‘park’ when the sentence actually mentioned *catching butterflies in the park*). This suggests that the association of the *wh*-phrase to the first verb may have happened incrementally, and this initial analysis is difficult for children to revise (see Trueswell et al., 1999; for related findings in French *wh*-questions, see Lassotta, Omaki, & Franck, 2015).

However, there are two caveats to this finding. First, no time course evidence was presented to support the proposed incremental process. Offline data used by Omaki et al. (2014) reflects the ultimate interpretation of the sentences, and is only suggestive of the real-time interpretative processes that occurred when the critical portion of the sentence (i.e., the first verb) was encountered. Thus, time course data would provide stronger evidence for children’s active dependency completion. Second, this particular study compares potential gap positions in two different clauses (main vs. embedded), while primary evidence for active dependency formation in adults comes from cases where two (or more) gap positions in the comparison set are within the same clause (cf. Aoshima et al., 2004; Frazier & Clifton, 1989; Gibson & Warren, 2004). As such, in order to provide time course evidence for adult-like incremental completion of filler-gap dependencies, it is ideal to use a fine time course measure that taps incremental filler-gap dependencies within the same clause.

1.4. Summary

In sum, filler-gap dependency processing provides a useful opportunity to investigate children’s use of top-down syntactic knowledge during real-time sentence comprehension. In addition, this investigation could shed light on the role of memory constraints and language experience in shaping the adult-like active dependency formation bias.

The present study uses a modified version of the visual world eye-tracking paradigm by Sussman and Sedivy (2003) to provide time course information on filler-gap dependency processing. In addition, unlike Omaki et al. (2014), we will use questions like *Can you tell me what Emily was eating the cake with __ ?* in which the two potential, competing gap positions are either the direct object (DO) gap position (…*what Emily was eating __ ?*), or the prepositional object (PO) gap position. This is more comparable to the structural environment that has been widely used in adult sentence processing research. If participants demonstrate the active dependency formation bias, they should show eye movement evidence for temporarily interpreting *what* as the direct object of the verb (see Section 3).

Before we proceed to the experimental investigation of filler-gap dependency processing, we will assess the nature of language experience for both adults and children (Section 2). The experimental predictions of the probabilistic prediction account of active dependency formation critically depends on the details of language experience, specifically, the distributional information of direct object gap and prepositional object gap in the input to adults and children.

2. Distributional analyses of filler-gap dependencies in the language input

In order to make precise the prediction of the probabilistic prediction account, we compared the distribution of *wh*-questions in adult and child corpora, and examined if there are differences in the distributional regularities of gap positions in their language experience.

2.1. Corpus information and coding scheme

The distributional analysis of filler-gap dependencies for adults was based on two naturalistic corpora of adult spoken language: the CallHome corpus (Kingsbury, Strassel, McLemore, & McIntyre, 1997) and a selection from the Switchboard corpus (Marcus, Santorini, Marcinkiewicz, & Taylor, 1999). These corpora were chosen because they consist of naturalistic, conversational speech between two adult participants. Spoken language corpora were used because the adult distribution will be compared to that of children, who only have limited experience with written language by the age of 5.

The distributional analysis of *wh*-questions in the child input was drawn from the CHILDES Treebank (Pearl & Sprouse, 2013), which provides a structure annotation for a subset of the corpora in the CHILDES database (MacWhinney, 2000).[1] Specifically, these corpora included Abe (Kuczaj, 1977), Naomi (Sachs, 1983), Nina (Suppes, 1974), and Adam and Sarah from the Brown corpus (1973) and are available on CHILDES (MacWhinney, 2000). This resulted in 143,353 lines of child-directed speech and 158,194 lines of child speech. Table 1 provides additional details about both the adult and child corpora including the number of lines of examined speech, children's age range, and the number of files / sessions.

Table 1
Details of corpora used in the current study.

Adult corpora

Corpus	Number of Files Examined	Number of Lines of Speech
CallHome	120	28,967
Switchboard	199	44,696

CHILDES corpora

Corpus	Child	Age Range	Number of Sessions	Number of lines of Child-Directed Speech
Brown (1973)	Adam	2;3 – 4;10	55	26,688
	Sarah	2;3 – 5;1	139	46,192
Kuczaj (1977)	Abe	2;4 – 5;0	210	22,156
Sachs (1983)	Naomi	1;1 – 5;1	93	12,251

[1] While the adult corpus data simultaneously accounts for comprehension and production frequency, the same is not true for child-directed speech, which only reflects frequency of structures in children's comprehension. See Atkinson (2016) for demonstration that the gap distributional pattern in comprehension is replicated in children's production.

Suppes (1974)	Nina	1;11 – 3;11	56	35,965

For the adult analysis, the search was limited to questions and relative clauses in syntactically parsed files. We used the Tregex utility (Levy & Andrew, 2006) to search the parsed trees for *wh*-phrases that indicate argument extractions: *who*, *what*, *which*, and *whose*. For the analysis of child-directed speech, sentences with an argument *wh*-word were extracted using a Python script written by one of the authors (e.g., DO gap: ‘I went for a walk tonight and what did I see?’, PO gap: ‘what are you crawling on?’, both from Adam’s corpus). These extracted sentences were then coded for embedding (main versus embedded question) and gap position (subject gap, DO gap, or PO gap). No differences were found between main and embedded clauses, so they are collapsed in the data below. Echo questions were also excluded.

2.2. Results

The analysis focuses on the filler-gap dependencies that are most relevant to those in the visual world experiment: *what* questions with post-verbal gaps. Of 73,663 lines of examined adult speech, only 546 contained a post-verbal *what* question (0.74% of the analyzed corpora). Of the 143,252 lines of child-directed speech in the examined CHILDES corpora, 3,737 of them contained *what* questions with a post-verbal gaps. This accounts for 2.6% of child-directed utterances. *What* questions with a subject gap were removed from the total, because they are not relevant for the current study. Table 2 presents the overall results of this analysis.

Table 2
Distribution of what *questions*

Adult corpora			
Corpus	DO Gap	PO Gap	Total
CallHome	369	55	424
Switchboard	105	17	122
Overall	474 (86.8%)	72 (13.2%)	546

CHILDES corpora			
Child	DO Gap	PO Gap	Total
Abe	304	28	332
Adam	618	118	736
Naomi	292	39	331
Nina	1,841	199	2,040
Sarah	274	24	298
Overall	3,329 (89.1%)	408 (10.9%)	3,737

Of the 546 *what* questions in the adult corpora, 474 contained a DO gap; this accounts for 86.8% of the questions. Clearly, the distribution of *what* questions that adults produce when speaking with other adults skews toward DO gaps; they are approximately 6.5 times more frequent than PO gaps. Thus, the distribution of gaps in adult's linguistic experience favors a DO gap interpretation and supports active dependency formation. Of the child-directed *what* questions, 89.1% of them were DO gap questions.[2] This distributional information is clearly skewed toward a preference for DO gaps, and adults and children are exposed to similar distributions of post-verbal gaps in their input: approximately 85% DO gaps versus about 15% PO gaps. These findings suggest that there is little difference between adults and children in distributional regularities of filler-gap dependency structures. As such, the probabilistic prediction account of active dependency formation predicts that adults and children should be equally biased towards incrementally adopting the DO gap analysis.

3. Experiment: Filler-gap dependency processing in the visual world

The present visual world eye-tracking study investigates the developmental trajectory of filler-gap dependency processing in adults and children. The child group consisted of 5- to 6-

[2] These counts include *wh*-questions that did not have an overt DO (e.g., What will he write on?) or a prepositional phrase (e.g., What did you say?), which differ from the target *wh*-questions in our experiment which always involved a DO and a prepositional phrase. These *what* questions are relatively uncommon: about 30% of questions with PO gaps included an overt object noun phrase, while 36% of DO gap questions included a prepositional phrase. Nonetheless, there is a similar distributional bias for DO gaps (90.6% vs. 9.4%).

year-old children, and this age range was selected based on the range used in the previous research on filler-gap dependency processing in children. It is also well documented that by this age children's syntactic knowledge of filler-gap dependencies is adult-like (e.g., de Villiers & Roeper, 1995; Gagliardi, Mease, & Lidz, 2016; Hamburger & Crain, 1982), so behavioral differences between adults and children (if any) are likely to reflect differences in their parsing mechanisms.

The experiment design uses a Question-after-Story paradigm (de Villiers & Roeper, 1995), and was modeled after Sussman and Sedivy (2003) and Omaki (2010). The story phase presents two critical events (e.g., eating a cake with a fork, washing the dishes with a sponge), and after the story, participants are presented with either a temporarily ambiguous *wh*-question like (1a) or a *yes-no* question with no filler-gap dependency like (1b).

(1) a. Can you tell me **what** Emily was eating the cake with __?

b. Can you tell me **if** Emily was eating the cake with the fork?

In (1a), if participants actively complete the filler-gap dependency and anticipate a gap at the earliest structural position, the fronted *wh*-phrase *what* should be temporarily interpreted as the direct object (DO) of the verb *eating* (DO gap analysis). Under this active DO gap analysis of the *wh*-question, participants should direct their gaze towards the patient image (the cake) at the verb region, because this image constitutes the answer to the question. The *yes-no* question counterpart (1b) was used to establish baseline fixations on the patient image during the verb. Here, when participants hear up to the critical verb region, the only syntactic difference between the two conditions is the presence of the fronted *wh*-phrase. As such, a reliable difference in patient fixations between (1a) and (1b) during the verb region serves as evidence for active formation of the filler-gap dependency.

Moreover, the object NP region (*the cake*) of the target sentences provides additional opportunities to tap active formation of the filler-gap dependency. First, it is possible that children are slower in lexical, discourse, or visual processing, so the anticipatory fixations towards the patient image that were planned during the verb region may not emerge until the object NP region. Second, this region may allow participants to demonstrate active dependency formation for the ultimate gap position. At the point of processing the object NP, this input will disconfirm the DO gap analysis, and given the story design of the study (see below), the only other plausible gap position is the prepositional object (PO) position after an upcoming preposition *with*. This (second) incremental PO gap analysis should yield anticipatory fixations on the instrument image before the preposition is presented in the auditory input.

The present study also explored whether children's filler-gap dependency processing may be mediated by individual differences in age and vocabulary size. It is plausible that age correlates with language experience or cognitive maturation that may be critical for the development of active dependency formation. Moreover, studies examining children's incremental use of verb information suggest that there is a reliable positive correlation between vocabulary size and processing speed, and that this relation holds from age 2 to 11 (Borovsky et al., 2012; Mani & Huettig, 2012; Nation et al., 2003). As mentioned above, processing of filler-gap dependencies in adults is known to be blind to statistical information associated with the verb, but variability in children's lexical processing speed could affect how quickly they can integrate words into structures. For this reason, we included a measure of children's receptive vocabulary size to explore its relation to filler-gap dependency processing.

Let us recap the explanations of active dependency formation in the adult literature, as well as their predictions for children's filler-gap dependency processing. The memory constraint account states that the active dependency formation bias results from a preference for a linearly shorter dependency, which reduces the memory load or chances of interference. The probabilistic prediction account states that adults use distributional cues to predict the upcoming structure, and more probable gap positions are preferentially pursued. Both accounts predict that children should also demonstrate active dependency formation, and hence prefer the DO gap analysis: this analysis provides a shorter dependency than the PO gap analysis, and it is more probable than the PO gap based on the corpus data in Section 2. However, if children do not incrementally adopt the DO gap analysis, it would imply that these accounts may need to be modified.

3.1. Method

3.1.1. Participants

Forty English-speaking children between the ages of 5;0 and 7;0 (mean age = 5;10, 21 females) participated in the study. Of these, 20 were in the 5-year-old group (mean age = 5;5, age range = 5;0-5;11, 9 females) and 20 were in the 6-year-old group (mean age = 6;5, age range = 6;0-6;11, 12 females). These children were recruited from the communities surrounding Johns Hopkins University and the greater Baltimore area. Three additional children participated in this study, but their data was excluded from analyses due to technical difficulty (n = 1) or lack of attention (n = 2).

In addition, 24 adult native speakers were recruited from the Johns Hopkins University community and were paid $5 for participating in this experiment. Four additional adult participants were tested but their data were excluded from analyses due to technical problems (n = 2) or lack of attention (n = 2).

3.1.2. Materials

Story and display design. A total of 20 stories with clipart animation were constructed. Ten of these were always used with target questions, and the remaining 10 were used with filler questions. These stories consisted of two events with different verbs, such that participants could not predict the content of the question until the verb was presented to them. This in turn makes the verb region the earliest moment in the *wh*-question condition where participants could potentially infer the answer to the question. The stories were followed by either a *wh*- or *yes-no* question about one of the events (see below for details of the question design). Each story followed the same basic structure: the character on the display introduced him or herself, the two events were mentioned, the character chose one and completed it, and the character completed the remaining event. A sample target story with a target question is provided in (2).

(2) Hi, my name is Emily. Today I'd like to eat some cake, but I also need to wash the dishes. Hmm, what should I do first? I think I'm gonna eat the cake, and for that I need a fork. Mmm! That cake was yummy. Now it's time to wash the dishes. I'm gonna need to use a sponge. Oh, those dishes are so clean. I did a great job today.

Question: Can you tell me what Emily was eating the cake with?

As stories involved a subject and two distinct events with associated patients and instruments, each display consisted of five pictures (see Figure 1): the agent (e.g., Emily), the patient and instrument from the first event (e.g., cake, fork), as well as the patient and instrument from the second event (e.g., dishes, sponge). The position of each type of picture was balanced across stories such that they appeared in all five locations. In order to make the display more engaging, animations were added to accompany each event. For example, when the event is about eating cake with a fork, the fork picture moved to the cake picture. Most of these animations left a "trace" of the event, so there was a visual representation of the events in the story. In the case of eating cake, a slice of cake was replaced by crumbs and the fork became dirty. These event traces were intended to facilitate relevant eye movements by increasing the available visual information about the completed events. Figure 1 shows the beginning (i.e., before either event) and end of the display associated with the story in (2). The list of target story scripts and questions can be found on the first author's website (http://www.emilyeatkinson.net).

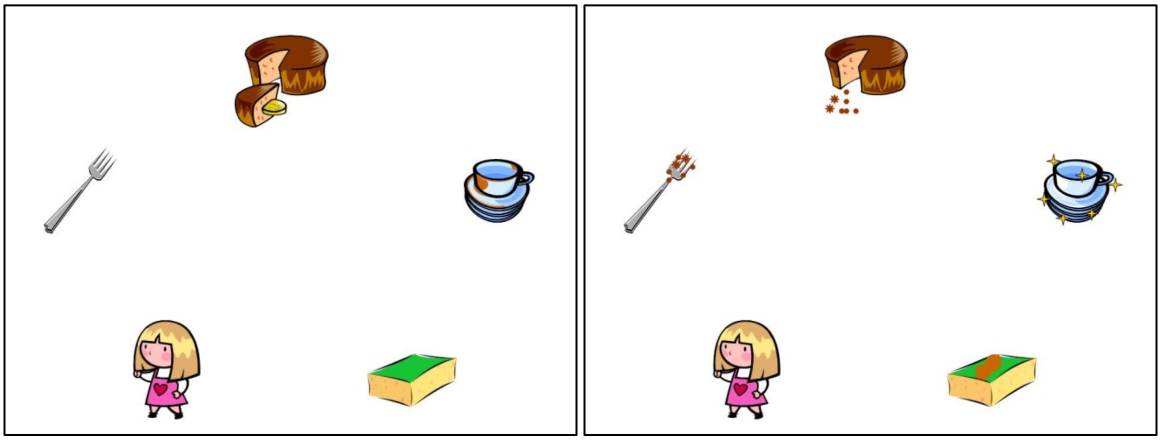

Figure 1. A sample story display. The initial phase is on the left, and the final phase is on the right. For a question like "Can you tell me what Emily was eating the cake with __ ?", the following image labels were used (from the top, clockwise): target patient (cake), distractor patient (dishes), distractor instrument (sponge), agent (Emily), target instrument (fork).

Questions. This experiment manipulated the question type and created *wh*-questions and *yes-no* questions (see (1) above). In order to maximize the structural similarity across conditions, all questions were embedded under a carrier phrase *Can you tell me*. This ensures that the only difference between the questions is the presence or absence of a filler-gap dependency, unlike main clause questions that require a different number of words prior to the critical verb region. All questions used the progressive form of the verb (*was VERB-ing*) to increase the verb duration, which helped to maximize the chances of observing potential changes in eye movements (if any) during the verb region. The average duration of the target verbs was 664ms (minimum = 475ms, maximum = 947ms).

For each of the 10 target stories, a *wh*-question and a *yes-no* question were constructed for each of the two events (e.g., cake eating event, dish washing event) in order to control for unexpected variance in visual salience or discourse prominence of the respective image. This resulted in 40 total target questions (4 per story, 2 *wh-* and 2 *yes-no* questions). In the *wh*-questions, the *wh*-phrase *what* was extracted from an instrument prepositional phrase headed by the preposition *with*. The *yes-no* questions also included an instrument prepositional phrase (e.g., *with the fork*). The answer to target *yes-no* questions was always "yes."

A single question was constructed for each filler story. The 5 filler *wh*-questions asked about the direct object rather than the instrument (e.g., *Can you tell me what Esmeralda was squashing __ with the magic wand?*). The 5 filler *yes-no* questions had the same structure as the target *yes-no* questions, but the correct answer to these questions was "no." In these questions, either the direct object or prepositional object from the non-questioned event was substituted for the correct noun phrase (e.g., *Can you tell me if Ethan was painting the TV with the brush?,* after a story in which Ethan actually painted the door).

Four lists were generated by counterbalancing the target questions such that each participant only heard one version per story. Each list consisted of 5 *wh*-targets and 5 *yes-no* targets. Half of these questions asked about the first event in the story, while the other half asked about the second event. The 10 targets were combined with the 10 fillers for a total of 20 story-question combinations.

Audio recording. The narratives were recorded by a female native speaker of American English. An additional female native speaker of American English recorded the questions. The narratives and questions were read with child-directed prosody and were recorded with a sampling rate of 44.1 kHz. The sound files and animations were incorporated into a single movie file for presentation.

3.1.3. Procedure

The 20 trials were grouped into four blocks of five trials. All trials began with a narrative and associated movie display. Following the story, a fixation cross appeared in the center of the screen and remained until the participant fixated on it for 1000ms. This fixation triggered the reappearance of the last display from the story. The display was accompanied by a question about one of the events in the story, and participants were prompted to answer this question aloud.

Participants were seated with their eyes approximately 24 inches in front of an EyeLink 1000 remote eye-tracker (SR Research, Toronto, Ontario, Canada), which is integrated in an LCD arm mount with a 17-inch computer monitor. The eye-tracker had a sampling rate of 500 Hz and a spatial resolution of less than one degree of visual angle. The audio was presented through free-standing speakers on both sides of the monitor. Participants were instructed to look at the pictures during the story and the question. Head movements were unrestricted, but participants were asked to minimize their movements and to look at the pictures by only moving their eyes. A 5-point calibration was performed before beginning the experiment. The entire procedure including consent, instructions, calibration, the experiment, and debriefing took approximately 25 to 30 minutes.

The procedure for the child participants was slightly modified to make the task more engaging and child-friendly. Before beginning any trials, children were given a practice story accompanied by practice questions, which did not have the same structure as the target questions. Calibration accuracy was checked at the beginning of each block to ensure a reliable recording of eye movement data. In addition, the fixation cross was replaced with a picture of a cartoon character of a comparable size. In order to ascertain that children's gaze was on the display at the onset of questions after stories, the questions were presented only after children fixated on the character for 1000ms. Finally, children received positive feedback after every trial, and they received a sticker as a reward after each block. This encouraged children to pay attention to the stories and stay engaged in the task.

Comprehension vocabulary measure. Child participants' vocabularies were assessed using the Peabody Picture Vocabulary Test, Fourth Edition (PPVT™-4, Dunn & Dunn, 2007). This test was administered after the children completed the visual world eye tracking experiment and took between 15 and 20 minutes to complete.

3.1.4. Data analysis

We screened the eye movement data and removed outlier trials from further analyses. First, because the main behavioral task in this eye-tracking experiment (i.e., answering questions out loud) does not require participants to fixate on the relevant image, some participants fixated disproportionately on the blank areas of the screen or on a single image throughout the question answering phase. Including such trials could skew the fixation proportion data and mask relevant effects, so we excluded trials in which the duration of fixations on any combination of the five pictures was below 35% of the question duration.[3] Additionally, trials were removed from analyses if the proportion of fixations to a single picture was 3 standard deviations above or below the mean proportion of fixations to a single picture during the question. Trials with incorrect answers to the question were also removed. These criteria affected 25 out of 240 target trials (10.4%) in adults, and 50 out of 400 target trials (12.5%) for the children. Two adults and

[3] To determine this criterion, the distribution of the duration of picture fixations per trial was examined for each age group. The point at which the distribution approached an asymptote was chosen as the cutoff percentage for each age group.

one child (a 5-year-old) were excluded from further analyses because these exclusion criteria removed more than half of their target trials.

For the purposes of statistical analyses, the data from the four lists was collapsed as there were no reliable differences between those lists. The overall time course data was arranged into 50ms (25 frame) windows (see below). Fixations were shifted by 200ms to account for the amount of time it takes to plan and execute a saccade (Altmann & Kamide, 2004; Matin, Shao, & Boff, 1993). In the figures to follow, the 0 time point on the x-axis represents the onset of the critical region of interest, but the actual region of analysis is indicated by the dotted lines that account for the 200ms of saccade latency.

The critical regions of interest were the verb region and the following object NP region. First, as discussed above, the verb region provides the earliest opportunity in the sentence to actively complete the filler-gap dependency and analyze the *wh*-phrase as the DO (Omaki 2010; Sussman & Sedivy, 2003), and for this reason the patient image fixation is the critical dependent variable. Second, we also investigated a potential spill-over effect of this active dependency formation by analyzing the patient image fixation in the subsequent object NP region. Third, we also examined the instrument fixation pattern in the object NP region, where the *wh*-phrase can be predictively analyzed as the PO. Because the average verb length was 664ms, 14 bins (700ms) were included in the analysis of the verb region. The average length of the object NP region was 889ms (minimum = 513ms, maximum = 1213ms), so 18 bins (900ms) were included in the analysis of this region.

Following suggestions by Barr (2008), we used 50ms analysis windows and the empirical logit transformation of fixation data for two reasons. First, because successive eye movements are affected by previous fixations, fixation data inherently violates the independence assumption of standard statistical analyses. One approach to filtering out these fixation dependencies is to aggregate the data into temporal bins (Barr, 2008). This way, each time bin represents a group of eye movement samples rather than an individual frame recording, and the fixation data across time bins are more independent than fixation data across each frame. Then, the categorical fixation data is transformed using the logit function to generate the log-odds of fixating on a particular image. The empirical logit is a quasi-logit transformation that handles cases where the probability of fixating on a particular image may be close to either 0 or 1, which would result in non-finite values under the standard logit function.

The empirical logit data was fit to a linear mixed effects model with population (adults versus children), question type, time and quadratic time as fixed effects. The inclusion of time as a fixed effect allows us to examine the effect of our manipulations on two different features of the eye tracking record: the intercept and the slope (see Barr, 2008). The intercept represents the likelihood of fixating on the image of interest at the onset of the word region. The slope, on the other hand, describes how fixations on the image of interest change over the course of the region, and thus sheds light on how rapidly fixations increase or decrease. Quadratic time (i.e., time squared) was also included as a factor for the models examining patient fixations in the verb region because of the expected pattern of fixations: participants fixate on the patient image associated with the linguistic stimuli, but tend to decrease their fixations toward the end of the region in anticipation of the next word region.

The model was fit with the maximal random effects structure that converged (Barr, Levy, Scheepers, & Tily, 2013), which minimally included random slopes for time and random intercepts for participants and items. This and all other models were run in the R environment (R Core Development Team, 2015) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). *P*-value estimates for the fixed and random effects were calculated using the Sattherwaite approximation in the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2015). Both categorical fixed effects, population and question type, were sum coded. For population, adults were coded positively, while children were coded negatively. For question type, *wh*-questions were coded positively and *yes-no* questions were coded negatively. Additionally, when there was a significant interaction of population type and question type, separate planned pairwise comparisons for adults and children evaluated the effect of question type within each level of population. The data from adults and children were isolated and individually fit to another linear mixed model with question type and time as fixed effects.

Finally, in order to explore the relation between the development of active dependency formation and vocabulary size or age, two separate linear fixed effect models were fit for the children. The first model used children's age in months, and the second one included children's raw vocabulary scores on the PPVT-4™.[4] Both of these continuous variables were centered for use as additional fixed effects. We used two separate models to explore the contribution of age and vocabulary size factors, because complex models with both of these factors as fixed effects failed to converge, most likely because of the (expected) significant correlation between these two variables ($r = 0.38$, $p = 0.01$).

3.2. Results

3.2.1. Vocabulary test scores

Children had an average raw score (i.e., total number correct) of 126.9 words ($SE = 2.5$ words), an average standardized score of 123.2 words ($SE = 1.8$ words), and averaged in the 89th

[4] We conducted the analyses using both raw and standardized scores, but the overall patterns were the same. This paper reports analyses based on raw vocabulary scores, as we think that the raw vocabulary score is a more veridical measure of the language processing proficiency in children.

percentile (*SE* = 2.4 percentiles). Overall, the children were fairly high achieving; only 4 children scored at or beneath the 75th percentile for their age in years and months.

3.2.2. Comprehension accuracy

The adults answered 99% of the questions accurately. The children had a mean accuracy of 97.6%, and the lowest accuracy was 85%. The participants' good offline performance suggests that they were attentive during the stories.

3.2.3. Eye movement data

The fixation data in the results section focuses on the results that are critical for assessing the presence of an active dependency formation bias. Patient or instrument fixations in this section refer to fixations on the patient or instrument images that are involved in the target event that is relevant to the question stimuli. Figure 2 presents fixation data from adults and children for all regions of interest.

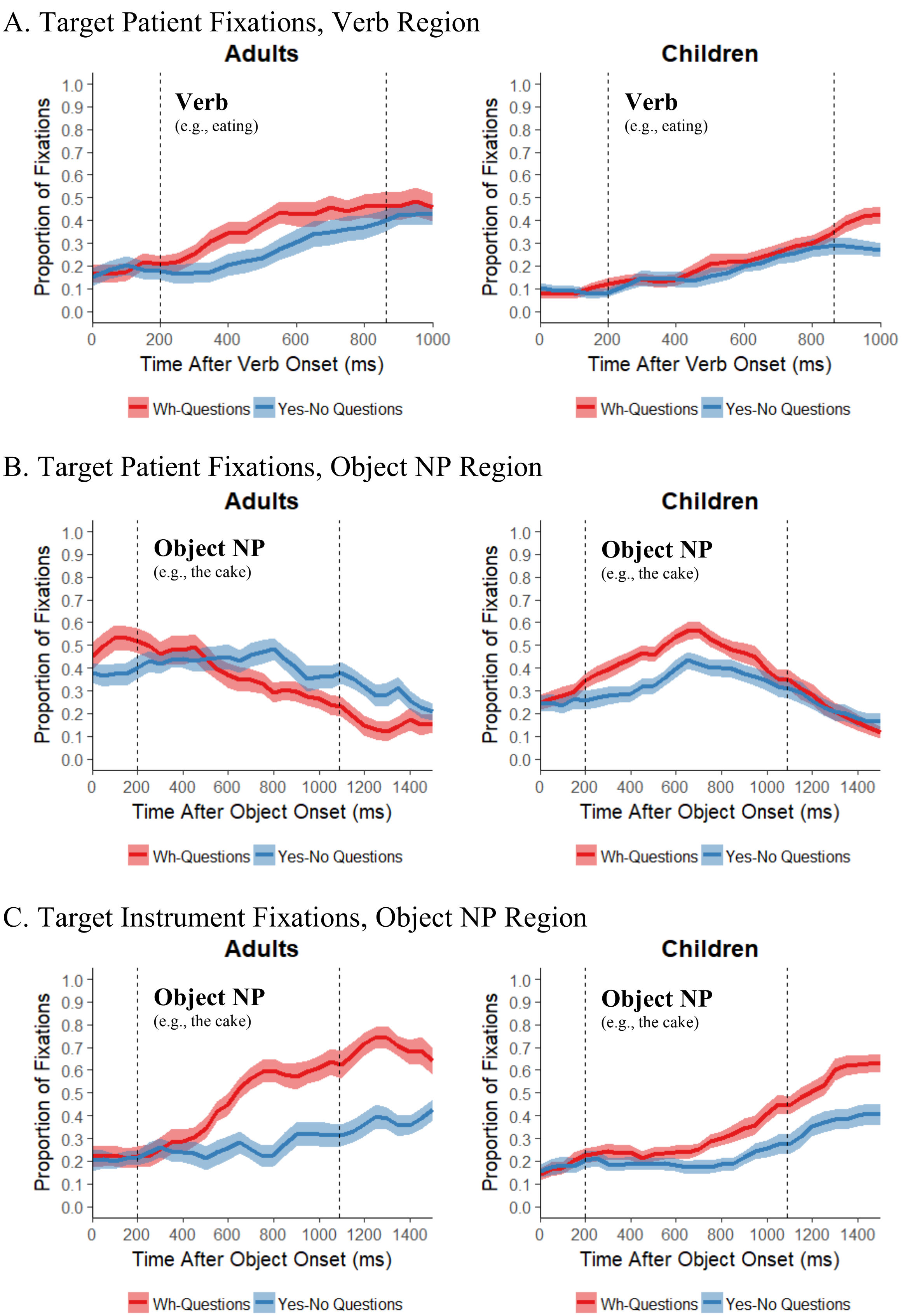
A. Target Patient Fixations, Verb Region
Adults
Children
Verb
(e.g., eating)
Proportion of Fixations
Time After Verb Onset (ms)
Wh-Questions
Yes-No Questions
B. Target Patient Fixations, Object NP Region
Adults
Children
Object NP
(e.g., the cake)
Proportion of Fixations
Time After Object Onset (ms)
Wh-Questions
Yes-No Questions
C. Target Instrument Fixations, Object NP Region
Adults
Children
Object NP
(e.g., the cake)
Proportion of Fixations
Time After Object Onset (ms)
Wh-Questions
Yes-No Questions

Figure 2. In all figures, shaded areas indicate ±1 standard error. The 0 time point on the x-axis represents the onset of the relevant region, and the dotted lines reflect the analysis region shifted by 200ms to account for saccade latency. Figure 2A: Proportion of fixations on the target patient image during the verb in both question type conditions, with data from adults on the left side and child data on the right side. Figure 2B: Proportion of fixations on the target patient image during the object NP in both question type conditions, with data from adults on the left side and child data on the right side. Figure 2C: Proportion of fixations on the target instrument image during the object NP in both question type conditions, with data from adults on the left side and child data on the right side.

Patient fixations in the verb region. Figure 2A presents the fixations on the patient image in the two question type conditions. The fixation data in Figure 2A illustrates that adults and children showed a different fixation pattern in the two question conditions: Adults fixated on the patient image more often in the *wh*-question condition than in the *yes-no* question condition, while children's fixations on the patient in the two conditions overlapped to a large extent. The linear mixed model analysis of fixation data in the critical regions is summarized in Table 3. There were no significant effects on the intercept. For the slope terms, there was a significant effect of population on time ($\beta = 5.32$, $SE = 2.34$, $p < 0.05$) and time squared ($\beta = -5.18$, $SE = 2.08$, $p < 0.05$), suggesting that adults' fixations on the target patient increased more quickly than children's did. Importantly, there was a significant interaction of question type and population on time ($\beta = 10.36$, $SE = 3.85$, $p < 0.01$) and time squared ($\beta = -10.28$, $SE = 3.41$, $p < 0.01$). These interactions showed that there was a difference in slopes based on the population, and indicated that the question type effect was different between adults and children.

Table 3

Fixed effect summary for the three regions of analysis: the verb region, the object NP region (target patient fixations), and the object NP region (target instrument fixations).

	Verb Region Target Patient		Object NP Region Target Patient		Object NP Region Target Instrument	
Predictor	*β*	*SE*	*β*	*SE*	*β*	*SE*
Intercept Effects						
Question type	0.002	0.60	0.33	0.56	-0.36	0.49
Population	-0.27	0.62	2.37***	0.68	0.34	0.35
Question type x Population	-1.81	1.14	1.28	1.02	-0.74	0.91
Time	3.88*	1.69	4.62**	1.49	1.37***	0.28
Time^2	-1.40	1.43	-3.93**	1.11	N/A	
Slope Effects						
Question type x Time	1.75	1.93	0.24	1.41	1.75**	0.64
Question type x Time^2	-1.78	1.71	-0.76	1.01	N/A	
Population x Time	5.32*	2.34	-7.08**	2.25	0.67†	0.39
Population x Time^2	-5.18*	2.08	4.35*	1.76	N/A	
Question type x Population x Time	10.36**	3.85	-5.74*	2.81	1.92	1.28
Question type x Population x Time^2	-10.28**	3.41	3.49 †	2.02	N/A	

Note: † $p < 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

To explore the details of these interactions, planned pairwise comparisons on the question type effect were conducted for each group. For adults, this analysis revealed no significant effects on the intercept, but there was a significant effect of the question type on the slope for both time (β = 7.31, *SE* = 3.02, $p < 0.05$) and time squared (β = -6.77, *SE* = 2.80, $p < 0.05$). Adults' fixations on the patient image increased more quickly during *wh*-questions compared to *yes-no* questions. Adults' higher proportion of fixations on the object in the *wh*-condition suggests that they actively completed the dependency and adopted a DO gap analysis. These findings are consistent with observations in previous eye-tracking experiments that the current study was modeled after (Omaki, 2010; Sussman & Sedivy, 2003).

On the other hand, planned pairwise comparisons for the child data demonstrated that children did not actively complete filler-gap dependencies in the verb region, as evident in

Figure 2A. There was a marginal difference at the intercept (β = 1.11, *SE* = 0.66, p = 0.09), suggesting that children were more likely to be fixating on the target patient at the beginning of the verb region during *wh*-questions. Question type also had a significant effect on the slope in the opposite direction for time (β = -4.54, *SE* = 2.26, $p < 0.05$) and time squared (β = 4.47, *SE* = 2.10, $p < 0.05$). This suggests that children's fixations on the target patient increased more quickly in the *yes-no* question condition. It is important to note, however, that these two effects counteract one another; because children were less likely to fixate on the target patient at the beginning of the verb region during *yes-no* questions, their fixations on this image increased more quickly to reach the same proportion of fixations as in the *wh*-question condition.

Patient fixations in the object NP region. Although children's fixations on the target patient during the verb did not reliably differ between conditions, visual inspection of the data in Figure 2A above suggests that the fixations on the target patient began to diverge at the end of the verb region. This raises the possibility that children actively associated the filler with the verb, but more slowly than adults did, and the verb region was too short to reveal this delayed active dependency formation process. In order to explore this possibility, fixations on the target patient during the object NP region were examined. If children indeed initiated the active dependency formation later, then the target patient fixations in the *wh*-question condition should be greater than in the *yes-no* question condition during the following NP region.

Figure 2B illustrates fixations on the target patient image in both question type conditions in the object NP region. These graphs are re-aligned such that 0ms represents the onset of the direct object (e.g., *the cake*). The outcome of the mixed effect model analysis with population and question type as fixed effects is summarized in Table 3. The model revealed a significant effect of population on the intercept ($\beta = 2.37$, $SE = 0.68$, $p < 0.001$) and the slope for both time ($\beta = -7.08$, $SE = 2.25$, $p < 0.01$) and time squared ($\beta = 4.35$, $SE = 1.76$, $p < 0.05$). These effects suggest that adults were more likely to be fixating on the target patient at the onset of the object NP region, while children increased their fixations on the target patient more rapidly than adults did. This was consistent with the fixation pattern in the verb region. Additionally, there was an effect of the interaction of question type and population on the slope, which was significant for time ($\beta = -5.74$, $SE = 2.81$, $p < 0.05$) and marginal for time squared ($\beta = 3.49$, $SE = 2.02$, $p = 0.08$).

The significant interaction between population and question type pointed in the opposite direction than the one observed in the verb region. This suggests that children reliably increased their fixations on the patient during *wh*-questions, while this was not the case for adults. This pattern was confirmed in pairwise comparisons of the question type effect within each population. The pairwise comparison for adults' fixation data indicated that there were no significant differences in the object NP region (intercept: $\beta = 0.80$, $SE = 0.68$, $p > 0.1$; time: $\beta = -1.84$, $SE = 3.04$, $p > 0.1$; time squared: $\beta = 0.25$, $SE = 1.56$, $p > 0.1$). Question type also had no significant effect on the intercept for the children ($\beta = -0.38$, $SE = 0.81$, $p > 0.1$), but for the slope, it had a marginal effect for time ($\beta = 3.27$, $SE = 1.89$, $p = 0.08$), and a significant effect for time squared ($\beta = -2.54$, $SE = 1.27$, $p < 0.05$). This adult-child contrast was likely driven by the difference in how soon they fixate on the patient image: in the *wh*-question condition, adults had already fixated on the patient image earlier than children, starting in the verb region. For this reason, adults started to look away toward another image (i.e., instrument; see below) earlier than children did (see Figure 2B). Critically, the reliable increase of patient fixations among children appears to reflect a delayed, 'spill-over' effect of active DO gap analysis.

Instrument fixations in the object NP region. The object NP region (e.g., *the cake*) provided an additional window into active formation of filler-gap dependencies. At this point in the question, the only potential gap position was the upcoming Prepositional Object (PO) gap position, so participants may have generated anticipatory fixations towards the target instrument (e.g., fork) without waiting to encounter the preposition in the input. Figure 2C presents the fixations on the target instrument in both question type conditions separated by population. During this region, children and adults showed a very similar fixation pattern: fixations on the target instrument increased more in the *wh*-question condition than in the *yes-no* question condition, which suggests that both children and adults incrementally adopted the PO gap analysis before the preposition was encountered in the input. The linear mixed effects model output of the instrument fixation data in the object NP region is summarized in Table 3 above. In this region, question type had a significant effect on the slope ($\beta = 1.75$, $SE = 0.64$, $p < 0.01$), and population also had a marginal effect on the slope ($\beta = 0.67$, $SE = 0.39$, $p = 0.096$). This suggests that fixations on the target instrument during the *wh*-questions had a greater positive slope than during the *yes-no* questions, and this difference in slopes was marginally more pronounced for the adults. However, importantly, question type and population did not show a significant interaction, suggesting that the question type effect was not modulated by the population factor.

Summarizing so far, these results suggest that adults' eye movement data demonstrated evidence for active DO gap analysis during the verb region, as well as active PO gap analysis during the object NP region. Children's eye movement data, on the other hand, showed evidence for active DO gap analysis during the object NP region. They also demonstrated an adult-like active PO gap analysis during the object NP region. Next, we investigated how the processing behaviors in children were modulated by age and vocabulary size.

Age and vocabulary size effect on children's eye movement. In the target patient fixation analysis above, data from all children were analyzed together as a group, but this leaves open the possibility that children at different developmental stages show different behaviors. To address this question, children's eye movement data in the verb and object NP region was submitted to two linear mixed effect models, one with age in months as a fixed effect and the other with the raw vocabulary score as a fixed effect.

Table 4 presents the summary of the model output for the linear mixed effect model with age in months. Age had no significant effect on the intercept in any of the regions regardless of the image being fixated, but there were a few age effects on the slope in the object NP region. For target instrument fixations in the object NP region, age in months had a marginal effect on the slope ($\beta = 0.06$, $SE = 0.04$, $p = 0.095$). This indicates that older children increased their fixations on the target instrument more rapidly than younger children, but the lack of interaction between age and question type suggests that the age effect only reflected a difference in the strengths of the same interpretation bias.

Table 4

Fixed effect summary for the age in months analysis including the three regions of analysis: the verb region, the object NP region (target patient fixations), and the object NP region (target instrument fixations).

	Verb Region		Object NP Region Target Patient		Object NP Region Target Instrument	
Predictor	*β*	*SE*	*β*	*SE*	*β*	*SE*
Intercept Effects						
Question Type	2.22	6.45	5.95	7.51	1.38	3.24
Age	0.05	0.05	0.08	0.06	-0.04	0.03
Question Type x Age	-0.02	0.09	-0.09	0.11	-0.02	0.05
Time	6.61	14.42	19.28	14.05	-3.46	2.67
$Time^2$	-2.08	12.74	-10.23	10.85	N/A	
Slope Effects						
Question Type x Time	-2.09	23.65	-30.45	19.32	0.13	2.91
Question Type x $Time^2$	3.26	22.08	22.34†	13.28	N/A	
Age x Time	-0.08	0.20	-0.16	0.20	0.06†	0.04
Age x $Time^2$	0.05	0.18	0.06	0.15	N/A	
Question Type x Age x Time	-0.04	0.33	0.47†	0.27	0.008	0.04
Question Type x Age x $Time^2$	0.02	0.31	-0.35†	0.19	N/A	

Note: † $p < 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

For target patient fixations in the object NP region, there was a marginal effect of question type on time squared (β = 22.34, *SE* = 13.28, p = 0.09). Importantly, there was a marginal interaction of question type and age in months on the slope for both time (β = 0.47, *SE* = 0.27, p = 0.08) and time squared (β = -0.35, *SE* = 0.19, p = 0.06), suggesting that the patient fixation pattern across the two question conditions was somewhat modulated by the age factor. Figure 3 presents eye movement data from 5-year-old children and 6-year-old children separately to illustrate the nature of this age effect. As the figure illustrates, 5-year-old children showed very little difference in patient fixation between the two question conditions, whereas 6-year-old children demonstrated an increase in patient fixations for the *wh*-question condition in comparison to the yes-no question condition similar to the adults' fixation pattern in the verb region. Linear mixed effects models for each age group confirmed this pattern. There was no significant effect of question type for 5-year-olds for the intercept (β = -0.26, *SE* = 0.82, p > 0.1) or slope for time (β = 1.31, *SE* = 2.37, $p > 0.1$) and time squared (β = -0.84, *SE* = 1.84, $p > 0.1$). On the other hand, the models for the 6-year-old group showed no main effect on the intercept (β = -0.33, *SE* = 0.72, $p > 0.1$), but revealed a significant effect of question type for slope for time (β = 4.58, *SE* = 2.26, $p < 0.05$) as well as time squared (β = -3.85, *SE* = 1.79, $p < 0.05$). This suggests that the evidence for children's active DO gap analysis in the object NP region that we observed earlier was mostly driven by the older children, with no clear sign of such active processing in the younger, 5-year-old group.

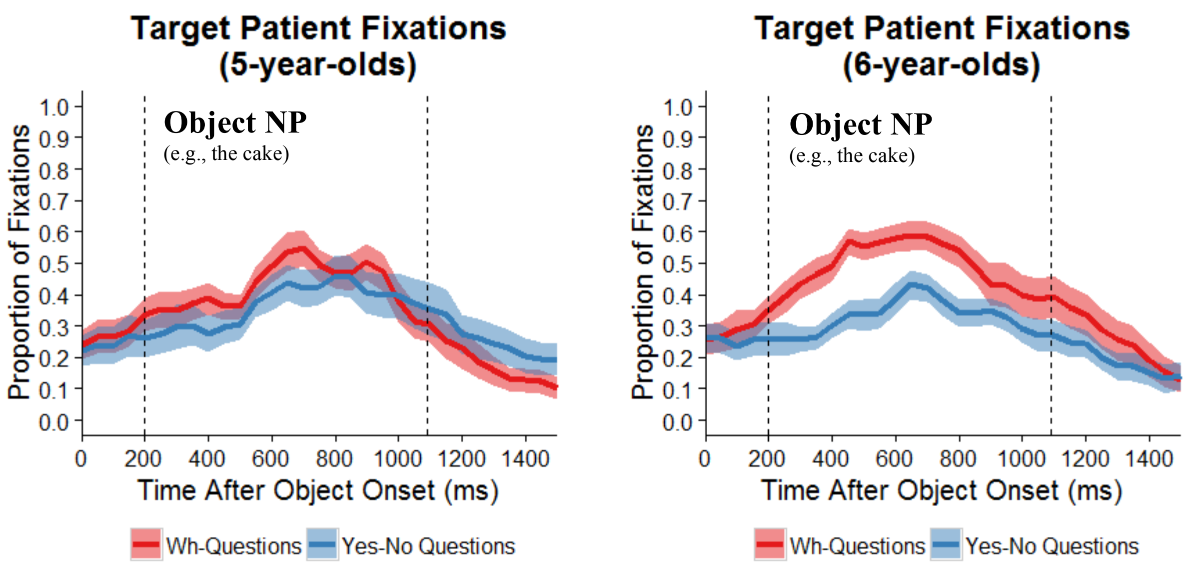

Figure 3. Proportion of fixations on the target patient image in both question type conditions, with data from 5-year-olds on the left side and 6-year-olds on the right side. Shaded areas indicate ±1 standard error. The 0 time point on the x-axis represents the onset of the object NP region, and the dotted lines reflect the analysis region shifted by 200ms to account for saccade latency.

Table 5 presents the summary of the model output for the linear mixed effect model with raw vocabulary score. For target patient fixations in both the verb region and object NP region, the results of this analysis mirrored those of the basic pairwise comparison described above.

Table 5

Fixed effect summary for the raw vocabulary analysis including the three regions of analysis: the verb region, the object NP region (target patient fixations), and the object NP region (target instrument fixations).

	Verb Region		Object NP Region Target Patient		Object NP Region Target Instrument	
Predictor	*β*	*SE*	*β*	*SE*	*β*	*SE*
Intercept Effects						
Question Type	1.12†	0.66	-0.20	0.54	0.03	0.38
Raw Vocabulary	0.03	0.02	0.0008	0.03	0.002	0.01
Question Type x Raw Vocabulary	0.01	0.04	0.05	0.03	-0.06**	0.02
Time	1.02	1.70	7.60***	1.65	1.11**	0.38
$Time^2$	1.46	1.49	-5.75***	1.28	N/A	
Slope Effects						
Question Type x Time	-4.52*	2.26	2.90†	1.65	0.71**	0.27
Question Type x $Time^2$	4.39*	2.09	-2.43†	1.30	N/A	
Raw Vocabulary x Time	-0.13	0.09	0.04	0.09	0.009	0.02
Raw Vocabulary x $Time^2$	0.12	0.08	-0.05	0.07	N/A	
Question Type x Raw Vocabulary x Time	-0.12	0.14	-0.03	0.10	0.09***	0.02
Question Type x Raw Vocabulary x $Time^2$	0.16	0.13	-0.02	0.08	N/A	

Note: † $p < 0.1$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

In the verb region, there was an effect of question type on both the intercept and the slope; it had a marginal effect on the intercept ($\beta = 1.12$, $SE = 0.66$, $p = 0.09$) and a significant effect on the slope both for time ($\beta = -4.52$, $SE = 2.26$, $p < 0.05$) and time squared ($\beta = 4.39$, $SE = 2.09$, $p < 0.05$). As in the basic model, these effects were in opposing directions, and resulted in an overall null effect of question type in this region.

For fixations on the target patient in the object NP region, there was a marginal effect of question type on the slope for both time ($\beta = 2.90$, $SE = 1.65$, $p = 0.08$) and time squared ($\beta = -2.43$, $SE = 1.30$, $p = 0.06$). Children's fixations on the target patient increased more rapidly during *wh*-questions, which also reflected the findings of the simpler model without the vocabulary score.

The inclusion of raw vocabulary score influenced the analysis of the target instrument fixations during the object NP regions. The interaction of raw vocabulary score and question type had a significant effect on the intercept (β = -0.06, *SE* = 0.02, $p < 0.01$). Children with larger raw vocabulary scores were more likely to be fixating on the target instrument at the onset of this region during *yes-no* questions. Importantly, we found a significant effect of question type on the slope (β = 0.71, *SE* = 0.27, $p < 0.01$) and a significant interaction of question type and vocabulary on the slope (β = 0.09, *SE* = 0.2, $p < 0.001$). This interaction suggested that the active PO gap analysis was modulated by vocabulary size.

To further explore the nature of this interaction, we used a median split to divide the child group into high (>126) vs. low (≤126) vocabulary groups, and analyzed eye movement patterns in each of these groups (Figure 4).

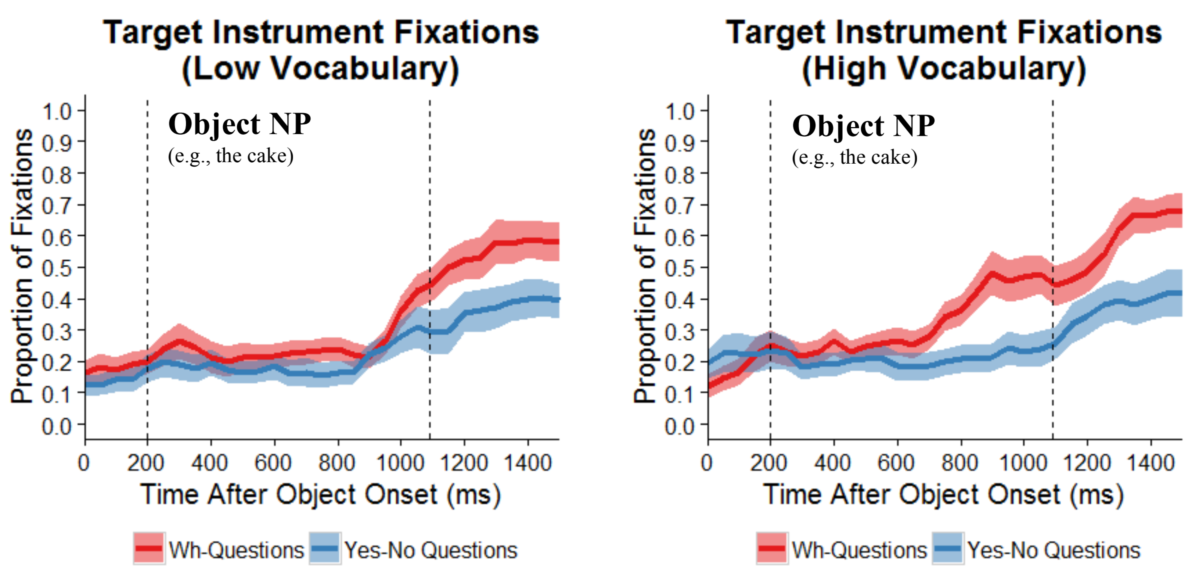

Figure 4. Proportion of fixations on the target instrument image in both question type conditions, with data from the low vocabulary group on the left side and high vocabulary group on the right side. Shaded areas indicate ±1 standard error. The 0 time point on the x-axis represents the onset of the object NP region, and the dotted lines reflect the analysis region shifted by 200ms to account for saccade latency.

The pairwise comparison revealed that there was no main effect of question type for the low vocabulary group for either intercept ($\beta = 0.57$, $SE = 0.57$, $p > 0.1$) or slope ($\beta = -0.24$, $SE = 0.37$, $p > 0.1$). For the high vocabulary group, on the other hand, there was no significant effect on the intercept ($\beta = -0.24$, $SE = 0.46$, $p > 0.1$), but there was a significant effect of question type on the slope for time ($\beta = 1.39$, $SE = 0.40$, $p < 0.001$). This pattern indicates that the high vocabulary group actively adopted the PO gap analysis during the object NP region, while the low vocabulary group did not show robust evidence for such anticipatory fixations.

It is possible that the low vocabulary group might demonstrate the PO gap analysis during the subsequent preposition region as a spill-over effect. This is plausible given the general trend of late emergence of incrementality that we observed for 6-year-old children's patient image fixations. In fact, children of all vocabulary sizes do fixate on the target instrument in this region, but this is for the trivial reason that this region corresponds to the end of the sentence, where the presence of the PO gap becomes evident due to the missing prepositional object. As such, this finding does not speak to a potential spill-over effect, nor alter our conclusions about the vocabulary effect in the object NP region. The data analysis and discussion of instrument fixation data in this preposition region is reported in the Appendix.

3.3. Discussion

The present visual world experiment investigated the time course of filler-gap dependency processing in adults and children. The findings of this experiment are summarized in Table 6. In the verb region, the eye movement data showed that adults actively completed filler-gap dependencies and adopted a DO gap analysis, but no such evidence was found for children, regardless of the age or vocabulary size. This replicates findings from earlier visual world studies by Omaki (2010) and Sussman and Sedivy (2003). However, evidence for children's active DO gap analysis was found in the subsequent object NP region, and this effect was only present in 6-year-old children. These results indicate that active DO gap analysis emerges around age 6, although the execution of this process appears to be slower than in adults. Finally, we also observed that by the end of the object NP region, children with high vocabulary scores demonstrated reliable anticipatory fixations for the target instrument, suggesting that the ability to actively adopt the PO gap analysis is somewhat dependent on having a large vocabulary.

Table 6

Summary of the results of the visual world eye tracking experiment by analysis region and population.

Population	Verb Region	Object NP Region
Adults	✓ active DO gap analysis	✓ active PO gap analysis
Children	× active DO gap analysis	✓ active DO gap analysis (6-year-olds only) ✓ active PO gap analysis (high vocabulary group only)

These findings highlight developmental changes in the filler-gap dependency processing mechanisms around age 6. The lack of evidence for active DO gap analysis in 5-year-old children is consistent with suggestions in the field that the ability to use top-down information during real-time processing develops later than their ability to use bottom-up, lexical information in incremental processing. Active dependency formation relies on top-down syntactic knowledge about what positions constitute the earliest potential gap position, and the integration of this top-down information and language input during real-time processing may plausibly require more language experience or cognitive maturation than is available for 5-year-old children. On the other hand, we also found that active PO gap analysis is modulated by children's vocabulary size, suggesting that either mature lexical knowledge or processing ability is somewhat required for actively interpreting the filler as the instrument. Finally, the absence of an active DO gap analysis was unexpected for the memory constraint account as well as the probabilistic prediction account of active gap dependency formation. We will return to the theoretical implications of these findings in General Discussion.

One alternative explanation for the lack of active DO gap analysis in 5-year-old children is that they were simply too slow in lexical or discourse processing to be able to demonstrate the question type effect within the critical regions. This explanation is not implausible given the complex processes that are required for linking active dependency formation to eye movement in our experiment: Participants must not only process the critical verb (e.g., *eat*) and associate the *wh*-phrase with it, but also recall the events in the story and locate the relevant object in the visual display, and executing this chain of processes may take longer than the duration of our critical regions. However, there are two pieces of evidence against this alternative account. First, previous studies have shown that children typically demonstrate incremental comprehension effects within 600ms of the verb region (e.g. Borovsky et al., 2012), and our critical regions had a longer duration (664ms for the verb region, 889ms for the object NP region). Moreover, unlike the instrument fixation data, we did not find a significant interaction of the vocabulary size and the question type effect on the patient fixation. On the assumption that lexical processing speed correlates with vocabulary development, this lack of interaction suggests that lexical processing was not a critical factor in the lack of active DO gap analysis.

4. General Discussion

The present study investigated children's processing of *wh*-questions to shed light on the development of incrementality in real-time sentence comprehension. The visual world eye-tracking experiment was designed so that two distinct active dependency formation processes (i.e., DO vs. PO gap analyses) could be attested. We found that active DO gap analysis only starts to emerge in 6-year-old children with a slight delay in its execution compared to adults, while the active PO gap analysis was not modulated by age, but rather by children's vocabulary size. These findings elucidate the critical developmental juncture where the child parser develops an incremental processing bias that was not previously available. Moreover, the distributional analysis of *wh*-questions in adult and child corpora suggested that the input contains the appropriate distribution of DO and PO gap questions to support active dependency formation at the verb region. This suggests that distributional information in the input somehow does not affect 5-year-old children's filler-gap dependency processing.

4.1. Comparison to previous works on filler-gap dependency processing in children

The current findings may appear to be at odds with the results of previous studies that reported apparent evidence for adult-like active dependency formation in 5-year-olds (Lassotta et al., 2015; Love, 2007; Omaki et al., 2014). However, as discussed in the Introduction, the cross-modal priming findings in Love (2007) are subject to an alternative explanation that has little to do with filler-gap dependency processing: the apparent reactivation of the filler at the gap position may simply reflect an integration of the display probe as the object of the preceding verb. Under this interpretation, their findings have little to do with filler-gap dependency processing, and therefore are not in conflict with our results. We also note that cross-modal priming results are often not replicated even for adults with presumably more stable processing mechanisms than children (e.g., McKoon, Ratcliff, & Ward, 1994). Taken together, the conclusion from the previous cross-modal priming study may need to be taken with caution.

We now turn to the evidence for a first verb association bias in cross-linguistic studies on complex *wh*-questions (Lassotta et al., 2015; Omaki et al., 2014). In these studies, the two gap positions that are in competition are in two separate clauses (e.g., *Where did Lizzie tell someone ___ that she was gonna catch butterflies ___ ?*), unlike the *wh*-questions in the current study where the two competing gap positions are within the same clause. One way to reconcile these findings is that this structural difference influenced 5-year-old children's ability to execute active dependency formation. For example, 5-year-old children may have an adult-like bias to associate the fronted wh-phrase to the first verb in the sentence, and they may indeed initiate this dependency formation process at the verb region. But this may not be sufficient for causing changes in eye movement data, as the changes in fixations are predicated on the execution of a number of processes that follow dependency formation decisions: memory retrieval of the *wh*-phrase, recollection of events in the story, identification of the relevant image in the scene or the visual working memory, and saccade planning, to name but a few. It may plausibly take a longer time for 5-year-old children to execute these processes than it takes for adults or 6-year-old children (though see Section 3.3 above for potential arguments against this view). Under this interpretation, having a long interval between the two competing gap positions that are separated by a clausal boundary may increase the chances of children's successful execution of the active dependency formation. Relatedly, the use of offline measures in the previous studies may have inadvertently made it easier for children to demonstrate evidence for the incremental processing bias, because there is ample time between the initiation of dependency formation and measurement of the interpretation. Further research is needed to investigate how much time is required for an execution of incremental parsing biases either for adults or children.

4.2. Language experience and active dependency formation

One of the major findings in this study is that 5-year-old children did not actively form filler-gap dependencies in real-time comprehension of *wh*-questions. This was surprising especially in light of the observation that the input for both adults and children contains the skewed distribution of filler-gap dependencies in favor of active dependency formation at the verb region. This observation thus raises a new question about why distributional regularities in the input did not affect 5-year-old children's biases in filler-gap dependency processing, and whether distributional regularities have any relation to the active dependency formation bias.

One potential explanation is that the adult-like active dependency formation bias is shaped by distributional regularities in language experience, but that children's ability to use probabilistic information in real-time comprehension is still immature. For example, children may not consider the distributional regularities as a robust and reliable information source until the biased distribution accumulates longer than 5 years. Plausibly, confidence in probabilistic inferences about upcoming syntactic structures could vary as a function of the reliability of the 'database' of syntactic distributions, and 5-year-old children may not have accumulated sufficient distributional information to make commitments to the active DO gap analysis.

Alternatively, the distributional data itself may be equally accessible and informative for 5-year-old children (see Snedeker & Trueswell, 2004), but they may generally have a higher internal threshold for being willing to make commitments. Suppose that adults and children access similar distributional information and compute the probability of DO gaps as .85 and PO gaps as .15. This probability ratio may be sufficient for adults to make interpretive commitments for the DO gap analysis, but children may not be willing to do so unless the probability of DO gaps reaches a higher value (e.g., .92). This *conservative dependency formation* account could explain why the active PO gap analysis was not modulated by age: once 5-year-old children encounter the object NP and integrate it into their syntactic representation, then the probability of a PO gap is effectively 1, because the story context constrains the possible structural continuation and leaves this PO gap analysis as the only viable option.

However, this account raises a new question as to why 5-year-old children's threshold for active dependency formation is higher than that of adults. We suggest that this conservativity may reflect children's strategy to compensate for their difficulty in revising interpretations that were assigned earlier in the sentence. It is important to note that children do in fact make incremental commitments to sentence interpretations and make blatant interpretation errors (Choi & Trueswell, 2010; Huang et al., 2013; Omaki et al., 2014; Trueswell et al., 1999; Weighall, 2008), as observed in studies on PP attachment ambiguity resolution (e.g., *Put the frog on the napkin in the box*). One possible explanation is that the filler-gap dependencies that we used in our study impose more burden on limited cognitive resources, and this additional demand leads the child comprehension mechanism to be more conservative in processing of filler-gap dependencies. A plausible candidate for this cognitive demand is the working memory resources; as reviewed in the Introduction, filler-gap dependency processing requires memory storage and maintenance of the filler information, as well as subsequent retrieval of this stored representation (e.g., Gibson, 1998; Lewis & Vasishth, 2005). We suggest that these additional demands may prevent the child parser from making risky interpretive commitments. Under this account, by age 6, children's working memory resources must be large enough to make risky incremental commitments in filler-gap dependency processing. This explanation predicts that children should avoid incremental processing for other temporarily ambiguous sentences if their memory resources are taxed.

Future research is needed to explore whether children could indeed flexibly attenuate their incremental processing biases as a function of the availability of working memory and other cognitive resources. In fact, the question of a cognitive pre-requisite for the use of probabilistic information is relevant for adult sentence processing as well. Future research is needed to investigate whether the incremental processing bias in adults is also modulated by individual differences in working memory resources (for some recent attempts with adults in this direction, see Johnson et al., 2016; Nicenboim, Vasishth, Gattei, Sigman, & Kliegl, 2015).

4.3. Working memory and the development of active dependency formation

The potential explanations that we have discussed so far assumed that biases in filler-gap dependency processing reflect probabilistic inferences that are derived from distributional regularities in the input. However, it is also possible to take the present findings to indicate that distributional regularities in the input have no bearing on active dependency formation biases, and rather that working memory constraints are the origin of the active dependency formation bias.

For example, much work has suggested that active dependency formation in adults is a predictive structure building process that is launched upon encoding of the filler information (e.g., syntactic category, animacy, argument vs. adjunct, etc.), and this structure building procedure will be actively maintained in limited working memory and continuously evaluated as the sentence unfolds (e.g., Frazier & Flores D'Arcais, 1989; Lewis & Vasishth, 2005; Wagers & Phillips, 2014). Under this account, robust encoding and maintenance of the filler information is a pre-requisite for active dependency formation, and our findings may indicate that this filler encoding and maintenance mechanism is immature in 5-year-old children. It is important to note that 5-year-old children do encode the filler information as they are able to accurately answer the *wh*-questions. Rather, our suggestion is that the filler information is not encoded robustly enough to trigger incremental evaluation of the structural analysis in real-time.

One explicit implementation of this suggestion is to represent the active status of the filler information in memory in terms of activation level, ranging from 0 (inactive) to 1 (highly active). Adults may be able to robustly and actively encode the filler during real-time comprehension, such that the resulting activation value of the filler is close to 1. However, children may only be able to assign a lower activation value, e.g., 0.5, perhaps due to their generally limited resources, or less efficient lexical and syntactic processes. This activation level may be sufficient for assigning an interpretation after identifying the gap later in the sentence, but the search process may not proceed in an adult-like, incremental fashion.

This memory encoding account makes two predictions. First, 5-year-old children should generally fail to show incremental processing biases for other long-distance dependencies that involve memory encoding and active maintenance of the head of the dependency, such as cataphora dependency (Kazanina, Lau, Lieberman, Yoshida, & Phillips, 2007; van Gompel & Liversedge, 2003). Second, if the memory representations of filler-gap dependency structures are primed through syntactic priming, 5-year-old children should be able to demonstrate an adult-like active dependency formation bias (for discussion, see Atkinson, 2016). Future research is needed to shed light on whether 5-year-old children's incremental processing is indeed restricted to structures that do not require such active maintenance of a long-distance syntactic dependency, and whether their encoding abilities can be manipulated through syntactic priming.

Additionally, future research that includes a measure of children's memory encoding and retrieval abilities may help to shed light on the role of memory capacity in filler-gap dependency processing. Some research has shown that individual differences in working memory span modulate children's processing behaviors in comprehension of complex relative clauses (Arosio, Yatsushiro, Forgiarini, & Guasti, 2012) or resolution of relative clause attachment ambiguities (Felser, Marinis, & Clahsen, 2003), and it remains to be seen whether such correlations extend to predictive structure building biases like active dependency formation bias.

4.4. Lexical processing efficiency and active PO gap analysis

Another important finding in the eye-tracking experiment was that vocabulary size was a reliable predictor of the active PO gap analysis, but not active DO gap analysis. The finding that vocabulary size did not modulate the active DO gap analysis is in fact consistent with research on adults' filler-gap dependency processing, which has repeatedly shown that the adult parser is blind to potentially informative lexical information encoded in the verb (e.g., Omaki et al., 2015; Pickering & Traxler, 2003; Staub, 2007; cf. van Schijndel, Schuler, & Culicover, 2014). On the other hand, the vocabulary size effect for the active PO gap analysis was unexpected.

We suggest that this vocabulary effect is a reflection of how efficiently children can access the lexical information of the object NP input and integrate it into the current parse. In order to use the top-down syntactic information to infer that the PO gap is the only remaining possible analysis, it is necessary for children to efficiently integrate the lexical information (e.g., *the cake*) into the current parse, and subsequently direct their gaze to the appropriate instrument (e.g., fork). It is in this process that lexical processing efficiency could play a role: If this lexical processing speed was not fast enough, then listeners may have already encountered the subsequent preposition *with* before generating the anticipatory fixation. A similar link between lexical processing efficiency and vocabulary size has been previously demonstrated in children as young as 18-months (Fernald & Marchman, 2012; Fernald, Perfors, & Marchman, 2006).

Under this interpretation, given the age effect that we observed for active dependency formation at the verb, the interpretation of the vocabulary size effect on the active PO gap analysis is different for the two age groups. For 5-year-old children who presumably do not postulate an object gap at the verb, integration of the object NP and anticipation of the instrument would constitute the first instance of active dependency formation in the sentence and, therefore, the first attempt at integrating the filler. On the other hand, for 6-year-old children with high vocabulary (as well as adults) who demonstrate active dependency formation in the object NP region, the instrument anticipation is the result of a revision process; the initial analysis of the filler as the direct object must be revised so that the filler can be associated with a PO gap. In fact, the visual inspection of the fixation data in Figures 3 and 4 is consistent with this interpretation: 6-year-olds' fixations on the target patient image peak about 600ms after the onset of the object NP and then begin to decline. On the other hand, instrument fixations by children with high vocabulary scores only begin to emerge around 700ms after the onset of the object NP region, and continue to increase towards the end of the region. In sum, this time course information is compatible with the proposed explanation that a subset of 6-year-olds may be demonstrating two types of incremental analyses (i.e., active DO gap analysis, and its revision to active PO gap analysis) within one region.

In this sense, the present finding can be understood as evidence that the timing of the syntactic revision process varies as a function of individual differences in their lexical processing efficiency. The efficiency in integration of error signals has been assumed to play an important role in sentence revision (for discussion, see Fodor & Inoue, 1994), but to our knowledge, the present finding from 6-year-old children with immature processing capacity may be a novel empirical demonstration of the link between lexical processing efficiency and speed of sentence revision.

5. Conclusion

Our eye-tracking study reported a developmental trajectory of active dependency formation bias: 5-year-olds do not incrementally complete filler-gap dependencies at the first available position while this incremental parsing bias begins to emerge at age 6. A corpus analysis of children's input revealed that children and adults are exposed to approximately the same distribution of gap positions, suggesting that the ability to use distributional information in real-time sentence comprehension may not be fully in place at age 5. We discussed how the current developmental findings provide novel insights on the origin of filler-gap dependency processing, probabilistic models of sentence processing, as well as the role of working memory that supports processing of long-distance dependencies. In sum, developmental investigations can provide a novel window into core properties of sentence processing mechanisms, and ultimately inform and advance theories of human sentence processing more broadly.

6. References

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

Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Meditating the mapping between language and the visual world. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 347–386). New York: Psychology Press.

Anderson, S. E., Farmer, T. A., Goldstein, M., Schwade, J., & Spivey, M. (2011). Individual differences in measures of linguistic experience account for variability in the sentence processing skill of five-year-olds. In I. Arnon & E. V. Clark (Eds.), *Experience, variation, and generalization: Learning a first language* (Vol. 7, pp. 203–221). Philadelphia: John Benjamins.

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language*, *51*, 23–54.

Arosio, F., Yatsushiro, K., Forgiarini, M., & Guasti, M. T. (2012). Morphological information and memory resources in children's processing of relative clauses in German. *Language Learning and Development*, *8*(4), 340–364.

Atkinson, E. (2016). *Active Dependency Completion in Adults and Children: Representations and adaptation* (Doctoral dissertation). Johns Hopkins University, Baltimore, MD.

Barr, D. J. (2008). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457–474.

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

Bates, D. M., Maechler, M., Bolker, B. M., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4* [R package version 1.1-9]. Retrieved from https://CRAN.R-project.org/package=lme4/

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*, 417–436.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Chacón, D. A., Imtiaz, M., Dasgupta, S., Murshed, S. M., Dan, M., & Phillips, C. (2016). Locality and word order in active dependency formation in Bangla. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01235

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272.

Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language*, *52*, 144–169.

Choi, Y., & Trueswell, J. C. (2010). Children's (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, *106*, 41–61.

Chow, W.-Y., Smith, C., Lau, E., & Phillips, C. (2015). A “bag-of-arguments” mechanism for initial verb predictions. *Language, Cognition and Neuroscience*, 1–20. https://doi.org/10.1080/23273798.2015.1066832

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. R. Dowty, L. Kartuunen, & A. M. Zwicky (Eds.), *Natural Language Parsing: Psychological, computational and theoretical Perspectives* (pp. 94–128). Cambridge: Cambridge University Press.

de Villiers, J., & Roeper, T. (1995). Relative clauses are barriers to wh-movement for young children. *Journal of Child Language*, *22*, 389–404.

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

Felser, C., Marinis, T., & Clahsen, H. (2003). Children’s processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition*, *11*(3), 127–163.

Fernald, A., & Marchman, V. A. (2012). Individual Differences in Lexical Processing at 18 Months Predict Vocabulary Growth in Typically Developing and Late-Talking Toddlers: Lexical Processing and Vocabulary Growth. *Child Development*, *83*(1), 203–222. https://doi.org/10.1111/j.1467-8624.2011.01692.x

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*(1), 98–116. https://doi.org/10.1037/0012-1649.42.1.98

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, *47*(2), 250–272.

Fine, A. B., & Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*, 1–14.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLOS One*, *8*(10), 1–18.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, *9*(3), 427–473.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, *27*(2), 285–319.

Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research*, *23*(5), 407–434.

Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language & Linguistic Theory*, *5*(4), 519–559.

Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, *4*(2), 93–126.

Frazier, L., & de Villiers, J. (Eds.). (1990). *Language processing and language acquisition*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Frazier, L., & Flores D'Arcais, G. B. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language*, *28*, 331–344.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178–210.

Gagliardi, A., Mease, T. M., & Lidz, J. (2016). Discontinuous development in the acquisition of filler-gap dependencies: Evidence from 15- and 20-month-olds. *Language Acquisition*, *23*(3), 234–260.

Gambi, C., Pickering, M. J., & Rabagliati, H. (2016). Beyond associations: Sensitivity to structure in pre-schoolers' linguistic predictions. *Cognition*, *157*, 340–351. https://doi.org/10.1016/j.cognition.2016.10.003

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*(1), 58–93.

Garnsey, S. M., Tanenhaus, M. K., & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research*, *18*(1), 51–60.

Gennari, S. P., & MacDonald, M. C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, *111*(1), 1–23.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1–76.

Gibson, E., & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, *7*(1), 55–78.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, *29*(2), 261–290.

Grodner, D., Gibson, E., & Tunstall, S. (2002). Syntactic complexity in ambiguity resolution. *Journal of Memory and Language*, *46*, 267–295.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second conference of the North American chapter of the Association for Computational Linguistics* (Vol. 2, pp. 159–166). Pittsburgh, PA: Association for Computational Linguistics.

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, *32*(2), 101–123.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, *30*(4), 643–672.

Hamburger, H., & Crain, S. (1982). Relative acquisition. In S. Kuczaj (Ed.), *Language Development* (Vol. 1: Syntax and Semantics, pp. 245–272). Hillsdale, NJ: Lawrence Erlbaum.

Huang, Y. T., Zheng, X., Meng, X., & Snedeker, J. (2013). Children's assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language*, *69*(4), 589–606.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L. R., & Trueswell, J. C. (2000). One frog, two frog, red frog, blue frog: Factors affecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research*, *29*(6), 597–626.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*(1), 57–83.

Johnson, A., Fiorentino, R., & Gabriele, A. (2016). Syntactic Constraints and Individual Differences in Native and Non-Native Processing of Wh-Movement. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00549

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*, 137–194.

Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, *56*, 384–409.

Kidd, E., Stewart, A. J., & Serratrice, L. (2011). Children do not overcome lexical biases where adults do: the role of the referential scene in garden-path recovery. *Journal of Child Language*, *38*(1), 222–234.

Kingsbury, P., Strassel, S., McLemore, C., & McIntyre, R. (1997). *CALLHOME American English Transcripts LDC97T14*. Web download, Philadelphia: Linguistic Data Consortium.

Kuczaj, S. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, *16*, 589–600.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *lmerTest: Tests for random and fixed effects for linear mixed effect models* [R package version 2.0-2.9]. Retrieved from http://CRAN.R-project.org/package=lmerTest/

Lassotta, R., Omaki, A., & Franck, J. (2015). Developmental changes in misinterpretation of garden-path wh-questions in French. *The Quarterly Journal of Experimental Psychology*, 1–26.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177.

Levy, R., & Andrew, G. (2006). Tregex and Tsurgeon: Tools for querying and manipulating tree data structures. In *5th International Conference on Language Resources and Evaluation*.

Levy, R., Fedorenko, E., Breen, M., & Gibson, E. (2012). The processing of extraposed structures in English. *Cognition*, *122*(1), 12–36.

Levy, R., Fedorenko, E., & Gibson, E. (2013). The syntactic complexity of Russian relative clauses. *Journal of Memory and Language*, *69*(4), 461–495.

Levy, R. P., & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, *68*(2), 199–222.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*, 375–419.

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198.

Linzen, T., & Jaeger, T. F. (2015). Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science*, 1–30.

Love, T. E. (2007). The processing of non-canonically ordered constituents in long distance dependencies by pre-school children: a real-time investigation. *Journal of Psycholinguistic Research*, *36*, 191–206.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703.

MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk (Version Third). Mahwah, NJ: Lawrence Erlbaum Associates.

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), 843–847.

Marcus, M., Santorini, B., Marcinkiewicz, M. A., & Taylor, A. (1999). *Treebank-3 LDC99T42*. Web download, Philadelphia: Linguistic Data Consortium.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*, 522–523.

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

McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 432–460.

McKoon, G., Ratcliff, R., & Ward, G. (1994). Testing theories of language processing: An empirical investigation of the on-line lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1219.

Mitchell, D. C., Cuetos, F., Corley, M. M., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, *24*(6), 469–488.

Myslín, M., & Levy, R. (2016). Comprehension priming as rational expectation for repetition: Evidence from syntactic processing. *Cognition*, *147*, 29–56.

Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, *86*, 314–329.

Nicenboim, B., Vasishth, S., Gattei, C., Sigman, M., & Kliegl, R. (2015). Working memory differences in long-distance dependency resolution. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00312

Omaki, A. (2010). *Commitment and Flexibility in the Developing Parser* (Doctoral dissertation). University of Maryland, College Park. Retrieved from ProQuest Dissertations and Theses. (Accession Order No. AAT 3426351).

Omaki, A., Lau, E., Davidson White, I., Dakan, M. L., Apple, A., & Phillips, C. (2015). Hyper-active gap filling. *Frontiers in Psychology: Language Sciences*, *6*(384). https://doi.org/doi.org/10.3389/fpsyg.2015.00384

Omaki, A., & Lidz, J. (2015). Linking parser development to acquisition of linguistic knowledge. *Language Acquisition*, *22*, 158–192.

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, *33*, 563–588.

Omaki, A., White, I. D., Goro, T., Lidz, J., & Phillips, C. (2014). No fear of commitment: Children's incremental interpretation in English and Japanese wh-questions. *Language Learning and Development*, *10*, 206–233.

Parker, D. (2017). Processing multiple gap dependencies: Forewarned is forearmed. *Journal of Memory and Language*, *97*, 175–186. https://doi.org/10.1016/j.jml.2017.08.003

Pickering, M. J., & Traxler, M. J. (2003). Evidence against the use of subcategorisation frequency in the processing of unbounded dependencies. *Language and Cognitive Processes*, *18*(4), 469–503.

Pickering, M. J., Traxler, M. J., & Crocker, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language*, *43*, 447–475.

Pinker, S. (1996). *Language learnability and language development* (2nd ed.). Cambridge, MA: Harvard University Press.

Pozzan, L., & Trueswell, J. C. (2015). Revise and resubmit: How real-time parsing limitations influence grammar acquisition. *Cognitive Psychology*, *80*, 73–108.

R Core Development Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent priming at trace positions in children's sentence processing. *Journal of Psycholinguistic Research*, *36*, 175–188.

Sachs, J. (1983). Talking about the there and then: The emergence of displaced reference in parent-child discourse. In K. E. Nelson (Ed.), *Children's language* (Vol. 4, pp. 1–29). Hillsdale, NJ: Lawrence Erlbaum.

Snedeker, J. (2013). Children's sentence processing. In R. P. G. Van Gompel (Ed.), *Sentence processing* (pp. 189–220). New York: Psychology Press.

Snedeker, J., & Huang, Y. T. (2016). Sentence processing. In E. L. Blavin & L. R. Naigles (Eds.), *The Cambridge handbook of child language* (2nd ed., pp. 409–437). Cambridge: Cambridge University Press.

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*, 238–299.

Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*, 227–267.

Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 550–569.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71–86.

Staub, A., & Clifton Jr., C. (2006). Syntactic prediction in language comprehension: Evidence from either...or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 425–436.

Stowe, L. E. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *1*(3), 227–245.

Straub, K., Wilson, C., McCollum, C., & Badecker, W. (2001). Prosodic structure and wh-questions. *Journal of Psycholinguistic Research*, *30*(4), 379–394.

Suppes, P. (1974). The semantics of children's language. *American Psychologist*, *29*, 103–114.

Sussman, R. S., & Sedivy, J. C. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes*, *18*(2), 143–163.

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*, *268*(5217), 1632–1634.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (2nd ed., Vol. 11, pp. 217–262). San Diego, CA: Academic Press.

Thothathiri, M., & Snedeker, J. (2008). Give and take: Syntactic priming during spoken language comprehension. *Cognition*, *108*(1), 51–68.

Tooley, K. M., & Bock, K. (2014). On the parity of structural persistence in language production and comprehension. *Cognition*, *132*(2), 101–136.

Traxler, M. J. (2008). Lexically independent priming in online sentence comprehension. *Psychonomic Bulletin & Review*, *15*(1), 149–155.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*, 454–475.

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.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 528–553.

Valian, V. (1990). Logical and psychological constraints on the acquisition of syntax. In L. Frazier & J. de Villiers (Eds.), *Language processing and language acquisition* (pp. 119–145). Dordrecht, The Netherlands: Kluwer Academic Publishers.

van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 407–430.

van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157–166.

van Gompel, R. P. G., & Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(1), 128–139.

van Schijndel, M., Schuler, W., & Culicover, P. W. (2014). Frequency effects in the processing of unbounded dependencies. In *Proceedings of the 36th annual conference of the Cognitive Science Society* (pp. 1658–1663).

Wagers, M., Borja, M. F., & Chung, S. (2015). The real-time comprehension of WH-dependencies in a WH-agreement language. *Language*, *91*(1), 109–144.

Wagers, M. W., & Pendleton, E. (2016). Structuring expectation: Licensing animacy in relative clause comprehension. In K. Kim, P. Umbal, T. Block, Q. Chan, T. Cheng, K. Finney, … L. Shorten (Eds.), *Proceedings of the 33rd West Coast Conference on Formal Linguistics* (pp. 29–46). Sommerville, MA: Cascadilla Proceedings Press.

Wagers, M. W., & Phillips, C. (2014). Going the distance: Memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology*, *67*(7), 1274–1304.

Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of Experimental Child Psychology*, *99*(2), 75–95.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, *56*(3), 165–209.

Woodard, K., Pozzan, L., & Trueswell, J. C. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology*, *141*, 187–209.

7. Appendix: Instrument fixations in the preposition region

Given that 6-year-olds only demonstrate incremental interpretation of *wh*-questions in the object NP region, it is possible that children with lower vocabulary scores will demonstrate the PO gap interpretation later in the question as a spill-over effect, i.e., during the preposition region. Figure 5A presents the fixations on the target instrument in both question type conditions, separated by population. Children and adults demonstrate similar fixation patterns: fixations on the target instrument are greater in the *wh*-question condition than in the *yes-no* question condition.

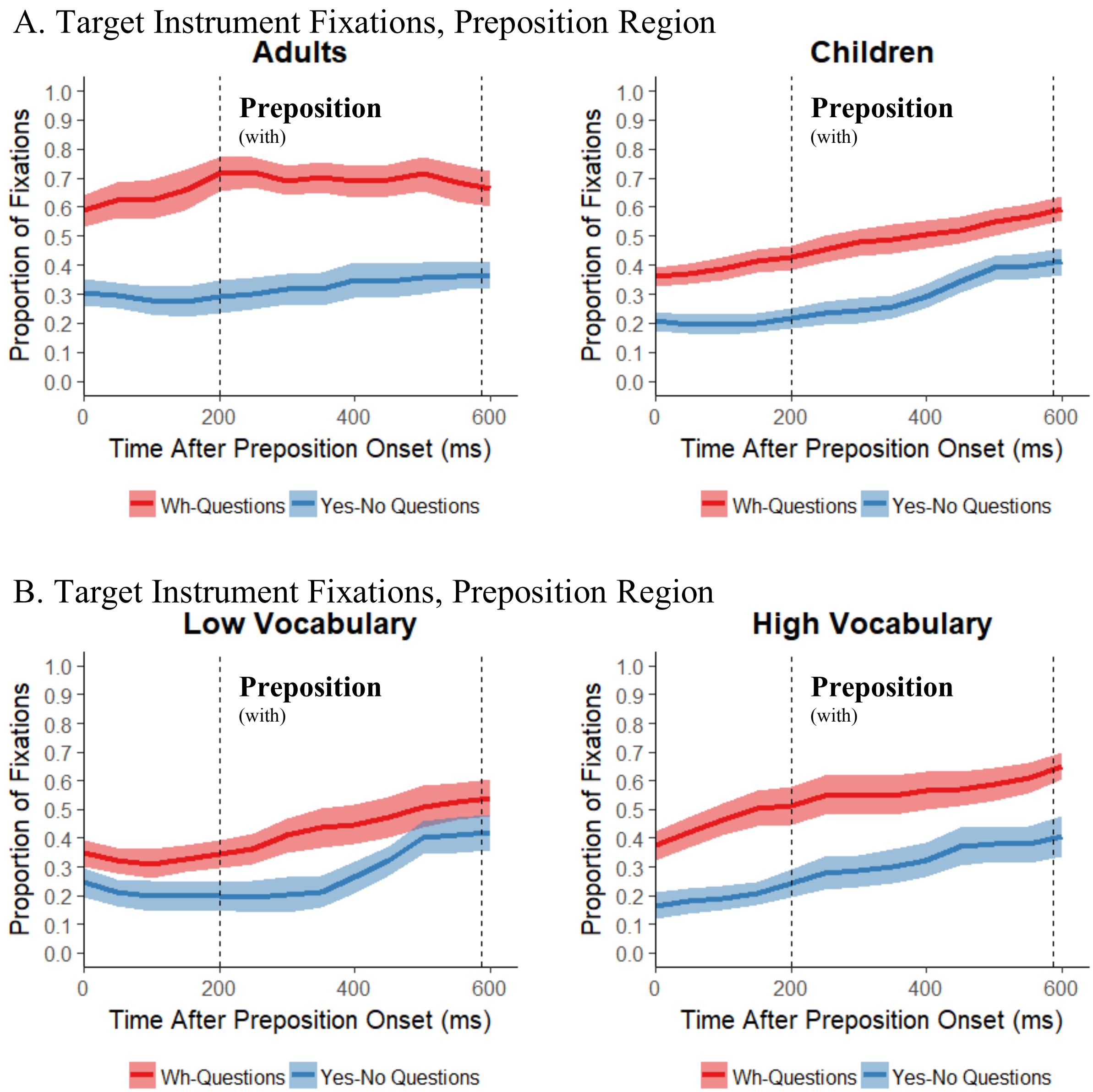

Figure 5. In all figures, shaded areas indicate ±1 standard error. The 0 time point on the x-axis represents the onset of the preposition region, and the dotted lines reflect the analysis region shifted by 200ms to account for saccade latency. Figure 5A: Proportion of fixations on the target instrument image in both question type conditions, with data from adults on the left side and children on the right side. Figure 5B: Proportion of fixations on the target instrument image in both question type conditions, with data from low vocabulary children on the left side and high vocabulary children on the right side.

A linear mixed effects model with question type and population as fixed effects was fit to the target instrument fixation data from the preposition region. On the intercept, there were significant effects of question type (β = 2.28, t = 4.81, $p < 0.001$) and population (β = 1.73, t =

4.24, $p < 0.001$). There were also significant effects of question type and population on the slope (question type: $\beta = -2.37$, $t = -3.16$, $p < 0.01$; population: $\beta = -3.04$, $t = -4.82$, $p < 0.001$). The interaction of question type and population was not significant for either the intercept or the slope. These results indicate that both children and adults were more likely to be fixating on the target instrument at the onset of the preposition during *wh*-questions, but adults were more likely to be fixating on the target instrument at the onset of this region in general. Also, fixations on the target instrument increased more rapidly during *yes-no* questions, and children's fixations generally increased more rapidly than adults.

Figure 5B presents the children's fixations on the target instrument in both question type conditions separated by vocabulary size. First, taking the children's data as a whole, a linear mixed effects model with question type and raw vocabulary score as fixed effects was fit to the target instrument fixation data from this region. On the intercept, question type had a significant effect ($\beta = 1.60$, $t = 3.07$, $p < 0.01$), while raw score had a marginal effect ($\beta = 0.03$, $t = 1.72$, $p = 0.09$). There were no significant slope effects or significant interactions. Children were more likely to be fixating on the target instrument during *wh*-questions at the onset of the preposition region. Thus, all children were associating the filler with the prepositional object position during the preposition region, though children with larger vocabularies were more likely to be fixating on the target instrument at the onset of the region in general. A linear mixed effect model with question type as a fixed effect was fit to each vocabulary group in order to examine variance within the two vocabulary groups (as divided by a median split of the raw vocabulary scores, see Section 2.3.2), but this analysis revealed no additional effects. For low vocabulary children (i.e., raw scores less than or equal to 127), question type did not have a significant effect on either the intercept ($\beta = 1.76$, $SE = 1.17$, $p > 0.1$) or the slope ($\beta = -1.89$, $SE = 2.03$, $p > 0.1$). Similarly, question type did not have a significant effect on either the intercept ($\beta = 0.94$, $SE = 0.79$, $p > 0.1$) or the slope ($\beta = 0.35$, $SE = 1.33$, $p > 0.1$) for high vocabulary children (i.e., raw scores greater than 127).

The fact that children of all ages and all vocabulary levels were fixating on the instrument image during the preposition is unsurprising, because the preposition region is the final region in the *wh*-questions. Given that children correctly answer the questions 97.5% of the time, they must determine that the target instrument is the correct answer at some point during processing. It is reasonable to assume that this decision occurs at the end of the question during the preposition. Thus, fixations on the target instrument during the *wh*-questions are expected during this region in preparation for answering the question.