Wonnacott, E, Joseph, H, Adelman, J and Nation, K

Is children's reading 'good enough'? Links between online processing and comprehension as children read syntactically ambiguous sentences

Wonnacott, E, Joseph, H, Adelman, J and Nation, K (2015) Is children's reading 'good enough'? Links between online processing and comprehension as children read syntactically ambiguous sentences. *Quarterly Journal of Experimental Psychology* , 69 (5). pp. 855-879.

doi: 10.1080/17470218.2015.1011176

Is children's reading 'good enough'? Links between online processing and comprehension as children read syntactically ambiguous sentences

Elizabeth Wonnacott[1], Holly S.S.L. Joseph[2], James S. Adelman[3] and Kate Nation[4]

University College London[1],

Oxford Brookes University[2],

University of Warwick[3]

University of Oxford[4]

Correspondence:

Dr. Elizabeth Wonnacott

Division of Psychology and Language Sciences

University College London

Chandler House, 2 Wakefield Street,

London WC1N 1PF

United Kingdom

e.wonnacott@ucl.ac.uk

Abstract

We monitored 8-and-10-year-old children's eye movements as they read sentences containing a temporary syntactic ambiguity to obtain a detailed record of their online processing. Children showed the classic garden-path effect in online processing. Their reading was disrupted following disambiguation, relative to control sentences containing a comma to block the ambiguity, although the disruption occurred somewhat later than would be expected for mature readers. We also asked children questions to probe their comprehension of the syntactic ambiguity offline. They made more errors following ambiguous sentences than control sentences, demonstrating that the initial incorrect parse of the garden-path sentence influenced offline comprehension. These findings are consistent with 'good enough' processing effects seen in adults. While faster reading times and more regressions were generally associated with better comprehension, spending longer reading the question predicted comprehension success specifically in the ambiguous condition. This suggests that reading the question prompted children to re-construct the sentence and engage in some form of processing which in turn increased the likelihood of comprehension success. Older children were more sensitive to the syntactic function of commas and overall, they were faster and more accurate than younger children.

Keywords*: Children's Sentence Processing; Syntactic Ambiguity; Eye movements; Reading Development; Good Enough Sentence Processing; Reanalysis,*

Word Count: 12909

Is children's reading 'good enough'? Links between on-line processing and comprehension as children read syntactically ambiguous sentences

Many experiments have explored how adults process language when reading sentences containing a temporary syntactic ambiguity. This has resulted in a rich and complex literature that has spawned considerable debate between competing theories of sentence processing (e.g., Frazier, 1987; Frazier & Rayner, 1982; MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; van Gompel & Pickering, 2007; Trueswell & Tanenhaus, 1994). This paper addresses two underexplored aspects of syntactic ambiguity effects on online processing in reading: how children are affected, and how such effects are reflected in comprehension. Relatively little is known about how children process syntactically ambiguous sentences as they read. In the psycholinguistic literature, measuring eye movements is considered to be the most appropriate method to tap online processing as people read (e.g., Rayner, 1998; Rayner, Pollatsek, Ashby, & Clifton, 2012). Strikingly, and in stark contrast to the plethora of studies on adults, only one published paper has used eye tracking to explore syntactic ambiguity resolution in children's reading (Joseph & Liversedge, 2013). Similarly, little is known about the relationship between online processing, considered to reflect the nature of the representations that people construct as they read, and eventual comprehension for what has been read. In this paper we begin to explore this link by recording children's eye movements as they read temporarily ambiguous sentences, and then asking them yes/no comprehension questions where the incorrect response is supported by the initial misanalysis of the temporary ambiguity. We also took the opportunity to explore developmental progression by comparing 7-8 year old children, chosen as being at the earlier stages of reading development, with 10-11 year olds, who are more independent readers. In addition to more advanced reading skills, the older children would have received teaching about the syntactic function of commas, allowing us to determine whether they were more sensitive to the presence of commas in sentences. Thus, our experiment addresses a set of novel questions concerning children's processing and comprehension of sentences containing a temporary syntactic ambiguity.

Temporary syntactic ambiguities in children's language processing

Several studies have explored children's processing of temporary syntactic ambiguities in spoken language, either via act-out tasks or via online methods such as the visual world paradigm and self-paced listening (e.g., Kidd & Bavin, 2007; Kidd, Stewart, & Serratrice, 2011; Snedeker & Trueswell, 2004; Trueswell, Sekerina, Hill, & Logrip, 1999). However, our focus here is on written language. By the time children start to read, they already have considerable experience with spoken language. Processing written language will, to some extent, be a consequence of children accessing their oral repertoire via orthography. However, written language differs from spoken language in important ways. It is lexically more dense and diverse, and it is much more complex: subordinate clauses, relative clauses, participle phrases and passive verb constructions are all substantially more frequent in written language than in speech (e.g., Baines, 1996; Hayes, 1988). Written language also includes punctuations such as commas which can serve a syntactic function. Thus, it is reasonable to assume that children have much to learn about how written language works, meaning that experiments are needed to investigate children's processing of sentences as they read.

Building on the tradition of psycholinguistic experiments with adults, Traxler (2002) reported three reading time experiments investigating 9-13 year-old children's processing of garden-path sentences containing a temporary ambiguity, as in sentence (1). Using self-paced reading, Traxler compared reading at the disambiguating verb (*fell*) in garden path sentences like (1), relative to sentences containing a comma, as in (2).

(1) When Sue tripped the girl fell over and the vase was broken.

(2) When Sue tripped, the girl fell over and the vase was broken.

The children in Traxler's experiments behaved similarly to adults in comparable experiments (e.g., Pickering & Traxler, 1998; Pickering, Traxler & Crocker, 2000), showing longer reading times in (1) than in (2). This indicates that, just like adults, children initially misanalyse a temporarily ambiguous region and that they are able to detect disambiguation online as soon as it occurs. Indeed children's performance was only non-adult like for stimuli in which semantic plausibility was low. Aside from this, the picture that emerges from this study is one in which processing in adults and children is highly similar. It is important to note, however, that Traxler (2002) used a self-paced reading methodology which may be different from naturalistic reading; with this limitation in mind, Joseph and Liversedge (2013) recorded children's eye movements across two experiments as they read sentences containing temporary syntactic ambiguities, comparing 6-to-11 year-olds with adults reading the same sentences. Both adults and children showed longer go-past durations (the sum of all fixations in a region, including regressive eye- movements to the left of the region, until the point of fixation progresses to the region to the right) and a greater likelihood of regressions (Experiment 2) in sentences where a temporary ambiguity had been detected, compared to control sentences without the ambiguity. However a developmental difference was detected: whereas adults showed an effect at the disambiguating word, the effect emerged in the subsequent region in children (Experiment 1), or in a later eye movement measure (Experiment 2) suggesting that they took longer to detect the initial misanalysis. Furthermore, the effect was longer-lasting in younger than older children (Experiment 2), showing that the time course of recovery from an initial misanalysis reduces with age. Joseph and Liversedge (2013) concluded that children's processing mechanisms are similar to those of adults but they are slower and less efficient. However, as this result is limited to a single published report, further investigation is warranted. As well as providing an opportunity to replicate Joseph and Liversedge's finding of delayed processing in children, our investigation addressed other theoretically-motivated questions.

'Good enough' effects in sentence processing

A limitation of the two studies of children's reading discussed above, and one which applies equally to the many experiments with adults exploring the online processing of temporary syntactic ambiguities, is that there was no attempt to link real-time reading processes with reading comprehension. A body of work by Ferreira, Christianson and colleagues suggests that where sentences contain misleading temporary ambiguities, these are not always fully resolved, indicating that initial misanalyses can linger and influence on-going comprehension. Temporarily ambiguous sentences may thus provide a testing ground for beginning to explore the links between online processing and comprehension. We briefly introduce the relevant adult literature before considering its implications for children's reading and the predictions to be tested here.

Christianson, Hollingworth, Halliwell, and Ferreira (2001) asked adults to read sentences containing a syntactic ambiguity. They used self-paced reading to present sentences and to assess the time course of reading. Unlike other studies, however, they asked comprehension questions that probed the initial misanalysis. For example after reading a sentence like (3), participants were asked a comprehension question, as in (4).

(3) While Bill hunted the deer ran into the woods.

(4) Did Bill hunt the deer?

Sentence (3) contains a garden path in which readers initially attach the second noun phrase (*the deer*) as direct object of the verb *hunt*, rather than the subject of the main clause. Consequently, reading times are typically longer on the disambiguating verb (*ran*) in this sentence, relative to a control sentence without a garden path. It is sometimes assumed that these longer reading times index identification of the misanalysis, and appropriate reanalysis, so that on-going comprehension is supported (e.g., Frazier & Rayner, 1982). However, Christianson et al. found that adults were highly likely to answer 'yes' to the question in (4) and more so than following control sentences in which the ordering of the main and subordinate clauses was reversed (5) so that there was no temporary ambiguity:

(5) The deer ran into the woods while Bill hunted.

They suggest that the greater number of "yes" responses following (3) is due to a lingering misanalysis of the temporary ambiguity and that initial thematic role assignments are surprisingly resilient to revision. Building on these findings, Ferreira, Bailey, and Ferraro (2002) described a 'good enough' approach to sentence comprehension, whereby processing is sometimes only partial and the semantic representation stemming from processing is incomplete. Complementary evidence in support of Ferreira et al.'s approach comes from studies using structural priming (van Gompel, Pickering, Pearson, & Jacob, 2006) paraphrasing (Patson, Darowski, Moon, & Ferreira, 2009) and eye-tracking (Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013; Sturt, 2007), suggesting that the good enough effect is not an artefact of Christianson et al. (2001) asking questions that probed the ambiguity directly.

To date, there has been no investigation of 'good enough' effects in children's reading. Although both of the previous studies examining children's online reading of syntactically ambiguous sentences (Traxler's,2002, self-paced study and Joseph and Liversedge's, 2013, eye-tracking study) included comprehension questions, these were designed to keep children's focus on the task. Questions occurred on a minority of trials (25%) and did not probe the ambiguity. However, evidence from spoken comprehension suggests that children in fact have even greater difficulty than adults revising incorrect parsing decisions (Trueswell et al., 1999; Weighall, 2008). There is thus good reason to predict that children also make 'good enough' errors in reading. Accordingly, an aim of our experiment was to establish that this was indeed the case by examining the extent to which children make 'good enough' errors when reading. As in Christianson et al. (2001), we used garden-path stimuli in which the subordinate clause verb was intransitive but the subject of the main clause was likely to be mis-analysed as its direct object. Also following Christianson et al., we manipulated verb type in our stimuli such that half of the subordinate clause verbs were optional transitive verbs (OT; e.g., *eat, drink, count*), and half were reflexive absolute transitive verbs (RAT; e.g. *dress, wash, hide*). Note that the critical difference between these verbs types is that, unlike OT verbs, RAT verbs retain their transitive argument structure even when no object is present, with the thematic patient theme of the verb being co-referential with the subject (e.g. "Anna dressed" is synonymous to "Anna dressed herself"). Example stimuli with an OT verb and RAT verb are given in (6) and (8) respectively. Questions probing the ambiguity are shown in (7) and (9).

(6) While Jim was eating the biscuits baked in the oven. (OT; garden-path)

(7) Did Jim eat the biscuits?

(8) While Anna was dressing the baby played in its cot. (RAT; garden-path)

(9) Did Anna dress the baby?

Performance in these conditions was contrasted against two types of non-ambiguous control. This is important as it allows us to detect whether incorrect answers to the comprehension questions are driven by general inference, rather than a lingering consequence of an initial syntactic misanalysis. Sentences in the comma condition, as in (10) and (11), provided an ideal control for the analyses of eye movement data as word order is identical. As little is known about whether children of this age can use commas to disambiguate ambiguous syntax (see below for further discussion), we included a reversed condition in which the order of the subordinate and main clause was switched, as in (12) and (13). Poor performance due to general inference should occur equally on reversed and garden path sentences, so that any greater difficulty in garden path sentences can be assumed to be due to the temporary syntactic ambiguity. Although it is clear good enough effects cannot be fully explained by an inference account in adults (see Christianson et al., 2001 for details), the fact that more errors are made when sentences contain an OT verb might be because OT verbs allow for more inference in their intransitive forms: as the object remains unspecified (i.e., Jim was eating SOMETHING), this leaves open the possibility that the something that was eaten was the biscuits. By contrast, the intransitive form of a RAT verb is reflexive and thus its direct object is specified (i.e., Anna was dressing HERSELF). Importantly however, Christianson et al. (2001) found that adults made good enough errors to sentences containing both types of verb relative to control sentences. If children's processing is adult-like, they should do the same. We thus include both verb types to gain a fuller picture of children's propensity to engage in good enough processing when reading.

(10) While Jim was eating, the biscuits baked in the oven. (OT; comma)

(11) While Anna was dressing, the baby played in its cot. (RAT; comma)

(12) The biscuits baked in the oven while Jim was eating. (OT; reversed)

(13) The baby played in its cot while Anna was dressing. (RAT; reversed)

Eye movement data and good enough processing

The fact that the people make comprehension errors that are indicative of good enough processing does not tell us why such errors are made. As a first step to address this issue, in addition to asking comprehension questions, we also recorded children's eye movements whilst reading sentences. In particular, we examined the processing of garden-path sentences, as in (6) and (8), and comma sentences, as in (10) and (11), which had an identical word order but no temporary ambiguity due to the presence of the comma. This allowed us to determine the effects of temporary ambiguity in children's sentence processing. Would we see increased reading times and regressive eye-movements at the disambiguating verb in garden path compared to comma sentences, as would be expected in adults? Would this effect show up in a later region than is typically seen in adults? Monitoring eye movements also allowed us to explore the role of regressive eye-movements in garden path sentences. Note that the targeting of regressions is a relatively unexplored area, even in the adult literature (notable exceptions are: Frazier & Rayner, 1982; Meseguer, Carreiras, & Clifton, 2002;

Mitchell, Shen, Green, & Hodgson, 2008; von der Malsburg & Vasishth, 2011). There is some evidence that children target regressions back to specific areas after encountering a syntactic ambiguity (Joseph & Liversedge, 2013), and that poorer comprehension skill in children is associated with fewer regressions when reading texts containing anaphors (Murray & Kennedy, 1988). As a preliminary investigation, we used the fact that our garden path sentences contained two different verb types, and looked to see if children showed differences in the patterns of regressions for garden-path sentences containing OT vs. RAT garden path sentences. Bearing in mind that a correct reanalysis of RAT verbs requires their subject noun to be given an additional thematic role in the re-analysis we might predict more regressions to the first noun (Anna) in (8) (Anna is the theme as well as the agent of dressing) than we would to the first noun (Jim) in (6) (Jim is the agent but not the theme of eating in).

Most critically, we were able to take advantage of the availability of both answers to the comprehension questions *and* eye-tracking data for the garden-path sentences, allowing us to relate patterns of eye movement behaviour to comprehension. If children detect an anomaly while reading and consequently engage in reanalysis, we predict that they should spend longer reading the disambiguating regions of garden path sentences, or make more regressions when reading such sentences. In turn, this pattern of reading behaviour should be associated with better performance on the comprehension questions. Against this prediction however, Christianson and Luke (2011) found remarkably little relationship between patterns of online reading latencies (as measured by self-paced reading) and offline comprehension for 'good enough' questions. To our knowledge, no published study includes analyses relating online behaviour as assessed by eye-movements with offline comprehension, even in adult skilled readers. Preliminary data reported by Christianson and Luke (2011) again found a lack of association between online reading behaviour as assessed by eye movements, and offline comprehension. Interestingly, Christianson and Luke (2011) did find a relationship between how long adults spent reading questions such as (4) and comprehension accuracy, with longer reading times being associated with better performance. Moreover, this relationship was only observed in the garden-path condition. This observation fits comfortably with the idea that full reanalysis might not happen until the reader is forced to examine specific aspects of the initial 'good enough' parse. If the reader can remember the sentence and engage in this form of reconstruction while reading the question, this might well impact reading times such that more time spent reading the question is predictive of accuracy of response. In our experiment, we also recorded reading times on the question, providing an opportunity to test this prediction. Furthermore, the availability of eye-movement data (rather than self-paced reading data) provided an opportunity to relate offline accuracy to both reading time measures and to the probability of making regressions into particular regions.

Children's sensitivity to commas when reading

A final aim of our experiment was to explore children's sensitivity to commas when reading. In adults, the comma in a sentence such as (10) or (11) induces initially longer reading times on the comma region (Hirotani, Frazier, & Rayner, 2006; Rayner, Kambe, & Duffy, 2000). This is likely to be a reflection of influences from inner speech, with prosodic phrasing being used strategically to aid efficient processing and comprehension. Consistent with this, increased reading times at the comma are associated with faster reading times for the sentence overall, an observation which in turn is consistent with data showing that commas can block syntactic misanalysis (e.g., Hill & Murray, 2000; Mitchell & Holmes, 1985). Once again however, there is a dearth of studies examining how commas influence children's reading. We know of no eye movement data, and only one online study – Traxler's (2002) self-paced reading experiments with children, discussed earlier. He reported that 9-13 year-old children spent longer reading the determiner immediately following the comma in sentences like (2), relative to the same word in sentence (1), but were faster at reading the disambiguating verb (e.g. *fell*) in comma sentences. This increase in processing time after the comma suggests that children are sensitive to the role of the comma. Beyond this however, very little is known about how and when children begin to process commas in the service of comprehension as they read. There is some evidence that experience matters, with adults who have strong 'punctuation habits' being more sensitive to commas in online reading than those with less advanced knowledge of punctuation (Steinhauer & Friederici, 2001). In the UK, children are first taught about commas when they are in Year 2 (6-7 years of age). At this time, they are taught that commas are used to separate items in a list. However, it is not until Year 4 (8-9 years of age) that children are taught that commas are also used to mark clauses. By comparing younger (Year 3) and older (Year 6) children, our experiment was able to determine whether there are age differences in children's sensitivity to commas. Measures of reading time at the comma region will indicate if children, like adults, spend longer in this region. More importantly, if children's eye-movements show signature effects of syntactic ambiguity in sentences like (6) and (8) compared to (10) and (11), we can conclude that they are indeed aware of the role of the comma in separating the clauses. In addition, since we included comprehension questions following each different sentence type (garden-path, comma and reversed) we have an additional offline measure of the extent to which the commas fully block the temporary ambiguity. Assuming that there are more errors with garden-path sentences, indicating that initial syntactic misanalysis does indeed affect comprehension, we can look to see whether comma sentences pattern like garden-path sentences (indicating that the comma did not fully block the initial temporary ambiguity) or reversed sentences (indicating that it did), thus providing important information not only about whether children notice commas during online reading, but whether this impacts on their eventual comprehension of what they've read. This is an area of reading skill that has been ignored in theories of reading development, but it seems reasonable to predict that older children will be more sensitive to the syntactic function of commas in written language than younger children.

In summary, our experiment allowed us to investigate a number of novel factors concerning children's reading of sentences containing a temporary syntactic ambiguity. We monitored children's eye movements as they read garden-path and control sentences. This provided a moment-by-moment window on processing, allowing us to determine whether and when children detected the ambiguity, and the effect this had on their patterns of eye-movements. Comprehension questions were also included and these probed the potentially lingering effects of the initial misanalysis - so called 'good enough' processing effects. Critically, the availability of eye movement data and comprehension data on the same items allowed us to relate indices of sentence processing with subsequent comprehension; this has not been done before and is thus exploratory, although our analyses were guided by a priori predictions. We included two non-ambiguous control conditions: sentences with clauses reversed to provide a strong baseline for investigating comprehension performance and sentences where a comma served to disambiguate the syntax; as comma sentences shared an identical word order with garden-path sentences, these provided an appropriate comparison for examining online processing, as well as allowing us to explore children's sensitivity to commas as they read.

Method

Participants

Two groups of primary school-aged children completed this experiment: 42 younger children recruited from Year 3 (age *M* = 8;4, *range* 7;7- 9;1; 21 female) and 39 older children recruited from Year 6 (age *M* =10;8, *range* 10;0 to 11;6; 22 female). All were monolingual English speakers. An additional 11 children were recruited but their data were not included in this report (due to calibration problems (*N*=1), lazy eye preventing accurate tracking (*N*=1), technical problems (*N*=4) and two children not reaching minimum standards in reading skill (see below for details); a further three children were excluded on a 'last out' policy to maintain equal numbers of children completing each counterbalanced list of stimuli). The experiment was part of a battery of unrelated experiments conducted with the same children. All children received a sticker and a certificate in return for taking part in the experiment.

To ensure that all children were proficient enough readers to cope with the experimental materials, we administered the Test of Word Reading Efficiency (TOWRE; Torgeson, Wagner, & Rashotte, 1999), a standardized assessment of word level reading ability. In this test, children read aloud lists of words or nonwords and the score is calculated as the number of items read correctly in 45 seconds. Two children scored below normal range (equating to a standard score less than 80); consequently, they were excluded from this experiment. The children whose data were included in the experiment showed average-for-age reading skills (population norm for standard scores is 100; $M_{\text{younger}} = 105$, $SE_{\text{younger}} = 2.0$; $M_{\text{older}} = 101$, $SE_{\text{older}} = 1.6$).

Design and Materials

The design comprised one between-groups factor, age (younger vs. older) and two within-group factors, sentence-structure (garden-path vs. comma vs. reversed) and verb-type (OT verbs vs. RAT verbs). Eye movements were recorded as children read the sentences. After each sentence, a comprehension question appeared on the next screen. Children responded yes/no via a button press; reading time for the question was computed from the onset of the question to the button press.

It should be noted that to fit around the children's classroom routines, the experiment needed to be relatively short, restricting the number of items we could include. Twelve garden-path sentences were constructed, each comprising a subordinate clause preceding a main clause (e.g., *While Jim was eating the biscuits baked in the oven*). The 12 comma sentences were identical except that the main and subordinate clauses were separated by a comma (e.g., *While Jim was eating, the biscuits baked in the oven*). In the reversed condition, clause order was switched so that the main clause always preceded the subordinate clause (e.g., *The biscuits baked in the oven while Jim was eating*). Across all three conditions, half the sentences contained an OT verb (e.g., eat) and half a RAT verb (e.g., dress). For OT sentences, where we predict that children will be particularly likely to make incorrect inferences in-line with the misanalysis of the garden path, we constructed our sentences to minimize support for this inference; in the example above, what Jim is eating could not actually be the biscuits, since they are currently in the oven. Each sentence was followed by a yes/no question which probed the interpretation of the subordinate clause, consistent with the initial misanalysis (e.g., *Did Jim eat the biscuits?*). The correct answer to this question was always no.

The factorial combination of verb type and sentence structure yielded 36 sentences. These were counterbalanced so that each child saw each of the 12 verbs in only one of the three different sentence structures. This meant that each child saw two items per cell: 2 OT verbs and 2 RAT verbs in each of garden-path, comma, reversed conditions (we address potential issues with power in the General Discussion). In addition to the experimental items, 36 filler items were presented, with each child seeing every filler, each again followed by a yes/no question (e.g., sentence: Henry kept score while Jack was kicking the ball. Question: Did Jack kick the ball?). The fillers had a similar structure to the experimental sentences, each comprising a main clause and subordinate "while" clause. They were created so as to maintain a balance in the ordering of main and subordinate clauses, and the presence of commas. There were no garden path sentences in the fillers. Across experimental and filler sentences, the correct answer to the comprehension questions was yes or no equally often. The full set of both experimental and filler sentences, with their corresponding questions, are given in Appendix A.

Procedure

Stimuli were presented using EyeTrack software (taken from http://www.psych.umass.edu/eyelab/software) and eye movements were recorded using an Eyelink 1000 eye tracker (SR Research; Mississauga, Canada). This is an infrared video-based tracking system combining hyperacuity image processing with a spatial resolution of 0.4 degrees. Eye movements were monitored at a rate of 1000Hz. Sentences were presented in white, Courier New font, size 18, on a black background and children read the sentences silently from a computer monitor at a viewing distance of 62 cm and each character covered 0.24° of horizontal visual angle. Although the children read binocularly, only the movements of the right eye were monitored. The children sat in a customised chair and chin and forehead rests were used to minimize head movements and ensure comfort.

To calibrate, children were instructed to look at each of three horizontal fixation points. They then fixated a small box at the left of the screen and the sentence appeared, contingent on their gaze. The children were asked to read the sentence normally and to press either of two buttons on a response box when they had finished reading the sentence. This served to terminate the display and generate the presentation of the question on a new screen. After reading the question, children chose from two options (yes or no) by clicking the left or right button on the response box.

The experiment was divided in half and conducted over two sessions on separate days, with fillers and experimental sentences spread evenly across the sessions and presented in a randomized order within each session. Each session lasted approximately 20 minutes.

Results and Discussion

Our results are organized into four sections, each accompanied by some discussion of the key findings. We begin with analyses on the comprehension questions, addressing whether children are susceptible to 'good enough' errors, and how this might be moderated by the experimental manipulations. Second, we summarize the eye movement data, examining reading times and regressions. Our third section reports analyses relating eye movement indices to performance on the comprehension question. Finally, we examine the relationship between the time spent reading the question and answering it correctly.

Data were analysed in the package *lme4* (Bates, Maechler, & Bolker, 2013) for the R computing environment (R Development Core Team, 2010) using linear mixed effects models (LMEs; Baayen, 2008; Jaeger, 2008; Quené & van den Bergh, 2008).Binary data were analysed with logistic models and time data with Gaussian models. Given the large number of models in this paper, we do not describe all of the details of each model at each point in the text, instead focusing on the specific fixed main effects and interactions of interest. However, our approach was to include all experimentally manipulated variables (and all the interactions between them) as fixed factors in a given model, regardless of whether they played a role in the specific predictions for that model, or whether they contributed significantly to the model (i.e., we did not use stepwise model comparison). Despite including all effects, we primarily considered only the effects relevant to our questions and predictions.

Random effects in the model were participants (n=81) and items (n=12). The fixed factors were *year-group* being between participants (Year 3 N=42, Year 6 N=39) and within items, *verb-type* being between items (RAT N=6, OT N=6) and within participants and *sentence-type* being both within items and within participants. We also aimed to use a full random slope structure in each model; that is, by-subject slopes for all within-subject main effects and interactions, and by-item slopes for all within-item main effects and interactions (as Barr, Levy, Scheepers, & Tily, 2013, showed this has the best statistical properties). A known problem with using maximal random slopes structure is that these more complex models may not converge. For the current analyses, the models only including experimentally manipulated factors as predictors (sections 1 & 2 below) all converged with maximal random slopes. However, those including reading time measures as predictors of offline performance (section 3) did not. For these non-converging models, we identified a converging solution with the same fixed effects and intercepts but a simpler random slopes structure (identified by successively removing random slopes for the factors of least interest). All analyses reported in the text relate to models that converged. In general, the critical pattern of significances from the converging/non-converging versions was identical. Details of the particular random slope structure used in each case is given as a footnote, along with details of where the converging and non-converging models differed for the few instances where this occurred.

All predicting variables (including discrete factor codings) were centred to reduce the effects of collinearity between main effects and interactions, and in order that main effects were evaluated as the average effects over levels of the other predictors[i].

For each analysis we report t/z statistics alongside p-values for fixed effects of specific interest (note that as the lme4 package does not provide p-values automatically for mixed effects models with a continuous outcome variable, these were calculated using a model comparison procedure).

1. Do children make comprehension errors indicative of 'good enough' processing?

Before discussing the performance of the children on the experimental sentences by condition, it is useful to review performance on the filler items: fillers were the only items in the experiment where the correct answer was yes, and as such, they are useful to identify whether children have a strong 'yes' bias (which might contribute to poor performance on experimental items). Overall, accuracy on the fillers was good, averaging 86% (younger children: yes questions $M = 83\%$, $SE = 2\%$, no questions $M = 79\%$, $SE = 3\%$; older children: yes questions $M = 93\%$, $SE = 1\%$, no questions $M = 87\%$, $SE = 2\%$). A linear mixed effects model with accuracy of response (1 or 0) as the predicted measure, found a reliable effect of age ($z = 4.5$, $p < .00001$) but no difference for the yes ($M = 88\%$) vs. no ($M = 83\%$) answers ($z = 1.48$, $p = .14$); nor was there a reliable interaction of answer-type with age ($z = 1.71$, $p = .09$).

The children's performance on the experimental questions is summarized in Figure 1. The data were analysed in an LME with accuracy of response (1 or 0) as the predicted measure. Recall that the correct answer to each question for the experimental items is no. Performance was generally poor across all conditions, although overall older children marginally outperformed younger children (younger $M = 49\%$, older $M = 58\%$; $z = 1.8$, $p = .07$). Turning to the contrasts between the different sentence types, children made fewer errors following reversed sentences ($M = 61\%$) than garden-path sentences (*M* = 42%; z =2.70, p <.01); this contrast did not interact with either verb-type, age or the interaction between them (*ps* >.4). This suggests that, like adults in previous studies, both age groups are sensitive to the garden path for both verb types, and that this lingers. Children also performed reliably better on comma (*M* =57%) than garden path sentences (*M* = 42%; z =3.52, p <.0005), indicating that they are sensitive to the syntactic function of the comma. This contrast did not interact with verb-type or verb-type by year-group (*ps*>.6) but there was a marginal interaction with age (z =1.80, p =.07). Given that we predicted that the older children would be more sensitive to commas, we investigated this interaction further. Both age groups were reliably more error prone on garden path sentences than comma sentences (older $M_{\text{garden path}}$ =44%, M_{comma} = 67%; z =3.23, p <.005; younger $M_{\text{garden path}}$ =40%, M_{comma} = 49%; z =2.07, p <.05). While the older children were significantly better than the younger children in the comma condition (z =3.01, p <.01), the two groups performed equivalently in the garden path condition (z = 0.56, p =.58). Thus, both groups were able to use commas to block the formation of the garden path meaning to some extent but the finding that the older children outperformed the younger children in the comma condition suggests a fuller mastery of the role of commas for the older children.

To further probe whether the comma completely blocked the garden-path, we compared performance with reversed sentences (where there is no opportunity to garden path) and comma sentences where full sensitivity to the comma should block the garden path). This contrast was not reliable (M_{reversed} = 61%, M_{comma} = 57%; z = 0.66, p =.51) and it did not interact with verb-type or verb-type by age (*ps* >.7). There was however a reliable interaction with age (z =2.0, p <.05). Younger children showed a tendency to perform better in the reversed than comma condition (*M* = 59% vs. *M* =49%; z =1.66, p <.1) but this was not true for older children (M_{comma} =67%, M_{reversed}= 63%; z =0.71, p >.4). While older children showed stronger performance than younger children on commas sentences, there was no age effect for reversed sentences (M_{older} =63%, M_{younger}= 59%; z =-

.79, $p = .43$). Thus the general pattern to have emerged showed older children to be somewhat more sensitive to the syntactic commas.

Turning to explore the role of verb-type, as in previous experiments with adults (Christianson et al., 2001) overall performance was reliably better for RAT sentences (61%) than OT sentences (45%; $z = 4.35$, $p < .00005$). We saw above that verb-type did not interact with the contrasts between sentence type; however there was marginal interaction between verb-type and age ($z = 1.91$, $p = .06$). The contrast between OT and RAT was reliable for both groups (older $M_{OT} = 47\%$, $M_{RAT =}$ 68%; $z = 3.98$, $p < .0001$; younger $M_{OT} = 44\%$, $M_{RAT =}$ 54%; $z = 2.5$, $p < .05$), the difference between the age-groups was only reliable for RAT sentences (18% difference; $z = 2.48$, $p < .05$ and not the OT sentences 3% difference; $z = .43$, $p = .66$). This suggests that older children may have a stronger understanding of the nature of reflexive verbs; however given that the interaction was marginal (and not predicted) it is not further considered.

In summary, it is clear that children are susceptible to 'good enough' effects when answering questions that probe the initial (but eventually incorrect) analysis of garden-path sentences containing a temporary syntactic ambiguity. While both age groups seemed to be aware of the role of the comma, there was some evidence that this this knowledge may be more secure in the older children. In particular, older children outperformed younger children on comma sentences, but not on either garden path or reversed sentences, and only the older group performed equivalently with the two sentence types. Like adults (Christianson et al., 2001), children found OT verbs harder. This was the case for all sentence types, including the unambiguous reversed condition. There was some evidence that this effect was modulated by age, with both groups showing the effect but older children benefitting more from the presence of the reflexive verb in RAT sentence. This may reflect development in the sensitivity to the reflexive syntax, although caution must be taken in interpretation given this was unpredicted and rests on a marginal interaction.

2. Online reading of garden-path and comma sentences as revealed by eye movements

Garden-path and comma sentences are identical in structure apart from the presence of the comma. Comparing the eye movement record across these two conditions allowed us to answer three questions. First, do children show adult-like processing at the region containing the comma (i.e. evidence of longer reading times at the subordinate verb when they encounter the comma than when they do not)? Second, do children show a garden path effect, that is, longer reading times and more regressive eye movements when they encounter the disambiguating material in a garden-path sentence, relative to a comma sentence? Third, do children look back to informative regions following disruption to processing in garden-path sentences?

These questions were explored in a series of mixed effects models with the relevant eye movement measure (reading times or number of regressions) in a particular region as the predicted measure. Our regions of interest varied across analyses but included: the whole sentence (*While Anna was dressing the baby played in its cot*), the first noun phrase (NP1) (*Anna*), the subordinate verb (*dressing*), the second noun phrase (NP2) (*the baby*), the disambiguating verb region (*played*) and the post-disambiguating verb region (*in its*). For all analyses, fixations longer than 1200ms were excluded from the data set. Fixations less than 80ms which were within a character of another fixation were summed, otherwise they were excluded. In addition, for each analysis cells were coded as missing if (i) the participant made more than 60 fixations in a particular region (ii) there were no first pass fixations at all in that region. On average of 14% of data was missing data across the models (min 6%, max 19%).

(i) Do children show longer reading times at the subordinate verb when they encounter the comma?

If children are processing the comma in an adult-like manner, we would expect longer reading times on the region that includes a comma, that is, longer reading times at the subordinate verb immediately preceding (and including) the comma, compared to sentences without a comma (e.g., Hirotani et al., 2006). In addition to this prediction, we anticipated that if older children are more sensitive to the utility of commas, they should show a greater increase in reading times in this region for comma vs. garden-path sentences, relative to younger children. The reading time measures which we explored were first pass duration (the sum of all fixations in a region until a saccade out of the region) and go-past duration (the sum of all temporally contiguous fixations in a region, including regressive eye-movements to the left of the region, until the point of fixation progresses to the region to the right).

Reading times (in ms) for the *subordinate verb region* are included in Table 1. (This table summarizes, for different regions, a number of different measures. For information, note that the table also includes first fixation duration; however, we do not include analyses for this early processing measure as generally it is not sensitive to syntactic manipulations (Joseph & Liversedge, 2013)). To explore eye movement behaviour at this first region, three LME models were run with each of the different reading time measures as the dependent variable. For both measures, there was a significant effect of age, with older children showing shorter reading times than the younger children (first pass: *M*s= 379 and 527, $t = 3.87$, $p < .0005$; go-past: *M*s= 556 and 772, $t = 3.14$, $p < .005$). Reading times were also significantly longer in this region for comma sentences than garden-path sentences for both measures (first pass: *M*s= 479 and 428, $t = 2.38$, $p < .05$; go-past: *M*s= 738 and 591, $t = 2.38$, $p < .05$). There was no interaction between age and sentence type for any measure and no other significant (or near significant) main effects or interactions. In summary, younger children showed longer reading times overall, but both groups of children showed equivalently longer reading times at the subordinate verb (*dressed*) when it was followed by a comma.

(ii) Do children show a garden-path effect?

Our hypothesis was that reading times should be longer at or following the disambiguating region in garden-path than comma sentences. Although effects are seen in adults on the disambiguating verb, Joseph and Liversedge (2013) found that children showed slightly delayed effects, emerging in the immediate post-disambiguating verb region. We examined measures of first pass duration and go-past duration[ii] noting that effects might emerge only in later measures. Potentially, older children might show effects on earlier measures and in earlier regions, leading to an interaction between sentence-type and age. Finally, following previous studies, we predicted that younger children would show longer reading times in all conditions (McConkie et al., 1991; Rayner, 1986). Reading times for each measure at the disambiguating verb-region and post disambiguating verb-region are included in Table 1. Four LME models were run, predicting each of the two reading time measures in each of the two regions. At the *disambiguating verb region*, there were significant main effects of age for each measure, reflecting longer reading times for the younger vs. older children (first pass: *M*s = 470 and 367, t =3.21, $p < .005$; go-past: *M*s = 680 and 502 , t =2.87, $p < .01$). Though reading times were longer in garden-path sentences than in comma sentences, the differences were not reliable for either measure (first pass: *M*s = 425 and 414, t =0.36; go-past: *M*s = 616 and 570, t =0.64), and no other significant (or near significant) main effects or interactions[iii].

At the *immediate post-disambiguating verb region*, for first pass reading times, the older children showed significantly faster reading times than the younger children (*M*s=409 and 520, t =3.52, $p < .001$). However there was no main effect of sentence type (comma = 462 and garden-path=469, t =0.15), and no interaction between age and sentence type was evident (t =1.6, p =.10). Turning to go-past duration on the post-disambiguating verb region[iv], here we saw a main effect of sentence-type, with significantly longer reading times for garden-path sentences than comma sentences (*M* =1114 vs. 760, t =4.06, $p < .0005$). Neither the main effect of age (t =0.36), nor its interaction with sentence-type (t =0.79) were significant; as before, no other main effect or interaction was significant or near significant. For both groups of children, longer go-past durations for the garden-path vs. comma sentences is clear evidence of a garden-path effect. To examine this further, we looked to see if children were more likely to make a regression out of this region in the garden-path condition. Relevant data are shown in Table 1, where we can see that children made 9% more regressions out of the post-disambiguating verb region in garden-path sentences than in comma sentences. Means are again shown in Table 1. There was a main effect of sentence type reflecting a greater likelihood of making a regression in garden-path sentences than comma sentences (*M* = 33% vs. 24%, z =2.1, p <.05; a simple test of the 2x2 contingency was also significant, χ^2=5.61, p <.05). There were no other significant or marginal main effects or interactions.

In summary, older children showed faster reading times than younger children overall. In line with other findings (Joseph & Liversedge, 2013), children were delayed (relative to adults) in their detection of and recovery from syntactic misanalysis with a garden-path effect not being evident on the disambiguating verb itself, nor in the earlier processing measure of first pass reading times. But, a garden-path effect was clearly evident in the post-disambiguating verb region for go-past durations. Consistent with this, children also made more regressions out of this region in the garden-path condition.

(iii) Do children look back to informative regions following disruption to processing in the garden-path condition?

The analyses reported above demonstrate that children spent longer reading when they encountered evidence to suggest that their initial analysis of the sentence was incorrect. One interpretation of this is that they are attempting reanalysis at this point. If correct, we might see more regressions back into those specific regions of the garden-path sentence which require reanalysis (e.g., Frazier & Rayner, 1982; Inhoff & Weger, 2005; Mitchell et al., 2008). The disambiguating information in our garden-path sentences indicates that, for both verb types, the second noun phrase needs to have its thematic role reassigned: in *While Anna was dressing the baby played in its crib,* the *baby* needs to be reanalysed so that it becomes the subject of *played* rather than the direct object of *dressing;* in *While Jim was eating the biscuits baked the oven*, *biscuits* need to be reanalysed as the subject of *baking* and not the direct object of *eating*. Thus we predicted relatively more regressions into the second NP in garden-path sentences. However the nature of OT vs. RAT verbs led us to different predictions for the first noun phrase. For RAT-verb sentences such as *While Anna dressed the baby played in its crib*, Anna is not only the subject but also the object of dressing. In contrast, in the OT sentence, Jim remains the subject of *eating*, with an implied but non-specified object. Thus, if regressions are targeted at regions for reanalysis, we predicted that we would see more regressions back to the first noun phrase in garden-path sentences than comma sentences for RAT verbs, but not for OT verbs.

Table 1 shows the number of regressions into each of these regions as a function of sentence-type, verb-type and age. Logistic LME models predicting whether a regression was made to the second NP or the first NP generated similar results, with main effects of sentence-type, reflecting more regressions in garden-path (M_{NP2} = 40%; M_{NP1} = 30%) sentences than comma (M_{NP2} = 27%; M_{NP1} =22%) sentences (NP2: z =4.1, p <.0001; NP1: z = 2.8, p <.01) and verb-type, reflecting more regressions for OT (M_{NP2} = 40%; M_{NP1} = 29%) verbs than RAT verbs (M_{NP2} = 28%; M_{NP1} = 23%), NP2: z =2.7, p <.01; NP1: z =2.2, p <.05. However, there was no support for the prediction that there would be more regressions into NP1 in the garden-path condition than comma condition specifically for RAT verbs (M_{GP} = 25%, M_{COM} = 20%) and not OT verbs (M_{GP} = 35%, M_{COM} = 23%), with no interaction between sentence type and verb type: z =0.01.

In summary, children made more regressions back to both the first and the second noun phrases in garden-path sentences than comma sentences. However, contrary to our specific prediction about the first noun phrase (i.e., the subject of the subordinate verb), the increase in number of regressions was not limited to sentences containing a RAT verb: an equivalent number were made in sentences containing an OT verb, even though the noun did not need to be reanalysed in those sentences. This suggests that although children made regressions in the garden-path condition, these were not targeted specifically to those regions requiring re-analysis. That children made more regressions to both noun phrases in sentences containing OT verbs presumably reflects the greater difficulty children had with those sentences, and is consistent with their poorer performance on the comprehension questions for OT verbs across conditions.

3. Relating children's eye movement patterns with their performance on the comprehension questions

Two aspects of our data are consistent with children having detected an initial misanalysis when reading garden-path sentences: longer reading times in the post-disambiguating verb region for garden-path vs. comma sentences, and more regressions in the garden-path condition than comma condition. Plausibly, these aspects of the eye movement record might also reflect children attempting to reanalyse the ambiguity. If correct, there should be an item-level association between patterns of eye movements and whether or not the comprehension question was answered correctly. The data in the upper part of Table 2 are relevant to addressing this prediction. Shown here are the go-past reading durations for the post-disambiguating verb region of sentences, as well as the proportion of trials where there was a regression out of this region, both as a function of whether or not the subsequent question was answered correctly. We also include the total reading time for the whole sentence for correct vs. incorrect answers, in order to provide a less-focused test of our general prediction.

Starting with the reading time measures, if longer reading times at the post-disambiguating verb region are indicative of reanalysis, longer times should be associated with better comprehension, especially in the garden-path condition. Numerically, we can see some support for this prediction in the garden-path condition: the mean go-past duration at the post disambiguating verb region for correct sentences was 1140ms, compared with 1095ms for incorrect sentences, with the effect actually reversed in the comma sentences with 679ms for correct sentences compared with 872ms for incorrect sentences. This was explored statistically in a logistic LME model predicting comprehension accuracy (correct vs. incorrect) from go-past duration in the post-disambiguating verb region and including the effect of sentence-type, verb-type and age-group, and all interactions.[v] Of interest was whether go-past reading time in the post-disambiguating verb region predicted comprehension performance and, most critically, whether there was an interaction with sentence-type. There was no main effect of go-past duration (z =-1.47, p =.14) and the overall means are in the opposite direction to our prediction (correct M =866, incorrect M =1001).The interaction with sentence-type was marginal, z =1.84, p

=.07). Since we specifically predicted a positive relationship between this measure and comprehension in the garden-path sentences but not the comma sentences, we ran a version of the model such that it estimated slopes for go-past duration in this region for comma and garden-path sentences separately. For comma sentences, there was a marginal effect, suggesting that longer reading times were associated incorrect answers (z =-1.85, p =.06, correct M =679, incorrect M =872); for the garden-path sentences, however, there was no relationship (z =0.42, p =.67, correct M = 1140, incorrect M = 1095). Thus there is no evidence that longer reading times in this region led to more accurate answers in the garden-path condition.

Arguably, a less-focused eye movement measure might show a closer association with comprehension performance. In fact, as is clear in Table 2, longer reading times for the entire sentence were associated with poorer comprehension for both types of sentence. A further LME model[vi] found a significant negative main effect of reading time (z =-3.43, p <.001: correct answers M = 4135, incorrect answers M =4920), indicating an overall relationship whereby longer reading times (i.e., slower reading) were associated with more comprehension errors. This did not interact with sentence type (z =0.65, p =.52). There was a significant interaction with verb-type (z =-2.35, p <.05). Breaking this down, there was a reliable disparity of reading times between correct and incorrect sentences for RAT sentences (correct M = 3924, incorrect M = 5020, z =3.7, p <0.0005) but not OT sentences (correct M = 4429, incorrect M = 4847, z =1.14, p =.25). One possibility is that comprehension following OT sentences is sufficiently hard that it masks the general relationship between slow reading and poor comprehension.

The overall relationship is likely to be indicative of a general relationship between reading fluency and reading comprehension, such that children who read more slowly are poorer readers, and therefore less able to answer the questions. To further test this possibility, we asked whether children's performance on an independent assessment of reading ability (taken from the TOWRE, a standardized test of word reading fluency) was related to (i) their sentence reading times in the experiment (ii) their comprehension scores in the comprehension test. We used LME's to predict these measures from raw scores on the TOWRE, along with the experimental factors used in previous models and all of the interactions[vii]. TOWRE reading skill predicted both experiment reading time, $t = -3.79$, $p < .0005$, and comprehension performance ($z = 4.5$, $p < .00001$). There were no significant interactions with any of the other factors (*t*s < 1.6).

Taken together, these analyses support the idea that the relationship seen in the experiment between slow sentence reading and lower levels of comprehension is due to the fact that poor readers are slow, and have less good comprehension. This also explains the finding that longer reading times in the post disambiguating verb region in comma sentences are associated with worse comprehension; this was not the case in the garden-path condition, presumably because longer reading times at this region in garden-path sentences may be caused by a separate phenomenon. Most importantly however, there is no evidence that longer reading times in disambiguating regions of garden-path sentences is specifically associated with *better* comprehension of those sentences, and thus no evidence to support the view that the increased time spent reading this region reflects time spent reanalysing the temporary ambiguity, at least to the level needed to support answering the comprehension question correctly.

As noted above, the other aspect of eye-movement behaviour that might be indicative of reanalysis and therefore might be associated with comprehension performance is the number of regressions made out of the disambiguating region. Inspection of the mean number of regressions out of this region as a function of correct vs. incorrect comprehension, shown in the lower part of Table 2, does not offer initial support to this prediction: overall, 29% of incorrectly answered trials showed a regression out vs. 28% of correctly answered trials. For the garden-path condition specifically, the results pattern in the opposite direction with 2% more regressions in incorrect vs. correctly answered trials. An LME model[viii] predicting comprehension accuracy (correct vs. incorrect) found that whether a regression was made from the post-disambiguating verb region was not a reliable predictor ($z = 0.11$, $p = .91$), and no interaction between number of regressions and sentence-type ($z = 0.16$, $p = .87$).

Since our earlier analyses examined regressions into specific regions (NP1 and NP2), we also addressed whether these regressions were associated with comprehension performance. Relevant data are summarised in the lower part of Table 2 and were analysed via two models predicting comprehension accuracy with the predictors of interest being (i) whether a regression was made back to the second noun phrase (ii) whether a regressions was made back to the first noun phrase[ix]. Regressions to NP2 was not a significant predictor of comprehension success (z =0.69); nor did it not interact with other factors. However, regressions back to NP1 was a significant predictor, with more regressions being associated with correct (M =30%) than incorrect (M =22%) performance (z =2.1, p =.04). The interaction with sentence-type was not significant (z =0.87, p =.39), nor was the interaction with verb type (z =1.53, p =.13). To further probe why regressions into this region might be associated with correct answering, we considered whether this also held for other regions of the sentence. The percentage of trials where a regression was made into each region in the sentence is shown in Table 3. From inspection, it appears that the relationship in fact holds across the first four regions of the sentence suggesting that in general, making a regression into an early part of a sentence is associated with comprehension success.

In summary, longer reading times were associated with poorer performance on the comprehension questions. This effect was general and held equally for garden-path vs. comma sentences. There was no evidence that making a regression out of the post-verbal region, or into the second noun phrase, predicted comprehension success. In contrast, the number of regressions into the first noun phrase was associated with comprehension accuracy. However, this needs to be modified by the observation that the effect was a general one, evident in comma as well as garden-path sentences; furthermore, inspection of the means suggests that the result is likely to be associated with making a regression into an early part of the sentence, rather than specifically targeted to the first noun phrase. Temporary ambiguity also lead to more regressive eye-movements out of the post disambiguating region, however there was no evidence from our data that these regressions were specifically associated with comprehension performance.

4. Reading times for questions and the ability to answer the question correctly

Our final set of analyses examined the relationship between time spent reading the comprehension question and accuracy of response, motivated by the idea that children engage in re-analysis of garden-path sentences when explicitly prompted by a question. If this was the case, correct answers should be associated with longer reading times on the question, as children actively reflect and reconstruct the sentence. An LME[x] predicting comprehension accuracy for garden-path and comma trials from question reading time (along with other experimental factors and interactions) found no main effect of question reading time ($z = 1.34$, $p = .18$). Importantly however, there was an interaction between sentence-type and question reading time ($z = 3.41$, $p < .001$). Breaking this down, in the garden-path condition, longer reading times were reliably associated with *correctly* answered questions (correct $M = 2805$, incorrect $M = 2559$; z =2.87, p <.005) whereas longer reading times were actually associated with *incorrect* answers in the comma condition (correct $M = 2641$, incorrect $M = 2988$; z = -1.56, p =.12). This was qualified by a significant three-way interaction between question reading time, sentence-type and age group (z =2.73, p < .01). The interaction between reading times and year-group was reliable for comma sentences (z =-1.97, p < .05), reflecting the fact that though both age groups had numerically longer reading times for incorrect answers, this was only reliable for older children (older $M_{incorrect} = 2681$, $M_{correct} = 2228$; z =2.12, p <.05; younger $M_{incorrect =}$ 3174, $M_{correct =}$ 3164; $z = .35$, $p = .73$). For garden path sentences, the interaction between age-group and reading time was marginal (z =1.76, p =.08), with both groups showing a numerical relationship between increased reading time and *correct* answers, but this was only reliable for the older group (older $M_{incorrect} = 1922$, $M_{correct} = 2390$, $z = 2.62$, $p = .009$; younger $M_{incorrect =}$ 3113, $M_{correct =}$ 3226; $z = 1.19$, $p = .23$).

In summary, the critical finding was that longer question reading times were associated with answering the comprehension questions incorrectly in the comma condition (reminiscent of the relationship between slower sentence reading time, poorer performance on the comprehension questions and lower reading skills as measured by the TOWRE, discussed earlier) but *correctly* in the garden-path condition. This was most clear for the older children. Recall that the questions were exactly the same across the two conditions, meaning that differences in reading time cannot be a consequence of differences in the difficulty of reading the question. Instead, we suggest that children spent longer reading the question in the garden-path condition as they were using time to re-visit and reanalyse the sentences from memory and that this was then associated with response accuracy.

General Discussion

By monitoring children's eye movements as they read sentences we obtained a detailed record of children's online processing of temporary syntactic ambiguities. Children showed the classic garden-path effect in online processing, with their reading being disrupted following disambiguation. Replicating Joseph and Liversedge (2013), children did not show a garden-path effect on the disambiguating verb itself, nor in the early processing measure of first pass duration; instead, an effect was clearly evident in the post-disambiguating verb region for go-past reading times and consistent with this, children also made more regressions out of this region in the garden-path condition. Taken together, these findings demonstrate that children are slower and less efficient than adults at processing sentences that contain syntactic ambiguities.

We also investigated whether regressions would be targeted to particular regions of the sentence. If children are monitoring their comprehension as they read, we predicted they would make more regressions back into the first noun phrase in sentences containing a reflexive verb (e.g., into *Anna* in the sentence *While Anna was dressing the baby played in its cot)* than into the first noun phrase in sentences with an optionally transitive verb (e.g., into *Jim* in the sentence *While Jim was eating the biscuits* *baked in the oven).* The rationale here was that when a verb is reflexive, the first noun phrase needs to have its role reanalysed; that is, since *the baby* is not the theme of the dressing action, *Anna* should now be assigned to that role. Although we saw more regressions in garden-path sentences than comma sentences, there was no evidence to suggest that this was more the case for RAT than OT verbs. Note that though this is a null effect, and should thus be treated with caution, the fact that the pattern of means actually reversed from our predictions suggests that it is not due to lack of power. This suggests that regressions were not especially targeted in the garden-path condition. This may reflect a developmental difference between children and skilled adult readers. Some previous research with adults has shown that following a syntactic misanalysis, adults selectively target their regressive saccades to a previously ambiguous region in order to repair their incorrect analysis (Frazier & Rayner, 1982; Meseguer et al., 2002). However, this is not always the case (Hyönä, Lorch Jr., & Kaakinen, 2002; Mitchell et al., 2008), and it is clear that the nature of regressions varies a good deal. For example, experiments that have examined scanpath patterns have described a great deal of variability in regression paths, with some people and reading conditions producing targeted regressions, whilst others showing a more global strategy and re-reading from the start of the sentence (von der Malsburg & Vasishth, 2011). There is also limited evidence that children can make targeted regressions following syntactic misanalysis (Joseph & Liversedge, 2013), although the conditions under which they do so are not clear. Further research specifically exploring factors that cause regressive eye movements and re-reading (both in adults and children) is required.

In addition to monitoring eye movements, we asked children comprehension questions that probed the syntactic ambiguity directly. It is clear that children were susceptible to making 'good enough' errors: they made more errors on comprehension questions that followed garden-path sentences than control sentences in which the clauses were reversed so as to avoid the ambiguity. This demonstrates that the initial incorrect parse of the garden-path sentence, in which the first verb was parsed as transitive with the following noun as its direct object, continued to influence comprehension. Discussion continues in the adult literature as to what causes good enough effects in processing. Slattery et al. (2013) demonstrated that the effect is not likely to be a consequence of the processor failing to construct a syntactic analysis that is complete and faithful to the input. Instead, good enough effects seem to be a form of semantic persistence (Sturt, 2007) in which initial syntactic structures decay over time and potentially exert influence on the form of semantic representation constructed; this in turn may cause interference, especially when comprehension is probed directly, as in the classic good enough paradigm adopted in our experiment. Whether semantic persistence is a consequence of factors within the sentence processing system (e.g., Sturt, 2007) or a more general property of a fallible memory system (e.g., Kaschak & Glenberg, 2004) remains to be seen. Two points are relevant and noteworthy from our data. First, good enough comprehension errors were evident regardless of whether sentences contained verbs that were optionally transitive (e.g., *eat*) or reflexive absolute transitive (e.g., *dress*). Overall however, sentences with OT verbs were harder to comprehend, even in the non-ambiguous control conditions. Presumably, this is because the theme of an OT verb remains unspecified and thus provides some support for an incorrect inference, even though we constructed our stimuli to make such an inference implausible. Nevertheless, the fact that garden-path sentences led to more comprehension errors than control sentences regardless of the nature of the verb demonstrated that children's good enough errors were not a consequence of reliance on general inference over structural information, a conclusion consistent with observations in the adult literature (Christianson et al., 2001; Slattery et al., 2013).

The second noteworthy finding relevant to questions about the bases of good enough effects concerns the relationship between online processing and offline comprehension, where longer reading times on the question predicted comprehension success, specifically for the garden-path sentences. This observation is consistent with the idea that full re-analysis might not happen until the reader is forced to examine specific aspects of the initial 'good enough' parse. In our experiment, the question served as the prompt for some form of re-analysis, resulting in an association between longer question reading times and comprehension success. Importantly, this relationship was observed *only* in the garden-path condition; in the control condition, faster reading times predicted comprehension success, despite the questions being identical (this finding is in line with the more general relationship between fast reading and effective comprehension, discussed below). This suggests that specifically where there has been a temporary ambiguity, additional processing time is needed to actively reconstruct information already read. One possibility is that this additional time reflects a process whereby children notice and resolve two conflicting parses (i.e. semantic persistence). Interestingly, the association between question reading time and comprehension success was tighter for the older children, suggesting that the ability to detect an inconsistency, and to devote resources to resolving it, may develop with reading skill and experience.

In contrast to the effect of question reading time, we found no evidence of specific associations between eye movement behaviour while reading the garden-path sentences and comprehension success. Although we saw inflated reading times in the post-disambiguating verb region in the garden-path condition, this was not predictive of comprehension success. In addition, while the number of regressions into the first noun phrase was associated with comprehension, this was equally true for comma vs. garden-path sentences. Thus, there was no evidence in the current data that the eye movements which occur when a reader encounters material which disambiguates a temporary ambiguity are actually indicative of re-analysis of that ambiguity. This is suggestive and may seem to support a view whereby good-enough comprehension effects do not result from processes that occur as the parse is constructed. However, as for any null result within a single experiment, caution is needed. An additional factor here is that is our power was low (12 experimental items, distributed across conditions). In combination with data loss, there is a possibility that we lacked sufficient power to see a relationship between disambiguating eye-movements and comprehension. In addition, any benefit of greater reading time post-disambiguation must work against the more general reversed relationship whereby longer reading times predicted *incorrect* responses, as seen in a number of our analyses. This relationship between fast reading and better comprehension is consistent with much literature (e.g., Perfetti, 2007) and with our own observation that individual differences in reading skill, as measured by the TOWRE, predicted total reading time and comprehension accuracy, with more fluent readers spending less time reading the sentences but nevertheless showing better comprehension. Relatedly, we also found that if a regression was made during sentence reading, comprehension was more likely to be successful. Plausibly, regressions indicate mindful and purposeful reading (Ehrlich, Remond & Tardieu, 1999; Murray & Kennedy, 1983; Schroeder, 2011) and on this view, it is not surprising that they are associated with comprehension success.

It is worth highlighting that this reversed relationship *wasn't* apparent was for (go-past) reading time at the post-disambiguating verb, where time for correct trials was numerically greater than for incorrect trials (although the difference was not reliable, $p > .5$). One possibility is that there *is* a benefit of longer reading times at this region but that it is cancelled out by the general relationship between slow reading and poor reading comprehension, exacerbated by a lack of power. Alternatively, there may be no benefit per se; however the fact that there is an additional reason for readers to slow down in this region (i.e. hesitation due to disambiguation) may disrupt the general relationship between slow reading and poor comprehension. It is not possible to pull these explanations apart using the current data; additional experiments are needed with increased statistical power (i.e., more items) as well as methods for controlling the children's reading speed and ability. However, it's worth noting that the finding of lack of relationship between online garden path effects and good enough comprehension is in line with the findings reported by Christianson and Luke (2011) in adults. The relationship between online and offline processing is clearly an area which warrants further exploration with both children and in adult readers.

The literature is surprisingly silent about children's sensitivity to commas and consequently we know very little about how children process commas as they read. By comparing 7-8 and 10-11 year-old children and examining the effect of commas on both comprehension accuracy and online reading, our data permit a number of conclusions to be drawn. Both age groups were worse at answering questions following a garden-path sentence than the same sentence with a comma, indicating that the comma had played a role in blocking the formation of a temporary ambiguity. However, there is reason to suspect this skill is not complete in the younger children. Their comprehension was worse in the comma condition than the reversed condition, suggesting that they are not able to use the comma to fully block the formation of a temporary ambiguity. Although this difference was small, younger children were reliably different from older children, who showed no differences in accuracy for comma vs. reversed sentences, suggesting that for this older group the knowledge is more secure. This developmental difference might be in part a consequence of the older children having been taught that commas may be used to divide clauses; the recommendation in the UK national curriculum is that this is taught in Year 4 (our children were drawn from Year 3 and Year 6 classrooms). The fact that the younger children showed a difference in comprehension for garden-path vs. comma sentences does demonstrate some sensitivity to the role of commas however; plausibly, this may have been abstracted from reading experience. Turning to online effects, the region containing the comma (the first verb) showed longer reading times than the equivalent region in garden-path sentences without a comma, and this held for both older and younger children, indicating that both groups were sensitive to its presence. Moreover, our finding of a temporary syntactic ambiguity effect in sentence processing when comparing garden-path and comma sentences is evidence that the comma played its expected disambiguating role, as children read the sentences.

A limitation of our experiment is that there were far fewer items than would generally be used in an experiment with adults (6 verbs of each type occurring across the sentence types). The fact that many of our results are in line with findings from previous literature (e.g. the pattern of offline data is in line with findings with adults; the relationship between slow reading and poor comprehension is in line with previous literature on reading development) makes us confident that our results are not just statistical overreaching. However, our findings must be interpreted cautiously, especially where null results are reported. Nevertheless, this study does represent an important first step, given the novelty of the questions explored: few studies have explored children's online reading behaviour, and there is a paucity of work mapping between online processing and offline comprehension not just in developmental studies, but also in studies of adults.

In conclusion, few studies have explored online processing in children's reading and ours is the first to explore links between patterns of reading behaviour as children read sentences and comprehension questions, and their subsequent success at answering those questions. Children showed classic garden-path effects in online processing and these lingered, influencing offline comprehension. Overall, faster reading times and more regressions were generally associated with comprehension success. In our data however, there was no evidence pointing to a relationship between longer reading times being associated with online disambiguation or increased offline comprehension success. In contrast, there *was* a relationship between the time spent reading the question and comprehension success, specifically for garden-path sentences. We argue that this reflects that the question served to prompt subsequent processing. One possibility is that this process reflects the resolution of a conflict between persisting semantic representations. An important question for future research will be the extent to which "good-enough" parsing effects result from insufficient initial parse, or semantic error in reconstructing the sentence from memory. The methodology adopted in our experiment affords a detailed examination of the reading process from which developmental and individual differences can be mapped.

Acknowledgements

We thank Rachel Loomes, Paul Forbes, Georgina Bremner and Nicholas Cooper for research assistance along with the staff and pupils at the Primary schools in Oxfordshire who participated in this experiment. The manuscript was prepared while Elizabeth Wonnacott was supported by the British Academy and Economic and Social Research Council, and Holly Joseph and Kate Nation were supported by the Economic and Social Research Council. Please address correspondence to Elizabeth Wonnacott (e.wonnacott@ucl.ac.uk) or Kate Nation (kate.nation@psy.ox.ac.uk).

Figure Caption

Figure :1 Proportion correct (SE) answers to comprehension questions for 8 year olds (left) and 10 year olds (right).

References

Baayen, R. H. (2008). *Analyzing linguistic data*. Cambridge, UK: Cambridge University Press.

Baines, L. (1996). From page to screen: When a novel is interpreted for film, what gets lost in the translation? *Journal of Adolescent & Adult Literacy*, *39*, 612-622.

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., Maechler, M., & Bolker, B. (2013). lme4: Linear Mixed Effects Models using S4 Classes. R package version 0.999999-2. Retrieved from http://cran.r-project.org/web/packages/lme4/index.html.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368–407.

Christianson, K., & Luke, S. G. (2011). Context strengthens initial misinterpretations of text. *Scientific Studies of Reading*, *15*, 136-166.

Ehrlich, M. F., Remond, M., & Tardieu, H. (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in metacognitive monitoring. *Reading and Writing*, *11*(1), 29-63.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-Enough Representations Language Comprehension. *Current Directions in Psychological Science*, *11*, 11–15.

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

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.

Hayes, D. P. (1988). Speaking and writing: Distinct patterns of word choice. *Journal of Memory and Language, 27*, 572-585.

Hill, R. L., & Murray, W. S. (2000). Commas and spaces: Effects of punctuation on eye movements and sentence parsing. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 565-589). Oxford: Elsevier.

Hirotani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, *54*, 425-443.

Hyönä, J., Lorch Jr., R. F., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, *94*, 44-55.

Inhoff, A. W., & Weger, U. W. (2005). Memory for word location during reading: Eye movements to previously read words are spatially selective but not precise. *Memory & Cognition*, *33*, 447-461.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434-446.

Joseph, H. S., & Liversedge, S. P. (2013). Children's and adults' on-line processing of syntactically ambiguous sentences during reading. *PLoS ONE, 8, e54141. doi:10.1371/journal.pone.0054141*

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, *133*, 450.

Kidd, E., & Bavin, E. L. (2007). Lexical and referential influences on on-line spoken language comprehension: A comparison of adults and primary-school-age children. *First Language*, *27*, 29-52.

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*, 222-234.

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

McConkie, G. W., Zola, D., Grimes, J., Kerr, P. W., Bryant, N. R., & Wolff, P. M. (1991). Children's eye movements during reading. In J. F. Stein (Ed.), *Vision and visual dyslexia* (pp. 251–262). London: Macmillan Press.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*, 283-312.

Meseguer, E., Carreiras, M., & Clifton, C. (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition*, *30*, 551-561.

Mitchell, D. C., & Holmes, V. M. (1985). The role of specific information about the verb in parsing sentences with local structural ambiguity. *Journal of Memory and Language*, *24*, 542-559.

Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the Selective Reanalysis hypothesis. *Journal of Memory and Language*, *59*, 266-293.

Murray, W. S., & Kennedy, A. (1988). Spatial coding in the processing of anaphor by good and poor readers: Evidence from eye movement analyses. The Quarterly Journal of Experimental Psychology, 40(4), 693-718.

Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 280–285.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357-383.

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 940–961.

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.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*, 413-425.

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of experimental child psychology*, *41*(2), 211-236.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372-422

Rayner, K, Kambe G, Duffy SA. (2000) The effect of clause wrap-up on eye movements during reading. *Q J Exp Psychol* (4):1061-1080.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *The psychology of reading* (2nd ed.). New York: Psychology Press.

Schroeder, S. (2011). What readers have and do: Effects of students' verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology*, *103*, 877-896.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, 69, 104-120.

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.

Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, *30*, 267-295.

Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, *105*, 477-488.

Torgesen, J. K., Wagner, R., & Rashotte, C. (1999). *Test of Word Reading Efficiency (TOWRE)*. Austin, TX: Pro-Ed.

Traxler, M. J. (2002). Plausibility and subcategorization preference in children's processing of temporarily ambiguous sentences: Evidence from self-paced reading. *The Quarterly Journal of Experimental Psychology: Section A*, *55*, 75-96.

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. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In Clifton, C., Frazier, L. and Rayner, K. (Eds.), *Perspectives in sentence processing* (pp. 155-179). Hillsdale, NJ: Erlbaum.

Van Gompel, R. P. & Pickering, M. J. (2007). Syntactic parsing. In M.G. Gaskell (Ed.) *Oxford handbook of psycholinguistics,* (pp. 289-307). Oxford, England: Oxford University Press.

van Gompel, R. P., Pickering, M. J., Pearson, J., & Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language*, *55*, 335-362.

von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*, *65*, 109–127.

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

Footnotes

[i] To implement the evaluation of effects at the average of the three levels of the sentence-type factor for use with lme4, we first manually calculated two dummy-coded variables to implement the factor, and then centred those.

[ii] A potential problem with this measure in this design is that there is evidence that a comma may act like a period, meaning that there will be fewer regressions to the region before the comma for reasons other than the presence of ambiguity (Hirotani et al., 2006). One solution is to use an adjusted measure of go-past duration which only includes all reading time in the region in question (i.e., all fixations in the region before moving rightwards, but not regressions). We therefore conducted the equivalent analysis with this adjusted measure of go-past: the pattern of results remains the same, with significantly longer go-past durations in garden-path than comma sentences, only for the post-disambiguating verb region (t =2.04, p <.05); reading times were longer, but not significantly so, in the disambiguating verb region for garden-path vs. comma sentences. Thanks to Steven Frisson for discussion of this point.

[iii] A possible concern is that some of our items had relatively short critical verb regions which might increase the likelihood of not seeing the effect at that critical verb. However inspection of the means for go-past showed that the difference between garden path and comma sentences was numerically larger (rather than smaller) for items with shorter critical verbs: items with critical verb length >= 6, 42 ms difference; items with critical verb <6, 52ms difference.

[iv] Due to an oversight in stimulus creation, in one of our sentences ("While Betty was waking up(,) the neighbours slept soundly"), the post-disambiguating verb region was also the final region of the sentence. To check whether this item was driving this effect, we re-ran the analysis with it removed. The pattern of results remained identical, with a significant effect of sentence type (t =3.7, p <.01) only.

[v] Random slopes included in the model were: by participant slopes for go-past reading time and go-past reading time by sentence-type. In a non-converging version of the model with maximal random slopes the (negative) main effect of go-past reading time was significant (z =-2.07, p =.04) but the interaction with sentence-type was not (z =1.05, p =.3).

[vi] Random slopes included were: a by participant slope for reading time. In a non-converging version of the model with maximal random slopes the pattern of results was identical to those reported except that the interaction of reading time by verb-type was marginal (t =-1.95, p =.05) and the simple effect of reading time for OT verbs was marginal (z =-1.7, p =.08).

[vii] Random slopes included in the model predicting reading times were: by participant slopes for sentence-type and sentence-type by verb-type; by item slopes for TOWRE score and year-group. Random slopes included in the model predicting comprehension were: by-participant slopes for verb-type and sentence-type and by-item slopes for TOWRE, year-group, sentence-type and sentence-type by year-group. Non-converging models with full random slopes structures showed an identical pattern of results except that in the model predicting comprehension there was a marginal interaction between TOWRE score and verb-type (t =1.83, p =.07)

[viii] Random slopes included in the model were: by participant slopes for regression and the interaction between regression and sentence-type; by item slopes for regression, the interaction between regression and sentence-type and the interaction between regression and year-group. The non-converging models with full random slopes structures showed an identical pattern of results

[ix] Random slopes included in the model with regressions into the second NP as a predictor were: by-participant slopes for regressions, regressions by sentence-type, sentence-type and verb-type; by item slopes for regressions and regressions by sentence-type. Random slopes included in the model with regressions into the first NP as a predictor were: by participant slopes for regressions, regressions by sentence-type and sentence-type; by-item slopes for regressions, regressions by

sentence-type and sentence-type. Non-converging models with full random slopes structures showed an identical pattern of results to those reported in the text apart from a reliable interaction between making a regression into the first noun phrase and verb-type (z =2.16, p=.03) which was NS in the converging model.

[x] Random slopes included in the model were: by participant slopes for question reading time and the interaction between question reading time and sentence type. A non-converging model with full random slopes structures showed an identical pattern of results to those reported in the text except that the interaction between question reading time and age was not reliable for comma sentences.