

Children's processing and comprehension of complex sentences containing temporal connectives: The influence of memory on the time course of accurate responses.

First Submission date: 1st December, 2015

First revision: 9th May, 2016

Second revision: 22nd June, 2016

Abstract

In a touch screen paradigm, we recorded 3- to 7-year-olds' ($N = 108$) accuracy and response times to assess their comprehension of two-clause sentences containing *before* and *after*. Children were influenced by order: performance was most accurate when the presentation order of the two clauses matched the chronological order of events: '*She drank the juice, before she walked in the park*' (chronological order) vs '*Before she walked in the park, she drank the juice*' (reverse order). Differences in response times for correct responses varied by sentence type: accurate responses were made more speedily for sentences that afforded an incremental processing of meaning. An independent measure of memory predicted this pattern of performance. We discuss these findings in relation to children's knowledge of connective meaning and the processing requirements of sentences containing temporal connectives.

Keywords: temporal connectives, incremental processing, memory, language acquisition, response times.

Children's processing and comprehension of complex sentences containing temporal connectives: The influence of memory on the time course of accurate responses.

Successful comprehenders form a coherent mental representation of the events described in spoken or written text (Johnson-Laird, 1983; Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). The construction of a coherent mental representation is guided by the presence and understanding of connectives, which aid the integration of clauses by signalling how events link together (Bestgen & Costermans, 1997; Cozijn, Noordman, & Vonk, 2011). In this paper, we focus on children's processing of sentences containing the temporal connectives *before* and *after*, which encode the relation between events on a temporal dimension (Cain & Nash, 2011; Gennari, 2004). Whilst *before* and *after* appear regularly in speech from as young as 3 years of age (Diessel, 2004), 12-year-olds demonstrate difficulties in comprehending these connectives in specific sentence structures (Pyykkönen & Järvikivi, 2012). In the current study, we investigate the influence of memory and language on 3- to 7-year-old's comprehension of complex sentences containing temporal connectives by investigating the influence of these skills on the accuracy and speed of responses using a touch-screen comprehension task.

Our mental representation of event order corresponds to the chronological order in which the events occur in real world situations: the first occurring event is followed by the second, and so forth (Coll-Florit & Gennari, 2011; Givón, 1991; Zwaan & Radvansky, 1998). However the order in which events are described does not necessarily map onto actual order. Temporal connectives allow us to describe the events in both a chronological order, such as '*She played in the park, before she drank the juice*', or in a reverse order '*Before she drank the juice, she played in the park.*' Therefore, reverse order sentences violate the default expectation that newly encountered information follows the most recent event in the existing representation (Zwaan & Radvansky, 1998). This has implications for developmental

differences in the comprehension and processing of these sentences: children are more accurate at comprehending sentences which describe events in a chronological order, compared to sentences which describe events in a reverse order (Clark, 1971).

Previous work has provided two developmental reasons for difficulty with reverse order sentences. First, young children may have a fragile understanding for the meaning of the connective. If so, they will be more likely to use a non-linguistic strategy to represent the sequence of events based on the assumption that language order maps onto real world order, rather than using the linguistic information provided by the connective to guide the construction of their mental representation. In line with this, several studies have reported that young children who display a poor knowledge of *before* and *after* comprehend reverse order sentences at below-chance accuracy, but are significantly more accurate on chronological sentences (Clark, 1971). The second reason is based on previous adult studies which show that, even when knowledge of temporal connectives is robust, reverse order sentences are still more difficult to process than chronological sentences. This difficulty is attributed to the greater processing costs required to create a chronological mental representation from events that are described in a reverse order relative to when events are described in their actual chronological order (Münste, Schiltz, & Kutas, 1998; Ye et al., 2012). For children, the differences in comprehension of chronological vs reverse order sentences are modulated by the development of memory and vocabulary (Blything, Davies, & Cain, 2015). This set of previous findings motivated the current study to contrast memory capacity-constrained (e.g., Just & Carpenter, 1992) and language-based (e.g., Van Dyke, Johns, & Kukona, 2014) accounts in relation to children's processing and comprehension of sentences containing temporal connectives.

A memory capacity-constrained framework (e.g., Just & Carpenter, 1992) attributes the difficulties for reverse order sentences to the requirement to hold more information active

in working memory, and to the available memory capacity of the individual. Children and adults process complex sentences incrementally, word by word and clause by clause (e.g., Cain & Nash, 2011; Traxler, Bybee & Pickering, 1997). As a result, reverse order sentences such as ‘*Before she drank the juice, she played in the park*’, are more difficult to process than chronological order sentences, because comprehenders do not process the first occurring event (*played in the park*) until part way through the sentence. As a result, they must then revise their mental representation. Conversely, a chronological order sentence such as ‘*She played in the park, before she drank the juice*’, allows incremental construction of the mental representation. Due to the memory demands associated with reverse order sentences, the memory capacity-constrained account would predict that individuals with low memory capacity would experience comprehension difficulties specifically for these constructions. Support for the memory capacity-constrained explanation comes from studies of both adults and children, with the difficulty for reverse order sentences being more pronounced in those who score low on a working memory capacity test (Blything et al., 2015; Münte et al., 1998).

In addition, even when children’s understanding for the connectives is robust enough to no longer rely on a non-linguistic strategy to understand and represent order (Clark, 1971), the connective might influence processing of these two-clause sentences because it varies the demands on working memory resources. Young children have poorer knowledge of *after* as a connective compared to *before* because it has more complex semantics (Clark, 1971), and is used in ways other than as a connective (e.g., *She is only after your money*, see Leech, Rayson, & Wilson, 2001). Words that are typically more difficult to learn - as reflected by a late age of acquisition, a low frequency of occurrence, or a high ambiguity in meaning - are processed by adults more slowly and less accurately than their less difficult to learn counterparts (Carroll & White, 1973; Juhasz, 2005). Most important for the predictions of the memory capacity-constrained account, these processing costs are more pronounced in

comprehenders with low working memory span compared to comprehenders with high working memory span (Gunter et al., 2003). Therefore, due to the complexity of *after*, sentences containing this connective may be more difficult to process because it is more taxing on working memory resources when activating knowledge of *after* as a temporal connective compared to *before*. Specifically, the influence of the connective on sentence processing should be driven by working memory capacity.

Also, the position of the connective in the sentence may influence the amount of information that must be held active in working memory. By manipulating both order and connective, the position of the connective varies across sentences. For example, *before* occurs in a medial sentence position when events are spoken in a chronological order, but in an initial sentence position when events are spoken in reverse order. The reverse is true for *after* sentences. Position of the connective has also been hypothesised to influence the amount of information held active in working memory. A medial position provides the information of the connective roughly when it is required to link the two meanings of the two adjacent clauses. Conversely, when the connective is provided at the beginning of the sentence, individuals must maintain the meaning of the connective while processing the first clause, and then link the clauses together (Diessel, 2004). In support of the proposal that the connective and its sentence position influence processing, Blything et al. (2015) reported that 4- to 6-year-olds displayed an advantage for chronological order sentences only when the sentence structure did not include these extra features which may increase demands on working memory resources. That finding was modulated by individual memory span, further supporting a memory capacity-constrained account (e.g., Just & Carpenter, 1992).

An alternative hypothesis for how memory influences the processing of these complex sentences is that the effect is actually driven by the quality of language knowledge rather than by the quantity of information that can be maintained within working memory

(e.g., Kidd, 2013; Klem, Melby-Lervåg, Hagtvet, Lyster, Gustafsson, & Hulme, 2015; Van Dyke et al., 2014). The language-based account draws on the notion that, rather than being separate systems (Baddeley, 2003; Baddeley & Hitch, 1974), working memory and long term memory are part of a unitary architecture in which working memory is a temporarily active portion of long term memory (Ericsson & Kintsch, 1995; McElree, 2006). Therefore, the current processing capacity of working memory is determined by the extent to which processing resources are devoted to the retrieval of target concepts from long term memory. That is, the ability to represent information in working memory is modulated by language knowledge. Poor language knowledge is likely to result in a fragile memory representation because the understanding for the meaning of target concepts is less distinct and robust, so the retrieval process is more susceptible to competition from other related concepts. Conversely, rich language knowledge supports the construction of a memory based mental representation because individuals can quickly access and accurately retrieve the precise target concepts. This reduces the likelihood of interference from related concepts, and frees up resources for constructing and maintaining an accurate mental representation.

The language-based account of sentence processing contrasts the memory capacity-constrained account (Just & Carpenter, 1992), which views working memory as independent from language (e.g., Baddeley, 2003; Baddeley & Hitch, 1974). In support of a language-based account of sentence processing, recent research with adults has examined the specificity or distinctness of retrieval cues in the text, for example how well the meaning of the target connective is activated in relation to competing temporal connectives, and how well other words in the sentence are activated in relation to competing words with similar meanings. This work shows that such information, rather than the number of individual text elements that must be held active in memory, can account for why some sentences are more

difficult to process than others (Van Dyke et al., 2014; Van Dyke & Johns, 2012; Van Dyke & McElree, 2006).

Research to date has explained children's difficulty in processing reverse order sentences using the framework of the memory-capacity constrained account (Blything et al., 2015; Pyykkönen & Järvikivi, 2012). However, those studies have used tasks that measure only response accuracy, in which children as young as 6 to 7 years can perform at ceiling. These findings motivate the need for a more sensitive assessment of children's sentence processing to study developmental and individual differences in performance. Studies of adults, for whom response accuracy is at ceiling, have used EEG and fMRI to index real-time processing (Münte, Schiltz & Kutas, 1998; Ye et al., 2012). This work demonstrates differences in the effort required to process chronological and reverse order sentences. Such findings have been explained within a memory-based account: reverse order sentences place higher demands on working memory. However, those studies used stimuli in which the connective was presented only in the sentence initial position, such that connective (*before*, *after*) was confounded with event order. This work has not included a design that compares order effects in sentences linked by both *before* and *after*. Further, the only previous studies that have examined online processing of these sentences have not included children, so they do not speak to developmental improvements. A fully factorial design is particularly important in developmental studies because children display developmental differences in their understanding of *before* and *after* (Clark, 1971).

The current study was motivated by our review of previous research on children's and adult's processing of multiple clause sentences including temporal connectives, to examine the role of memory and language in children's comprehension of such sentences. We measured the speed of children's responses using a touch screen comprehension task (for use of this method with preschool children, see Friend, Schmitt, & Simpson, 2012; Möhring,

Newcombe, & Frick, 2014), in addition to response accuracy. Here, we provided strict training and practice instructions to encourage speeded responses. Slower responses can be interpreted as a reflection of processing difficulties, which relate to the extra time needed to construct and revise a mental representation (Cain & Nash, 2011; Just, Carpenter & Wooley, 1982; Pérez, Paolieri, Macizo, & Bajo, 2014; Zwaan & Radwansky, 1998).

In addition to studying both accuracy and the time taken to make a response, our study differs from previous developmental studies by the nature of the task instructions. Pyykkönen and Järvikivi (2012) asked 8- to 12-year-olds to read a sentence reporting two events and to then indicate which occurred first, or whether they occurred at the same time. Even the oldest children were not at ceiling. In a study of the comprehension of similar sentences by much younger children, 6- to 7-year-olds were close to ceiling (Blything et al., 2015). Procedural differences between these studies may explain the age differences in reported competence: Blything et al. (2015) minimised processing demands by using a simple forced-choice touch screen comprehension task in which children were asked to select which event happened first from two images of the actions that were narrated in the sentence. However, Blything et al.'s (2015) "what happened first" instruction may have artificially increased accurate responses for (more complex) reverse order sentences. When children hear a two-clause sentence, the most recently heard event will be more recently activated in the child's memory than the first mentioned event. If children are asked "what happened first", the most recent event maps onto the answer for reverse order sentences, but not chronological sentences. This could boost response accuracy for reverse order sentences. By asking which event happened last, we can investigate whether children display the same levels and patterns of accuracy as found in previous studies, with a different set of instructions, and in so doing assess the reproducibility of the main findings.

The current study

Children listened to a two-clause sentence containing *before* or *after*, with events narrated either in a chronological or reverse order. During the narration, an animation of the event in each clause was shown, separately, on a touch screen monitor. Children were then asked to touch the picture that represented which of the two events happened last. We did not explicitly manipulate the position of the connective but it varied by the nature of our two within-subject factors: order and connective. Therefore, like others (e.g., Pyykkönen & Järvikivi, 2012), we can also relate our findings to connective position in the sentence.

We first hypothesized that the reason for the youngest children's poor comprehension would be that they use a non-linguistic strategy to compensate for a fragile understanding of the connective (Clark, 1971). Evidence for this would come from above chance performance for chronological sentences, but not for reverse order sentences. For the older children, we predicted a different pattern of performance, because they were expected to have more robust knowledge of the specific meaning of the connectives. Specifically, we expected these children to perform above chance for all sentence types, reflecting their ability to accurately encode the connective. However, we predicted that their accuracy for reverse order sentences would be lower than that for chronological order sentences, because of the higher processing demands of this sentence type (Just & Carpenter, 1992; Van Dyke et al., 2014).

Our second and third hypotheses relate to two different accounts: whether memory (Just & Carpenter, 1992) or language knowledge (e.g., Van Dyke et al., 2014) best explains processing difficulties. As memory and language skills both typically improve within the age range of interest, we also predict that whichever skill best explains performance should also explain unique variance over and above the effects of age, thus accounting for developmental improvements. Our use of a timed response measure, in addition to accuracy, provides a

sensitive means to assess whether different sentence structures differ in processing ease, as has been found for adults (Münte et al., 1998; Ye et al., 2012).

If a memory capacity-constrained account (e.g., Just & Carpenter, 1992) best explains processing difficulties, children should be more accurate and faster to respond to sentences that place the least demands on working memory. This account predicts the best performance for sentences with a chronological order that are linked by *before* (medial position) because these permit incremental word by word processing. All other sentence combinations (before-reverse, after-chronological, and after-reverse) carry two features that increase the amount of information that must be held in working memory (reverse order, more difficult connective, initial position). Critically, this pattern of performance will be predicted by an independent measure of memory.

If a language-based account (e.g., Van Dyke et al., 2014) best explains processing difficulties, then language knowledge, as measured by performance across connective (*before*, *after*) and by an independent measure of vocabulary, should modulate how well children can process and comprehend sentence structures that require more computational effort. More specifically, we would expect slower and less accurate responses to reverse order sentences linked by *after*, and for the pattern of performance to be driven by our measure of vocabulary knowledge. Critically, the influence of these measures of language knowledge would be expected to override the effects of working memory that would be proposed by the memory capacity account (Just & Carpenter, 1992; as demonstrated by Van Dyke et al., 2014).

Note that the influence of connective knowledge that is proposed by a language-based account of sentence processing (Van Dyke et al., 2014) differs to that proposed by the first (non-linguistic strategy) hypothesis (Clark, 1971). The first hypothesis focuses on whether young children display *below*-chance accuracy for reverse order sentences: this would be a

result of using a non-linguistic strategy, which is in turn a result of not having a basic appreciation for the meaning of the connective. Conversely, the language-based account of sentence processing (Van Dyke et al., 2014) relates to when children perform *above-chance* at all sentence structures. Therefore, it focuses on the period that follows children's appreciation for the meaning of the connective, which is a later period of interest to the first hypothesis and relates to a more fine-grained understanding of the connective that can be used to contrast only the predictions of a memory capacity-constrained account (Just & Carpenter, 1992).

Method

Participants

The sample comprised 108 children aged 3 to 7 years from schools in socially mixed catchment areas of North West England. There were 27 3- to 4-year-olds (aged 3;7-4;6, 16 boys), 28 4- to 5-year-olds (aged 4;8-5;7, 15 boys), 27 5- to 6-year-olds (aged 5;8 to 6;6, 15 boys), and 26 6- to 7-year-olds (aged 6;7 to 7;8, 11 boys). Data collection took place between March and June 2015. Written parental consent was obtained for all children, and assent was obtained from all children prior to assessment sessions. All children were native English speakers with no reported language disabilities.

Materials and Procedure

All children completed assessments of connective comprehension, memory, and receptive vocabulary. The connectives task was administered over two separate sessions. Each session lasted no longer than fifteen minutes. One session included the vocabulary assessment, the other the memory assessment.

Connective comprehension task. Comprehension of *before* and *after* was measured using a touch-screen comprehension task. There were 32 sentences that reported events that

are arbitrarily related (e.g., *He put on the socks, before he ate the burger*) (see Blything et al., 2015). These thirty-two two-clause sequences were counterbalanced across four lists so that they each represented one of four sentence constructions that vary by order of mention of events (chronological or reverse) and connective type (*before, after*). The four sentence constructions are shown in Table 1.

Table 1

Sentence conditions

Before		After	
Chronological	Reverse	Chronological	Reverse
He put on the socks, before he ate the pie.	Before he ate the pie, he put on the socks.	After he put on the socks, he ate the pie.	He ate the pie, after he put on the socks.

We created animated cartoons using Anime Studio Pro 9.1 (Smith Micro Software, 2012). Each cartoon depicted the actor, action and object of the event represented by a clause (e.g., Tom putting on socks; Tom eating a pie). For each item, the animations were presented in a sequential order with the animation on the right hand side of the screen shown first, followed by the animation on the left hand side of the screen. The presentation of the two animations was counterbalanced by both order of appearance and side of presentation. First, the animations were presented to the children. A recorded instruction was then played over headphones (*‘Listen carefully and touch the thing Tom/Sue did last’*), followed by a narration of the sentence itself (e.g., *‘Tom/Sue put on the socks before he/she ate the pie’*). A response window was opened with a short beep and was closed by a blank screen once the child had responded.

Practice trial instructions emphasized the importance of making judgments based solely on the meaning of the narrated sentence, not the visual stimuli. These practice trials happened prior to both of the sessions, so that children would be more attentive to the purpose of the task and therefore remember these instructions more easily. One sample t-tests revealed no significant preference for order or side of presentation ($ps > .15$).

The experiment was run using the PsyScript 3.2.1 (Slavin, 2013) scripting environment on a Macintosh laptop connected to a touch-screen monitor. Items were presented in a random order and no experimental conditions were presented twice on a run at any point, preventing potential priming effects (e.g., Allen, Haywood, Rajendran, & Branigan, 2011). A response was recorded as correct when the child touched the event that was described as happening last. Response time (RT) was the time between the audio beep following the sentence narration and the child’s response.

Vocabulary. Our measure of receptive vocabulary was the British Picture Vocabulary Scales – III (Dunn, Dunn, Styles, & Sewell, 2009), in which children have to point to one of four pictures that best illustrates the meaning of a word spoken aloud by the researcher. Testing was discontinued when a specified number of errors had been made, as per the guidelines in the manual. Raw vocabulary scores demonstrated age-related improvements: 3- to 4-year-olds = 64.85 (7.99); 4- to 5-year-olds = 78.71 (7.34); 5- to 6-year-olds = 91.26 (6.74); 6- to 7-year-olds = 98.67 (8.56). All children had a standardised score above 85 and the mean scores (SD) indicate that each age group was performing at an age-appropriate level: 3- to 4-year-olds = 108.89 (7.44); 4- to 5-year-olds = 104.43 (8.36); 5- to 6-year-olds = 100.56 (5.62); and 6- to 7-year-olds = 98.38 (7.44).

Memory. Each child completed the digit span subtest from the Working Memory Battery for Children (Pickering & Gathercole, 2001) to assess memory. This is the most suitable assessment of memory for our age range, because 4-year-olds perform at floor on more complex measures of working memory (Gathercole, Pickering, Ambridge, & Wearing, 2004). In this task, children were asked to recall a string of digits in the same order that they were spoken by the experimenter. The easiest level comprises strings of two digits, and the number of items in the string is increased once three trials on level were answered correctly. Raw scores were used for the analysis. The raw memory scores [mean (SD)] demonstrated age-related improvements: 3- to 4-year-olds = 19.11 (3.23); 4- to 5-year-olds = 22.71 (3.14); 5- to 6-year-olds = 25.78 (3.99); 6- to 7-year-olds = 26.81 (3.74). In addition, the standardised scores of memory were within the normal range of 85-115 for each age group: 4- to 5-year-olds = 103.86 (11.00); 5- to 6-year-olds = 108.70 (14.32); and 6- to 7-year-olds = 106.73 (15.84). Standardised scores are not provided for 3- to 4-year-olds. The test-retest reliability reported in the manual for children aged 5-7 years is good: $r = .81$.

Design

A 4 x 2 x 2 mixed design was used. The between-subjects independent variable was age group (3-4, 4-5, 5-6, and 6-7 years) and the within-subjects variables were order (chronological, reverse order), and connective type (*before*, *after*). By manipulating order and connective, we also by nature varied the position of the connective (see Table 1). The dependent variables were accuracy and response times.

Results

We report the results for accuracy and response times separately. For each, a series of Generalised Linear Mixed-effects models (GLMMs; Baayen, Davidson, & Bates, 2008) were fitted to the data in the R statistics environment (R Core Team, 2012) using *glmer* (for the binomial accuracy dependent variable) and *lmer* (for the continuous response time dependent variable) from package *lme4* (Bates, Maechler & Bolker, 2014). This method is essentially an extension of logistic regression, such that it allows both subject and item effects to be simultaneously treated as random. In other words, a GLMM simultaneously controls for (error) variance that is unexpectedly caused by specific items and specific participants rather than by the fixed effects themselves.

The aim for each model was to have a maximal random effects structure: random intercepts for subjects and items, and random slopes where applicable to the design (Barr, Levy, Scheepers & Tilly, 2013). However, this process highlighted the problems associated with obtaining a maximum model that have been recently outlined by Bates, Kliegl, Vasishth, and Baayen (submitted). Specifically, the information in typical data (i.e., the number of observations per subject and per item) is not sufficient to support the complexity of maximum models. As a consequence of this, our most complex models failed to converge. Using the recommendations of Bates et al. (submitted), fixed and random effects were incrementally added to a minimal model, and were justified by using the likelihood ratio test (Pinheiro &

Bates, 2000) for comparing models. In addition, the models were pruned so that non-significant factors were removed.

Accuracy analysis

We removed ten children from the analysis: four who performed at ceiling across the four sentences (100%), five who were identified as outliers in by-age by-sentence box plots, and one who was identified as the single outlier in by-age box plots of our independent measure of memory. This did not alter the main findings. Therefore, we report the main effects and interactions of memory, vocabulary, age, order and connective on the accuracy of responses by 98 children.

An initial model (Table A.1, see appendix) was built that only examined the effects of age, order and connective. This showed no difference between accuracy for *before* and *after* sentences, and no interaction effects between variables (all $ps > .15$). Therefore, following recommendations to allow more complex models to be clearly interpretable and to be better supported by the data (see Bates et al., submitted), these non-significant effects were pruned. The pruning of non-significant factors did not alter the reported findings (Table A.2, see appendix), and together with the removal of data points, ensured a normal distribution of the data that, in turn, allowed convergence of the final reported model that incorporated the effects of memory and vocabulary (Table 2). Memory and vocabulary were strongly correlated ($r = .69$), so were both centred. The addition of memory ($\chi^2(2) = 7.23, p < .03$) and vocabulary ($\chi^2(2) = 7.23, p < .03$) both improved the fit of the pruned model (Table A.2 in Appendix).

The inferential statistics are presented in Table 2. The first column provides the parameter estimates (b) which can be interpreted the same way as a regression, such that each shows the change in the log odds accuracy of response associated with each fixed effect on the dependent variable. A positive value indicates that the effect will benefit accuracy

whereas a negative value indicates that the effect will hinder accuracy. The by-age group mean (SD) accuracy scores for each sentence type are shown in Figure 1. There was a significant and sizeable effect of order, because chronological sentences were comprehended more accurately than reverse order sentences. There was also a main effect of memory, because children with higher working memory scores were significantly more accurate on the sentence comprehension task. There were no significant interactions between the variables. The influence of memory was over and above age and vocabulary, which were both non-significant. This contrasts with the finding reported in the initial models that had not incorporated memory and vocabulary (Table A.1 and Table A.2, see appendix): These had reported a main effect of age, with each of the three older age groups performing significantly more accurately than the 3- to 4-year-olds. This indicates that the effects of age in those initial models served as a proxy for the role of memory.

We also investigated a possible trade-off between accuracy and reaction times. However, the fit of the final reported model (Table 2), was not improved when reaction times were added as a fixed effect covariate ($\chi^2(2) = 0.34, p < .84$) or as item-wise random intercepts ($\chi^2(1) = 0.83, p < .36$). Similarly, these additions did not significantly improve the fit of the models reported in the appendix (Table A.1 and Table A.2), all $ps > .90$.

Figure 1

Mean (SE) proportion correct for each experimental condition by 3- to 7-year-olds

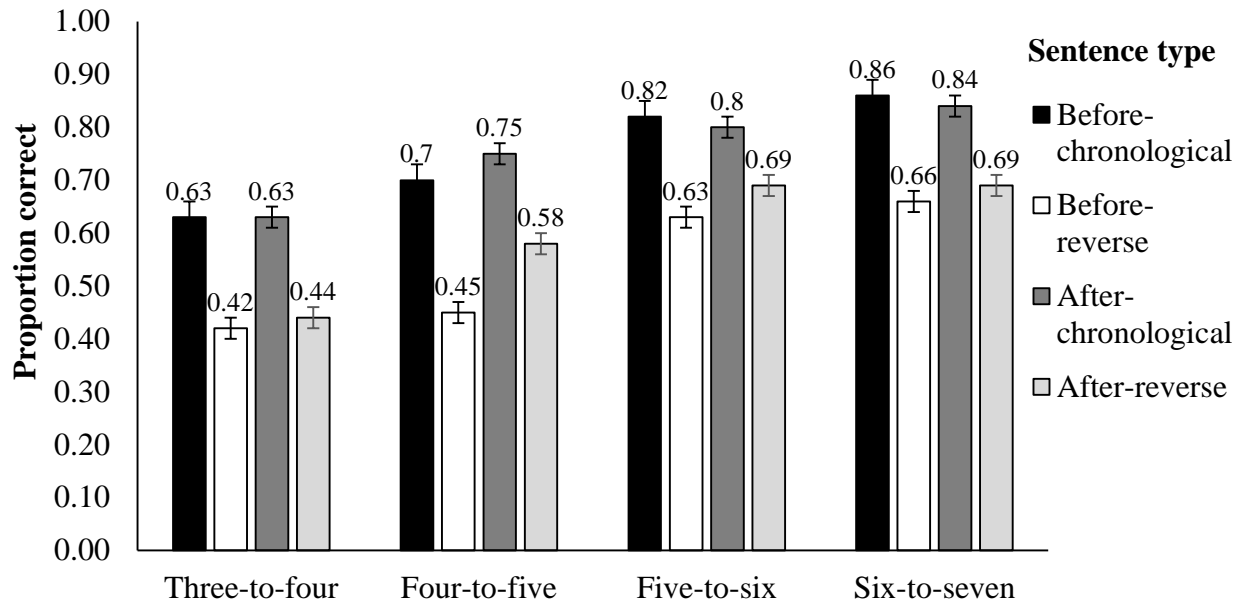


Table 2

Main effect and interactions of memory, vocabulary, age, and order on the proportion of correct answers by 3- to 7-year-olds.

Main model	M (b)	SE	z	CI		p(> z)
				2.5%	97.5%	
<i>Fixed effects</i>						
(Intercept)	0.20	0.22	0.90	-0.24	0.64	0.37
Memory	0.06	0.03	2.11	<0.01	0.11	0.04
Vocabulary	0.02	0.01	1.56	<0.01	0.04	0.12
Four-to-Five	0.02	0.23	0.09	-0.44	0.48	0.93
Five-to-Six	0.14	0.34	0.40	-0.53	0.80	0.69
Six-to-Seven	0.28	0.38	0.74	-0.46	1.02	0.46
Order	0.91	0.10	9.12	0.71	1.10	<0.01
Memory : Order	-0.03	0.03	-0.92	-0.09	0.03	0.36
Vocabulary : Order	0.01	0.01	0.57	-0.01	0.02	0.57
<i>Random effects</i>				Variance	SD	
Subject (intercept)				0.53	0.72	
Subject.1 (slope) order				0.29	0.54	

*Note: 1. Bold = predictor is significant at $p < .05$ or better.

2. Number of observations = 3136; groups = 98 subjects

We followed up the main effect of order with one-sample t-tests to examine whether each age-group performed above chance for chronological compared to reverse order sentences. Our youngest two age groups performed above chance for before-chronological sentences, [3- to 4-year olds: $t(26) = 2.93$, $p < .01$; 4- to 5-year olds: $t(27) = 4.21$, $p < .01$] and after-chronological sentences, [3- to 4-year olds: $t(26) = 2.82$, $p < .01$; 4- to 5-year olds: $t(27) = 5.82$, $p < .01$]. However, these children were not above chance level for before-reverse sentences [3- to 4-year olds: $t(26) = -1.60$, $p = .94$; 4- to 5-year olds: $t(27) = -0.85$, $p = .80$], or after-reverse sentences [3- to 4-year olds: $t(26) = -1.17$, $p = .87$; 4- to 5-year olds: $t(27) = -$

1.38, $p = .09$]. This pattern of performance indicates that their inaccuracy for reverse order sentences was likely a result of their fragile understanding for the meaning of *before* and *after*. Conversely, despite performing less accurately for reverse order compared to chronological sentences, our oldest two age-groups still performed above chance for before-reverse sentences [5- to 6-year-olds: $t(26) = 3.56, p < .01$; 6- to 7-year-olds: $t(27) = 3.20, p < .01$] and after-reverse sentences [5- to 6-year-olds: $t(26) = 2.88, p < .01$; 6- to 7-year-olds: $t(27) = 4.87, p < .01$]. This pattern of results indicates that the older children had a robust appreciation of the meanings of temporal connectives and understood both *before* and *after*. However, their performance was poorer when these connectives were used in sentences that expressed events in reverse order indicating that processing load may be a factor in children's connective comprehension.

Response time analysis

We did not include responses by 3- to 4-year-olds because their longer response times suggested that they were not able to follow the instruction to respond as quickly as possible. The 1816 correct responses by 4- to 7-year-olds were screened following recommendations from Baayen and Milin (2010) to remove potential distortions from the norm and improve the convergence of models. We first removed extreme response times that exceeded 2.5 standard deviations past the overall mean (49 responses over 9.5 seconds). Second, we removed remaining outliers that were more than 2.5 standard deviations above the mean response by subject (54 responses) and by item (42 further responses). Thus, a total of 8% of the original data points were removed as outliers. In addition, the data of one 6- to 7-year-old was removed because they were identified as an outlier in by-age box plots of our independent measure of memory. The mean (SD) response times in seconds by age-group were 1.75 (1.40) for 4- to 5-year-olds; 1.19 (1.17) for 5- to 6-year-olds; and 1.11 (1.27) for 6- to 7-year-olds. Mean response times for all correct responses in each experimental condition are

presented in Figure 2. Non-transformed means are reported for ease of interpretation. When 3- to- 4-year-olds were screened using this method, their response times were 2.96 (2.20) seconds, hence their exclusion.

A square root transformation was used for the inferential analysis so that the data were normally distributed. As in the accuracy analysis, an initial model was built which did not incorporate memory and vocabulary as covariates (Table A.4, see appendix). However, the response times model was not pruned, because age, order and connective each had either a significant main effect or were involved in an interaction. The same pattern of findings was found in a model of non-transformed response times (see Table A.3, Appendix), but our final model (Table 3) reports the square root transformation because the normal distribution reduced the stress on the model and, in turn, allowed the convergence of the additional effects of (centred) memory and (centred) vocabulary. In GLMMs of data with a continuous dependent variable, it is custom to present t-values and confidence intervals rather than p values because, for reasons beyond the current study, the statistical function `lmer` (from package `lme4`; Bates, Maechler & Bolker, 2012) does not provide p values. Reliably, a significant effect is indicated by a t-value exceeding 2, and when confidence intervals do not pass zero (Baayen, 2008).

Figure 2

Mean (SD) response times for each experimental condition by 4- to 7-year-olds

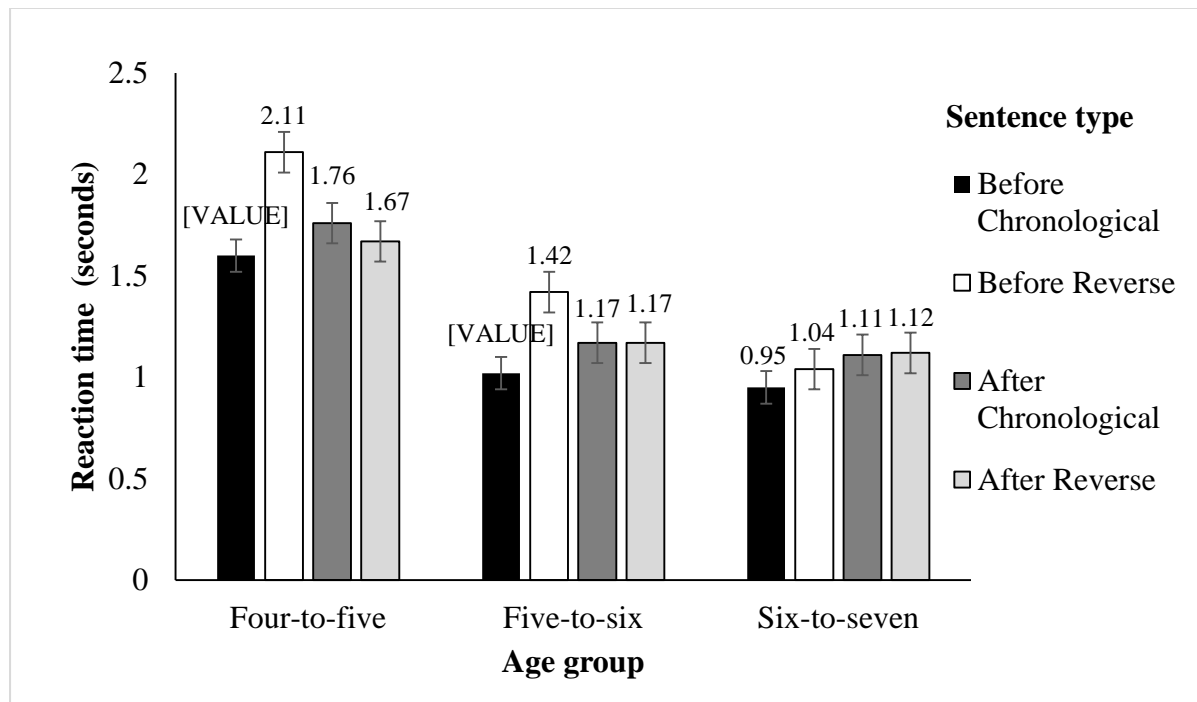


Table 3 summarises the main effects and interactions of memory, age, order and connective on response times. Similar to the accuracy analysis, there was no main effect of age once memory was added as a covariate, indicating that working memory was driving the developmental improvement in the processing of sentences overall. In contrast to the analysis of the accuracy data, there was a main effect of connective: Response times to sentences with *before* were faster than for sentences with *after*. Also in contrast to the analysis of accuracy data, the main effect of order was not significant: Response times to chronological sentences were not significantly different to those for reverse order sentences.

The main effect of connective was qualified by a three-way interaction between age, order and connective. The influence of age on the effects of order and connective indicates a developmental improvement in the processing of sentences. Therefore, the interaction was broken down by age. This is reported in Table 4 with by age-group models of the effect of order in a subset of each connective. The response times by 4- to 6-year-old's were

significantly influenced by an interaction between order and connective, whereas older children's response times were not. In the 4- to 6-year-olds, there was a main effect of order for *before* sentences, but not for *after* sentences. Specifically, before-chronological sentences were responded to significantly faster than before-reverse sentences, whereas response times to chronological and reverse order sentences containing *after* did not differ.

In line with the accuracy data, the addition of memory to the model significantly improved the fit of the data, $\chi^2(4) = 11.43, p = .02$. Children with higher memory capacity made faster (correct) responses overall. Most notably, there was a significant two-way interaction between memory and order, and also one between memory and connective. These interactions indicate that memory predicted the effects of both connective and order. Vocabulary did not improve the fit of the data, $\chi^2(4) = 6.53, p = .16$. Therefore, we do not report models of response times that incorporate vocabulary. This indicates that processing times were driven by memory capacity rather than vocabulary *per se*.

Table 3

Main effect and interactions of memory, age, order and connective on response times (with square root transformation) to correct answers by 4- to 7-year-olds.

Main model	M (b)	SE	t	CI	
				2.5%	97.5%
<i>Fixed effects:</i>					
(Intercept)	1.10	0.07	16.18	0.97	1.24
Memory	-0.03	0.01	-3.06	-0.05	-0.01
Five-to-Six	-0.06	0.09	-0.59	-0.24	0.13
Six-to-Seven	-0.13	0.10	-1.37	-0.32	0.06
Order	0.09	0.06	1.43	-0.03	0.21
Connective	0.28	0.07	3.86	0.14	0.42
Memory:Order	0.02	0.01	2.03	0.00	0.04
Five-to-Six:Order	-0.15	0.09	-1.75	-0.32	0.02
Six-to-Seven:Order	-0.13	0.09	-1.56	-0.30	0.03
Memory:Connective	0.02	0.01	2.49	0.01	0.04
Five-to-Six:Connective	-0.25	0.09	-2.62	-0.43	-0.06
Six-to-Seven:Connective	-0.33	0.10	-3.38	-0.52	-0.14
Order:Connective	-0.32	0.09	-3.49	-0.50	-0.14
Memory:Order:Connective	-0.02	0.01	-1.76	-0.05	<0.01
Five-to-Six:Order:Connective	0.22	0.12	1.75	-0.03	0.46
Six-to-Seven:Order:Connective	0.33	0.13	2.60	0.08	0.58
<i>Random effects:</i>			Variance	SD	
Subject (intercept)			0.06	0.24	
Item (intercept)			<0.01	0.07	
Residual			0.21	0.46	

*Note: 1. Bold = predictor is significant at $p < .05$ or better.

2. Number of observations = 1648 ; groups = 80 subjects and 64 items

Table 4

Simple effects age-group models of the effect of order by connective type on response times (with square root transformation) to correct answers.

	Age 4-5					Age 5-6					Age 6-7				
	(b)	SE	t	CI% 2.5	CI% 97.5	(b)	SE	t	CI% 2.5	CI% 97.5	(b)	SE	t	CI% 2.5	CI% 97.5
Before															
<i>Fixed:</i>															
(Intercept)	1.38	0.06	22.42	1.26	1.50	1.08	0.06	16.97	0.95	1.20	0.90	0.06	14.44	0.78	1.02
Order	-0.22	0.06	-3.35	-0.34	-0.09	-0.18	0.06	-3.21	-0.28	-0.07	-0.04	0.05	-0.67	-0.14	0.07
<i>Random:</i>				Var.	SD					Var.	SD				
Subject				0.02	0.18					0.06	0.24				
Item				0.03	0.12					0.00	0.06				
Residual				0.2	0.45					0.21	0.46				
After															
<i>Fixed:</i>															
(Intercept)	1.19	0.06	19.72	1.07	1.30	1.02	0.06	16.10	0.90	1.15	0.94	0.08	12.34	0.79	1.09
Order	0.04	0.06	0.61	-0.08	0.16	-0.04	0.05	-0.72	-0.14	0.06	-0.02	0.05	-0.43	-0.13	0.08
<i>Random:</i>				Var.	SD					Var.	SD				
Subject				0.04	0.19					0.06	0.25				
Item				<0.01	<0.01					<0.01	0.06				
Residual				0.25	0.5					0.18	0.42				

Notes. 1. Bold = predictor is significant at $p < .05$ or better. 2. Number of observations for ages 4-5 ‘before’ models = 230; groups = 28 subjects and 64 items. 3. Number of observations for ages 4-5 ‘after’ models = 267; groups = 28 subjects and 63 items. 4. Number of observations for ages 5-6 ‘before’ models = 292; groups = 27 subjects and 64 items. 5. Number of observations for ages 5-6 ‘after’ models = 267; groups = 27 subjects and 64 items. 6. Number of observations for ages 6-7 ‘before’ models = 282; groups = 25 subjects and 63 items. 7. Number of observations for ages 6-7 ‘after’ models = 290; groups = 25 subjects and 64 items.

Discussion

This study was designed to identify the reasons why children continue to experience difficulties in comprehending sentences containing *before* and *after* beyond the age that they have begun to display an early competence for these connectives. In general, there were developmental improvements in performance, such that sentences were understood more accurately and processed more quickly by older children. In relation to event order, children were less accurate at comprehending reverse order compared to chronological sentences. Our experimental manipulation of sentence type, together with independent measures of memory and language knowledge, enabled us to test between different theoretical accounts of children's difficulties with such sentences. The precise pattern of findings indicates different reasons for this effect in younger and older children. As discussed below, the evidence suggests that younger children's performance with reverse order sentences was limited because they displayed little or no understanding of the connective and instead relied on a non-linguistic strategy (Clark, 1971). In contrast, older children's overall performance indicated that they knew the meanings of the two connectives. A consideration of the pattern of performance and how this was related to individual differences in memory and language skills, suggests that older children's performance was limited by the processing demands of these sentences (Just & Carpenter, 1992; Van Dyke et al., 2014). We first examine the findings of the accuracy analysis and then turn to the analysis of response times, and discuss why variability in children's processing of these sentences is best explained by a memory capacity-constrained account (e.g., Just & Carpenter, 1992).

Our findings for response accuracy are convergent with the developmental findings reported by previous studies of children's comprehension of sentences with temporal connectives (Blything et al., 2015; Clark, 1971; Pyykkönen & Järvikivi, 2012). Children aged 3 to 5 years performed above chance on chronological sentences, but not for reverse order

sentences. This difference indicates that they did not take full advantage of the event order that is signalled by the connective, and compensated for this by defaulting to an expectation that language order maps onto the actual order of events (Clark, 1971). The 5- to- 7-year-olds performed above chance for all sentence types, which reflects an appreciation for the meaning of the connectives. However, they were in general poorer on reverse order sentences. Since older children displayed an appreciation for the meaning of the connectives, one reason for the lower accuracy for reverse order sentences is that these sentences have higher processing costs (Pyykkönen & Järvikivi, 2012).

Performance on the accuracy task was best explained by memory rather than chronological age or vocabulary. This finding provides partial support for the memory capacity-constrained account (Just & Carpenter, 1992). That is, performance was driven by whether children's memory capacity was sufficient to cope with the processing demands of our sentences in general. However, the account is only partially supported because the inaccurate comprehension of reverse order compared to chronological sentences did not interact with memory. We argue that the absence of this interaction could be attributed to the task requirement to provide speeded responses. When children are required to respond quickly, they have less time to reflect on and revise the representation that they have constructed and stored in memory (see Marinis, 2010). As a result, the ability to accurately store and manipulate the contents of memory may have a weaker influence on accuracy. Therefore, we turn to our response time measure, to better understand our pattern of data and the processing difficulties experienced by children with these sentence types.

Response times were analysed for only correct responses to determine if different connectives or structures differed in ease of processing. Thus, the pattern of data cannot be compared directly with the accuracy data. The response time analyses indicate that, even when sentences with temporal connectives are comprehended correctly, some are more

difficult to process than others (e.g., Cain & Nash, 2011; Ye et al., 2012). The response time data support the memory capacity-constrained account (Just & Carpenter, 1992). Children responded most quickly to chronological order sentences linked by *before* (medial position), which allow incremental word by word processing; and more slowly to before-reverse sentences, which do not afford incremental processing. There was no effect of order for sentences containing *after*. After-chronological sentences (initial position, later acquired connective) sentences and after-reverse sentences (reverse order, later acquired connective) each carry two features associated with taxing information to be held in working memory, and do not permit incremental processing. This may be the reason for the absence of response time differences between these two sentence types.

Importantly, the incorporation of memory significantly improved the fit of the model for response times, whereas vocabulary did not. Moreover, the main effect of age was no longer significant when memory was added to the model. Instead, the main effect of memory can account for developmental improvements in the processing of these sentences. This suggests that, as in the accuracy findings, age effects were partly a proxy for the influence of memory. Of particular note, the variation in response times across our sentence structures was predicted by our independent measure of memory span. This indicates that demands on working memory are driving these effects. That is, children with higher working memory spans are better able to cope with the higher memory demands of difficult sentences, and so experience fewer problems, as do in adults (Just & Carpenter, 1992).

In turn, the support we provide for a memory capacity-constrained account of sentence processing informs and maps onto our understanding of how the temporal information in these sentences is mentally represented (Gennari, 2004; Zwaan & Radvansky, 1998). We interpret the slower responses to sentences that do not afford incremental processing as a reflection of processing difficulties that relate to the extra time needed to

construct and revise a mental representation (Cain & Nash, 2011; Just, Carpenter & Wooley, 1982; Pérez, Paolieri, Macizo, & Bajo, 2014; Zwaan & Radwansky, 1998). Those sentences carry additional memory processing demands because more information must be maintained in working memory whilst the mental representation is revised. It follows that children who have lower working memory capacity will be less capable of revising the mental representation into the desired accurate linear order. This provides additional support to previous studies that have attributed children's inaccuracy with these sentence structures to a difficulty in mentally representing sentences that carry higher memory processing demands (Blything et al., 2015; Pyykkönen & Järvikivi, 2012).

Of course, we should not dismiss language effects *per se*. For example, the advantage for chronological sentences displayed by the younger children is a result of their *below-chance* accuracy for reverse order sentences. This suggests that when children do not have an appreciation for the meaning of a temporal connective, they will use a non-linguistic strategy to understand and represent the relation between two events (Clark, 1971). However, these findings are not relevant to the language-based account of processing (Van Dyke et al., 2014), which focuses on a more fine-grained understanding of the connective (and the other words in the sentence) in the immediate years that follow an appreciation for its meaning. We did report an effect of language knowledge on processing: *before* sentences were responded to faster than *after* sentences. However, children with a higher working memory capacity were less likely to display such effects. Therefore, these connective effects are interpreted in line with a memory capacity-constrained framework (Just & Carpenter, 1992), such that sentences linked by the more complex connective *after* carry additional demands on working memory compared to sentences linked by *before* (Clark, 1971; Leech et al., 2001). This fits the prediction that chronological sentences linked by *before* are processed most easily because it is the only sentence structure that does not carry any additional features that increase the

amount of information to be held in working memory (easier connective, chronological order, medial position).

A strength of our design was the manipulation of both memory and language processing requirements of our stimuli, in addition to the use of independent measures of memory and language to relate to performance. It is worth noting that language research is becoming increasingly aware of the need use an intensive battery of measures for individual differences in skills such as memory and vocabulary (Language and Reading Research Consortium, 2015). We selected a single measure of short-term memory with a low semantic load to better disentangle the effects of memory and language, noting that memory measures with greater semantic content are more strongly related to language processing ability in young children than digit based tasks (Cain, 2006; Seigneuric, Ehrlich, Oakhill, & Yuill, 2000). Because of our age range, we were not able to use a measure of complex memory span (Gathercole, et al., 2004) and note that such a measure may be more strongly related to language processing than our short-term memory measure (Daneman & Merikle, 1996). Similarly, we measured only the breadth of vocabulary (i.e., number of words known or not known), a measure used frequently with our age cohort (e.g., Silva & Cain, 2015). However, depth of vocabulary knowledge (i.e., the richness of knowledge for a particular word) is also highly predictive of comprehension ability (Cain & Oakhill, 2014; Ouellette, 2006). Therefore, future work should explore the sensitivity and inclusion of more complex measures of memory and vocabulary when assessing the relation between these skills and language processing to provide a more accurate assessment of these constructs to relate to sentence comprehension.

It is also worth noting that the accuracy findings inform us of the importance of the nature of the task itself. Children were less accurate overall relative to previous studies of the same age group (e.g., Blything et al., 2015). This is most likely a result of the requirement for

children to produce speeded responses. However, relative to previous studies, children also displayed lower accuracy for reverse order sentences. That poor performance cannot be attributed to the speeded instructions alone, because accuracy for chronological sentences was equivalent to previous studies.¹ In line with our predictions, we attribute this difference to the use of the “what happened last” question. Therefore the current study suggests that, in forced-choice paradigms for these sentences, accuracy may be distorted by false positive answers whereby children are more likely to choose the target answer because it maps onto the event that had been most recently activated in memory. This highlights the motivation of the current study to inform existing accuracy data with a measure of processing ease (response times) in addition to accuracy.

This is the first study to report a measure that indicates how efficiently children process two-clause sentences containing *before* and *after*. That is, it takes the first step to supporting previous forced-choice accuracy studies that have attributed children’s inaccurate comprehension to a difficulty in representing sentences that do not afford incremental processing (Blything et al., 2015; Pyykkönen & Järvikivi, 2012). The specific measure was chosen because the paradigm was analogous to the touch screen comprehension task used by Blything et al. The average response times were well within the range of those that have been previously reported by other touch screen paradigms as a reflection of children’s mental representations (Möhring et al., 2014); and previous studies have also interpreted response times to comprehension accuracy tasks as a reflection of the time needed to construct and revise a mental representation (e.g., Cain & Nash, 2011; Pérez et al., 2014; Zwaan & Radwansky, 1998). However, in order to gain a full picture of how children process these sentences, further research must assess real time moment by moment processing in sentence

¹ Older children did display a slight reduction in accuracy for before-chronological sentences relative to the ceiling level achieved in a recent study by Blything et al, but these reductions are likely a result of ceiling performance being a less realistic opportunity with speeded responses.

comprehension (and production). For example, the reason that our memory measures were less likely to influence response times in children with increasing age, may be that, at their more advanced developmental stage, they are more capable of revising the mental representation during sentence presentation. A paradigm that included measurement of ERPs might usefully indicate where the cognitive demands were greatest and whether processing effort for particular sentence regions are more strongly related to independent measures of memory, as has been shown with adults (Münte et al., 1998).

Overall, our analyses demonstrate age-related differences in 3- to 7-year-old's understanding of temporal connectives (e.g., Clark, 1971). The pattern of findings supports the conclusion that the 3- to 5-year olds were inaccurate because they had a poor appreciation for the meaning of the connectives, so could not appropriately use the linguistic information about temporal order. The 5- to 7-year-olds demonstrated a robust understanding of the connective but displayed evidence of processing difficulties. Our critical processing time measure provided evidence that the processing difficulty can be attributed to the memory load of the sentence structure and to the available memory resources of the individual (Just & Carpenter, 1992). Finally, we emphasise the need for future studies to test the generalisation of this conclusion with different independent measures of memory, more comprehension assessments of vocabulary knowledge, and online paradigms that provide an indicator of processing efficiency during the comprehension of the sentence itself.

Acknowledgements

This work was supported by a studentship from the Economic and Social Research Council awarded to the first author. For further details on materials or analysis methods please contact the corresponding author. The full dataset can be accessed online:

[10.17635/lancaster/researchdata/87](https://doi.org/10.17635/lancaster/researchdata/87)

References

- Allen, M. L., Haywood, S., Rajendran, G., & Branigan, H. (2011). Evidence for syntactic alignment in children with autism. *Developmental Science, 14*, 540-548. doi: 10.1111/j.1467-7687.2010.01001.x
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412. doi:10.1016/j.jml.2007.12.005
- Baayen, R. H. (2008). *Analyzing linguistic data*. Cambridge University Press.
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*, 12-28.
<http://www.redalyc.org/articulo.oa?id=299023506003>
- Baddeley, A. D. (2003). Working Memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829-839. doi:10.1038/nrn1201
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation Vol. 8* (pp. 47–89). New York: Academic Press.
- Barr, D. J., Levy, R., Scheepers, C., & Tilly, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bates, D. M., Maechler, M., & Bolker, B. (2014). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999999–0. <http://CRAN.R-project.org/package=lme4>.
- Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, H. (submitted to *Journal of Memory and Language*). Parsimonious Mixed Models. doi: arXiv:1506.04967

- Bestgen, Y., & Costermans, J. (1997). Temporal markers of narrative structure: Studies in production. In J. Costermans, & M. Fayol (Eds.), *Processing interclausal relationships: Studies in the production and comprehension of text* (pp. 201-218). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Blything, L. P., Davies, R., & Cain, K. (2015). Young children's comprehension of temporal relations in complex sentences: the influence of memory on performance. *Child Development, 86*, 1922-1934. doi: 10.1111/cdev.12412
- Cain, K. (2006). Individual differences in children's memory and reading comprehension: an investigation of semantic and inhibitory deficits. *Memory, 14*, 553-569. doi: 10.1080/09658210600624481
- Cain, K., & Nash, H. (2011). The influence of connectives on young readers' processing and comprehension of text. *Journal of Educational Psychology, 103*, 429-441. doi: 10.1037/a0022824
- Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Année Psychologique, 114*, 647-662. doi: 10.4074/S0003503314004035
- Carroll, J. B., & White, M. N. (1973). Word frequency and age-of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology, 25*, 85-95. doi: 10.1080/14640747308400325
- Clark, E. V. (1971). On the acquisition of the meaning of before and after. *Journal of Verbal Learning and Verbal Behavior, 10*, 266-75. doi: 10.1016/S0022-5371(71)80054-3
- Coll-Florit, M., & Gennari, S. P. (2011). Time in language: Event duration in sentence comprehension. *Cognitive Psychology, 62*, 41-79. doi: 10.1016/j.cogpsych.2010.09.002

- Cozijn, R., Noordman, L. G. M., & Vonk, W. (2011). Propositional integration and world-knowledge inference: Processes in understanding 'because' sentences. *Discourse Processes, 48*, 475–500. doi:10.1016/j.jml.2011.01.001
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review, 3*, 422-433. doi: 10.3758/BF03214546
- Diessel, H. (2004). *The acquisition of complex sentences in English*. Cambridge: Cambridge University Press.
- Dunn, L. M., Dunn, D. M., Styles, B., & Sewell, J. (2009). *The British Picture Vocabulary Scale III – 3rd Edition*. London: GL Assessment.
- Friend, M., Schmitt, S. A., & Simpson, A. M. (2012). Evaluating the predictive validity of the Computerized Comprehension Task: Comprehension predicts production. *Developmental Psychology, 48*, 136. doi: 10.1037/a0025511
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 177–190. doi: 10.1037/0012-1649.40.2.177
- Gennari, S. P. (2004). Temporal references and temporal relations in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 877–890. doi: 10.1037/0278-7393.30.4.877
- Givón, T. (1991). Isomorphism in the grammatical code: cognitive and biological considerations. *Studies in Language, 15*, 85–114. doi: 10.1075/sl.15.1.04giv
- Gunter, T. C., Wagner, S., & Friederici, A. D. (2003). Working memory and lexical ambiguity resolution as revealed by ERPs: A difficult case for activation theories. *Journal of Cognitive Neuroscience, 15*, 643-657. doi: 10.1162/089892903322307366

- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge, MA: Harvard University Press.
- Juhasz, B.J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, *131*, 684-712. doi: 10.1037/0033-2909.131.5.684
- Just, M., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149. doi: 10.1037//0033-295X.99.1.122
- Just, M. A., Carpenter, P. A., & Wooley, J. D. (1982). Paradigms and Processes in Reading Comprehension. *Journal of Experimental Psychology: General*, *111*, 228-238. doi: 10.1037/0096-3445.111.2.228
- Kidd, E. (2013). The role of verbal working memory in children's sentence comprehension. A critical review. *Topics in Language Disorders*, *33*, 208-223. doi: 10.1097/TLD.0b013e31829d623e
- Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S. A. H., Gustafsson, J. E., & Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science*, *18*, 146-154. doi: 10.1111/desc.12202
- Language and Reading Research Consortium. (2015). The dimensionality of language ability in young children. *Child Development*, *86*, 1948-1965. doi: 10.1111/cdev.12450
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Longman.
- Marinis, T. (2010). Using on-line processing methods in language acquisition research. In: S. Unsworth & E. Blom, (Eds.), *Experimental Methods in Language Acquisition Research* (pp. 139-162). Amsterdam: John Benjamins.

- McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.). *The psychology of learning and motivation Vol. 46* (pp. 155-200). San Diego: Academic Press.
- Möhring, W., Newcombe, N. S., & Frick, A. (2014). Zooming in on spatial scaling: Preschool children and adults use mental transformations to scale spaces. *Developmental Psychology, 50*, 1614-9. doi: 10.1037/a0035905
- Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature, 395*, 71–73. doi: 10.1038/25731
- Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*, 554-566. doi: 10.1037/0022-0663.98.3.554
- Pérez, A. I., Paolieri, D., Macizo, P., & Bajo, T. (2014). The role of working memory in inferential sentence comprehension. *Cognitive Processing, 15*, 405-413. doi: 10.1007/s10339-014-0611-7
- Pickering, S. J., & Gathercole, S. E. (2001). *Working Memory Test Battery for Children*. London: The Psychological Corporation.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-plus (Statistics and Computing)*. New York: Springer.
- Pyykkönen, P., & Järvikivi, J. (2012). Children and situation models of multiple events. *Developmental Psychology, 48*, 521–529. doi: 10.1037/a0025526
- R Development Core Team (2014). *R: A language and environment for statistical computing*. Austria: R Foundation for Statistical Computing Vienna. ISBN 3-900051-07-0.
- Seigneuric, A., Ehrlich, M. F., Oakhill, J. V., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing, 13*, 81–103. doi: 10.1023/A:1008088230941

- Silva, M., & Cain, K. (2015). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. *Journal of Educational Psychology, 107*, 321-331. doi: 10.1037/a0037769
- Slavin, S. (2013). PsyScript (Version 3.2.1) [Computer software]. Lancaster University. Retrieved March 15, 2013. Available from <https://open.psych.lancs.ac.uk/software/PsyScript.html>
- Smith Micro Software (2012). Anime Studio Pro (Version 9.1) [Computer software]. Retrieved December 3, 2012. Available from <http://anime.smithmicro.com/index.html>
- Traxler, M. J., Bybee, M. D., & Pickering, M. J. (1997). Influence of connectives on language comprehension: Eye-tracking evidence for incremental interpretation. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 50A*, 481–497. doi :10.1080/027249897391982
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic.
- Van Dyke, J. A., & Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Language and Linguistic Compass, 6*, 193-211. doi: 10.1002/lnc3.330
- Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition, 131*, 373–403. doi: 10.1016/j.cognition.2014.01.007
- Van Dyke, J. A. & McElree, B. (2006). Retrieval interference in sentence processing. *Journal of Memory and Language, 55*, 157-166. doi: 10.1016/j.jml.2006.03.007

Ye, Z., Kutas, M., St. George, M., Sereno, M. I., Ling, F., & Münte, T. F. (2012). Rearranging the world: Neural network supporting the processing of temporal connectives.

Neuroimage, 59, 3662–3667. doi: 10.1016/j.neuroimage.2011.11.039

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162–185. doi: 10.1037/0033-2909.123.2.162

Appendix

Table A.1

Justification for pruning the non-significant main effect and interactions of age, order and connective on the proportion of correct answers by 3- to 7-year-olds.

Main model	M (b)	SE	z	CI		p(> z)
				2.5%	97.5%	
<i>Fixed effects</i>						
(Intercept)	-0.25	0.22	-1.14	-0.68	0.18	0.25
Four-to-Five	0.64	0.31	2.04	0.02	1.25	0.04
Five-to-Six	1.21	0.33	3.67	0.56	1.85	<0.01
Six-to-Seven	1.21	0.35	3.51	0.54	1.89	<0.01
Order	0.81	0.23	3.49	0.35	1.26	<0.01
Connective	-0.13	0.25	-0.51	-0.61	0.36	0.61
Four-to-Five:Order	0.01	0.34	0.04	-0.65	0.67	0.97
Five-to-Six:Order	-0.27	0.36	-0.74	-0.97	0.44	0.46
Six-to-Seven:Order	0.02	0.39	0.06	-0.74	0.79	0.95
Four-to-Five:Connective	-0.51	0.35	-1.45	-1.21	0.18	0.15
Five-to-Six:Connective	-0.23	0.37	-0.63	-0.95	0.49	0.53
Six-to-Seven:Connective	-0.03	0.39	-0.07	-0.79	0.74	0.95
Order:Connective	0.14	0.29	0.49	-0.42	0.71	0.62
Four-to-Five:Order:Connective	0.29	0.42	0.68	-0.54	1.12	0.49
Five-to-Six:Order:Connective	0.39	0.45	0.86	-0.50	1.28	0.39
Six-to-Seven:Order:Connective	0.25	0.50	0.49	-0.74	1.24	0.62
<i>Random effects:</i>				Variance	SD	
Subject (intercept)				0.80	0.90	
Subject slope 1 order				0.33	0.57	
Subject slope 2 connective				0.54	0.73	

*Note: 1. Bold = predictor is significant at $p < .05$ or better.

2. Number of observations: 3136; groups = 98 subjects

Table A.2*Main effect of age and order on the proportion of correct answers by 3- to 7-year-olds.*

Main model	M (<i>b</i>)	SE	z	CI		p(> z)
				2.5%	97.5%	
<i>Fixed effects:</i>						
(Intercept)	-0.34	0.15	-2.22	-0.64	-0.04	0.03
Four-to-Five	0.47	0.20	2.37	0.08	0.86	0.02
Five-to-Six	1.02	0.21	4.91	0.62	1.43	<0.01
Six-to-Seven	1.25	0.22	5.60	0.81	1.69	<0.01
Order	0.90	0.10	9.12	0.71	1.10	<0.01
<i>Random effects:</i>						
					Variance	SD
Subject (intercept)					0.62	0.79
Subject.1 (slope) order					0.30	0.55

*Note: 1. Bold = predictor is significant at $p < .05$ or better.

2. Number of observations: 3136; groups = 98 subjects

Table A.3

Main effect and interactions of age, order and connective on response times (without square root transformation) to correct answers by 4- to 7- year-olds.

Main model	M (b)	SE	t	CI	
				2.5%	97.5%
<i>Fixed effects:</i>					
(Intercept)	1.66	0.17	9.62	1.32	2.00
Five-to-Six	-0.35	0.24	-1.47	-0.82	0.12
Six-to-Seven	-0.46	0.24	-1.91	-0.93	0.01
Order	0.12	0.16	0.75	-0.19	0.43
Connective	0.56	0.17	3.33	0.23	0.88
Five-to-Six:Order	-0.22	0.22	-1.01	-0.65	0.21
Six-to-Seven:Order	-0.19	0.22	-0.86	-0.62	0.24
Five-to-Six:Connective	-0.41	0.22	-1.89	-0.84	0.01
Six-to-Seven:Connective	-0.60	0.22	-2.72	-1.03	-0.17
Order:Connective	-0.69	0.21	-3.27	-1.10	-0.28
Five-to-Six:Order:Connective	0.44	0.28	1.56	-0.11	1.00
Six-to-Seven:Order:Connective	0.60	0.28	2.10	0.04	1.15
<i>Random effects:</i>				Variance	SD
Subject (intercept)				0.47	0.69
Subject (slope 1) order				0.12	0.35
Subject (slope 2) connective				0.03	0.18
Item (intercept)				0.03	0.18
Residual				1.24	1.11

*Note: 1. Bold = predictor is significant at $p < .05$ or better.

2. Number of observations = 1648 ; groups = 80 subjects and 64 items

Table A.4

Main effect and interactions of age, order and connective on response times (with square root transformation) to correct answers by 4- to 7- year-olds.

Main model	M (b)	SE	t	CI	
				2.5%	97.5%
<i>Fixed effects:</i>					
(Intercept)	1.18	0.06	18.52	1.05	1.30
Five-to-Six	-0.16	0.09	-1.83	-0.33	0.01
Six-to-Seven	-0.24	0.09	-2.75	-0.42	-0.07
Order	0.05	0.06	0.78	-0.07	0.16
Connective	0.22	0.07	3.20	0.08	0.35
Five-to-Six:Order	-0.08	0.08	-1.04	-0.24	0.07
Six-to-Seven:Order	-0.06	0.08	-0.81	-0.22	0.09
Five-to-Six:Connective	-0.16	0.09	-1.82	-0.33	0.01
Six-to-Seven:Connective	-0.23	0.09	-2.62	-0.41	-0.06
Order:Connective	-0.26	0.09	-3.08	-0.43	-0.10
Five-to-Six:Order:Connective	0.14	0.12	1.17	-0.09	0.36
Six-to-Seven:Order:Connective	0.24	0.12	2.05	0.01	0.46
<i>Random effects:</i>				Variance	SD
Subject (intercept)				0.06	0.24
Item (intercept)				<0.01	0.07
Residual				0.21	0.46

*Note: 1. Bold = predictor is significant at $p < .05$ or better.

2. Number of observations = 1648 ; groups = 80 subjects and 64 items