Reading comprehension in elementary school children: cognitive studies of the reader, the text, and the task

Karlsson, A.K.J.

Author: Karlsson, J.A.K.
Title: Reading comprehension in elementary school children: cognitive studies of the reader, the text, and the task
Issue Date: 2021-04-07

Chapter 4

Individual differences in children's comprehension of temporal relations: Dissociable contributions of working memory capacity and working memory updating

This chapter has been published as:
Karlsson, J., Jolles, D., Koornneef, A., van den Broek, P., & van Leijenhorst, L. (2019). Individual differences in children's comprehension of temporal relations: Dissociable contributions of working memory capacity and working memory updating. *Journal of Experimental Child Psychology, 185,* 1-18. doi:10.1016/j.jecp.2019.04.007

Individual Differences in Children's Comprehension of Temporal Relations: Dissociable Contributions of Working Memory Capacity and Working Memory Updating

Abstract

In two experiments, we examined 9- to 12-year-olds'comprehension and processing of two-clause sentences with a temporal connective (*before* or *after*) in sentence-medial or sentence-initial position. We obtained measures of individual differences in Working Memory capacity and Working Memory updating to test their contributions to comprehension. We measured the accuracy of children's responses to the questions "what happened first?" (Experiment 1; N = 74) or "what happened last?" (Experiment 2; N = 50) as well as their sentence reading times. Together, these experiments show continued development of comprehension of temporal relations in children in upper elementary school, and suggest that children's comprehension difficulties (i.e., more comprehension errors and longer reading times) were influenced by clause salience and recency effects rather than sentence chronology or the familiarity of the connective. Our findings are consistent with a memory resource-limited account and suggest that individual differences in WM updating and WM capacity make dissociable contributions to processing and comprehension of sentences with temporal order information.

Keywords: Clause Salience; Temporal Connectives; Recency; Reading Comprehension; Developmental Science

4.1 Introduction

Children in upper elementary school have to read and comprehend two-clause sentences describing a sequence of events on a daily basis. Sentences such as "Write down the answer to the question after you read this paragraph" or "Before you add or subtract you should solve the multiplication" are commonly used to instruct students at school. Cognitive theories of reading comprehension suggest that in order to understand these sentences, readers have to integrate the meaning of the individual words into a coherent mental representation (McNamara & Magliano, 2009; van Dijk & Kintsch, 1983). An essential aspect of creating a coherent representation is the processing of the temporal relations of the events represented in the text (Claus & Kelter, 2006; van den Broek, 1990; Zwaan, 1996). In sentences such as our example sentences above readers must use the temporal connectives *before* and *after* to establish the order in which the events represented in a sentence occurred and integrate this information in their emerging representation of the meaning of the sentence (Mann & Thompson, 1986; Van Silfhout, Evers-Vermeul, & Sanders, 2015). Comprehension of the correct temporal order between events in two-clause sentences is especially challenging because the clauses can occur in two orders. The sentences "You should solve the multiplication before you add or subtract" and "Before you add or subtract you should solve the multiplication" both describe the same order of events in the real world, however events are presented chronologically only in the first sentence.

Experimental studies have found that even though comprehension of spoken two-clause sentences starts developing in preschool and early elementary school (e.g. Blything, Davies, & Cain, 2015; Clark, 1971), comprehension of temporal relations in written two-clause sentences continues to improve throughout elementary school (Cain & Nash, 2011; Pyykkönen & Järvikivi, 2012). Different theoretical accounts have tried to explain these findings, but because the sample and methods used to examine these questions have varied between studies, our understanding of the circumstances that help or hinder comprehension of these type of sentences remains limited. Importantly, different factors might be involved during different stages of development. As prior research has focused mostly on preschoolers and early elementary school children, it is currently unclear what factors are most important during upper elementary school. Moreover, there are large individual differences in comprehension between children. A better understanding of the source of these differences could help educators identify and help those readers that need help the most.

Whereas in younger children semantic factors best explain comprehension difficulties for non-chronological sentences (de Ruiter, Theakston, Brandt & Lieven, 2018), recent studies suggest that a memory capacity-constrained theoretical account (e.g. Just & Carpenter, 1992) best explains performance differences in older children (Blything & Cain, 2016; de Ruiter, Theakston, Brandt & Lieven, 2018; Pyykkönen & Järvikivi, 2012). In the current study, we conduct two experiments and investigate the effect of textual factors (connective position, connective familiarity, and clause salience) as well as individual differences in Working Memory (WM) on the comprehension of temporal relations in 9-12-year-old children. We examine upper elementary school children because WM is still immature in this age range, while the demands on reading comprehension are relatively high. Additionally, prior studies suggest that comprehension of non-chronological sentences is more taxing, even for adults (Münte, Schiltz, & Kutas, 1998; Ye et al., 2012a, 2012b), and still immature in 12-year-olds (Pyykkönen & Järvikivi, 2012). More specifically we aim to further develop theoretical accounts of comprehension of temporal relations in upper elementary school readers by examining the influence of individual differences in both WM capacity and WM updating.

4.1.1 Working Memory and Reading Comprehension

The memory capacity-constrained framework (e.g. Just & Carpenter, 1992) explains difficulties that readers encounter in processing the temporal relations in non-chronological sentences in terms of the demand such sentences makes on their limited WM resources. Even though there are many different definitions of WM (for a review see Cowan, 2017), many converge on the notion that WM can be seen as a complex mental workspace in which one can keep, update, and manipulate information (e.g. Baddeley, 2003). In the context of language comprehension research, a definition of WM often entails both a limited-capacity store that holds information in a heightened state of availability, *and* the processes necessary to update the contents of this storage space, and is measured using complex WM span tasks (see e.g. Carretti, Borella, Cornoldi, & De Beni, 2009; Daneman & Merikle, 1996). WM develops throughout childhood and adolescence (e.g. Diamond, 2013; Gathercole, 1999; Huizinga, Dolan, & van der Molen, 2006). This development is thought to underlie the development of more complex cognitive abilities such as reading comprehension. Syntactically or semantically complex sentences such as two-clause sentences with temporal connectives require the reader to integrate different syntactical structures such as main and subordinate clauses in WM (e.g. Natsopoulos & Abadzi, 1986), and update their mental representation in response to linguistic cues. Comprehension of such sentences will be impeded if (1) the temporal sequence of events cannot be inferred from knowledge schemas that describe the typical order of the events (for example a sentence about a person who brushes her teeth before she goes to bed; e.g. French & Brown, 1977), and (2) if the reader does not have enough WM capacity and updating abilities to resolve the sequence of events based on textual cues. Hence, it could be argued that non-chronological sentences are more difficult compared to chronological sentences because participants have to use their WM to switch the order of events in their emerging mental representation. Crucially, theories within a memory capacity-constrained framework assume that individual differences in WM resources between readers can explain differences in comprehension of sentences containing temporal connectives.

Prior studies in 3-7-year-old children (Blything, Davies & Cain, 2015; Blything & Cain, 2016) and adults (Münte, Schiltz, & Kutas, 1998; Ye et al., 2012a, 2012b) have already shown that comprehension difficulties for non-chronological sentences were more pronounced in individuals with a more limited WM. For example, Blything, Davies and Cain (2015) and Blything and Cain (2016) found that individual differences in 3-7-year-olds comprehension of sentences with the temporal connectives *before* and *after* could be predicted by their performance on a simple WM span task. However, these studies were in younger children and used a listening comprehension, not a reading comprehension task. Furthermore, a number of studies in adults explained individual differences in brain activation in response to chronological and non-chronological sentences in terms of increased demands on WM for non-chronological sentences compared to chronological sentences (Münte, Schiltz, & Kutas, 1998; Ye et al., 2012a, 2012b). However, these studies did not include a behavioral measure of sentence comprehension.

Thus, there is some evidence that individual differences in WM influence comprehension of temporal relations in children and adults. However, the measures of WM that have been used in these prior studies do not allow for very specific conclusions, because they do not allow a disentangling of the roles of the development of WM capacity (the ability to store and process items in WM) and WM updating (the ability to continuously update and monitor the contents of WM). Moreover, to our knowledge, only two studies to date have examined comprehension of temporal relations in upper elementary school (Cain & Nash, 2011; Pyykkönen & Järvikivi, 2012). Cain and Nash (2011) compared 8-year-olds', 10-year-olds', and adults' comprehension of different types of connectives using online and offline tasks and found that 10-year-olds' ability to process and comprehend temporal connectives was still immature. Pyykkönen and Järvikivi (2012) found immature comprehension of two-clause sentences with temporal connectives in children up to 12 years of age. These studies with older children have not directly related comprehension to individual differences in WM. Our aim is to extend these prior studies by examining the relation between sentence comprehension and individual differences in WM in upper elementary school children. Moreover, for the first time, we will differentiate between WM capacity and WM updating abilities**.** Because prior studies in adults have shown that WM capacity and WM updating are not necessarily correlated (e.g. Jaeggi, Buschkuehl, Perrig, & Meier, 2010; but see Schmiedek, Hildebrandt, Lövdén, Lindenberger, & Wilhelm, 2009), and differentiation of WM has proven fruitful to better understand children's language comprehension in prior work (Finney, Montgomery, Gillam, & Evans, 2014). Using both a capacity- and an updating measure might elucidate how different aspects of WM influence comprehension of temporal relations in upper elementary school children.

4.1.2 Textual Characteristics Influencing Comprehension of Temporal Relations

Several textual factors have been found to influence comprehension of two-clause sentences with temporal connectives (e.g. Blything & Cain, 2016; de Ruiter, Theakston, Brandt & Lieven, 2018; Natsopoulos & Abadzi, 1986). In two-clause sentences, chronology results from the interaction of the temporal connective (*after* or *before*) and its position in the sentence (sentence-initial or sentence-medial). Results from previous studies appear contradicting regarding the role of the position of the connective in children. On the one hand, Blything and colleagues (2015) found that 3-7-year-old children who were asked to listen to two-clause sentences with temporal connectives and report the first event, made more errors on sentences with sentence-initial connectives than on sentences with sentence-medial connectives. The authors argued that a sentence-initial connective places a heavier load on comprehenders' WM capacity because the connective has to be kept in mind while attending to further information and can be used to order the events only at the end of the sentence. The additional demands that actively retaining this information places on WM may impoverish comprehension processes (Gibson, 2000; Gibson & Pearlmutter, 1998). On the other hand, Pyykkönen and Järvikivi (2012) found that 8-12-year-old children who were asked to read two-clause sentences with temporal connectives and report the first event, made more errors on sentences with sentence-medial connectives than on sentences with sentence-initial connectives. The authors suggest that sentence-medial connectives force readers to update the temporal order of events in their emerging mental representation in WM. Taken together, although both studies point to differences between conditions in the demands that are placed on children's limited WM resources, one study indicated that sentence-initial connectives are more difficult because they place a heavier load on readers' WM capacity (Blything et al., 2015), whereas the other suggested that sentence-medial connectives are more difficult because they demand readers to update their mental representations (Pyykkönen & Järvikivi, 2012). Critically, both studies did not directly test the relation between comprehension and WM with consideration of the influence of the position of the connectives.

Besides the position of the connectives, differences in familiarity of the connectives *before* and *after* itself could also be relevant. Even though semantic knowledge of temporal connectives seems to develop in early childhood (Blything & Cain, 2016; de Ruiter, Theakston, Brandt & Lieven, 2018; Evers-Vermeul & Sanders, 2009), differences in the use of the connectives *before* and *after* might continue to have an effect. For English speaking children, comprehension of *before* and comprehension of *after* have been suggested to develop differently due to a difference in linguistic complexity (e.g. Clark, 1971). In English the connective *after* is more ambiguous than the connective *before* because it is used in more ways than to indicate temporal relations (e.g. "Ben chased after a dog", or "Ben chased the dog after he played with the cat"). Words that are more ambiguous are typically associated with processing costs (e.g. Eddington & Tokowicz, 2015). Although there is no such difference in Dutch and Dutch children commonly encounter both words, *before* ('voordat') is somewhat more frequent than *after* ('nadat') (Tellings, Hulsbosch, Vermeer, & van den Bosch, 2015). Therefore, we include this familiarity factor of the connectives in our analyses.

Finally, most prior research concerning two-clause sentences with temporal connectives has required participants to indicate which event happened first in a forced-choice task (e.g. Natsopoulos & Abadzi, 1986; Pyykkönen & Järvikivi, 2012; but see Blything & Cain, 2016). In doing so, the answer was always situated in the main clause when the connective *before* was used, and in the subordinate clause when the connective *after* was used. This is important because clause salience could influence comprehension. According to a syntactic account of the processing of complex sentences, certain syntactic configurations facilitate or hinder comprehension. Sentences are thought to be easier to comprehend if they follow a main-subordinate clause order (Diessel, 2005), and readers are more likely to attend to, encode, and remember the information in a main clause than in a subordinate clause (Cooreman & Sanford, 1996; Miltsakaki, 2003; Sanford, 2002). For example, adults are more likely to detect false statements when these false statements are presented in the main clause than when they are presented in the subordinate clause (Baker & Wagner, 1987). Clause effects on comprehension do not depend on extensive communicative experience and are found as early as in preschool. For example, to comprehend the temporal order between events in a listening task, preschool children treat information in the main clause as more important (Trosborg, 1982). This preference for main clauses seems especially strong when children cannot rely on their world knowledge to infer the temporal order (e.g. French & Brown, 1977). Moreover, by asking which event happened first the most recently read event corresponds to the correct answer for non-chronological sentences, not for chronological sentences. This could inflate accuracy scores for non-chronological sentences. Changing the comprehension question to "what happened last?" shifts the types of sentences for which comprehension of temporal relations is facilitated (see Table 4.1a and 4.1b), which allows us to investigate if the effects found in previous studies replicate with different task demands. Blything and Cain (2016) asked their participants to indicate what happened last and found lower accuracy overall compared to previous studies, which they attributed to these task differences. To better understand clause effects and recency effects the current study therefore includes two experiments: in the first experiment children are asked to indicate what happened first and in the second experiment they are asked to indicate what happened last.

In sum, research to date has suggested that children's difficulty in processing non-chronological sentences in upper elementary school can be best explained by a memory capacity-constrained account. However, very few studies to date examined children in upper elementary school, and the studies that did examine this age group did not include all textual factors that could influence online processing of temporal connectives, and did not directly measure WM. We aimed to close this gap in the literature by including measures of individual differences in WM capacity and WM updating in our study of processing and comprehension of two clause sentences with temporal connectives.

4.1.3 The current study

In the current study, we examined 9- to 12-year-old children's comprehension of temporal relations in sentences with two successive events and temporal connectives. We designed a computerized self-paced reading task based on the paradigm used by Pyykkönen and Järvikivi (2012) in which children read two-clause sentences containing the temporal connective *before* or *after,* in sentence initial or sentence medial position. This manipulation caused events to be presented either in chronological or non-chronological order (see Table 4.1). Following each sentence children were shown pictures of both events and were asked to indicate which of these events occurred first (Experiment 1), or last (Experiment 2) (see Figure 4.1). It should be noted that our study differed from the Pyykkönen and Järvikivi study in that the current computerized task did not allow for rereading: on each trial the sentence appeared on the first screen and was then replaced by a second screen with the comprehension question. As a consequence, our task likely puts greater demands on the participants' WM. The present task's demands on

WM are more similar to the listening task used by Blything and colleagues (2015) with preschool children.

Our first hypothesis, in line with previous studies was that 9- to 12-year-old children would perform above chance for all sentence types, reflecting a general understanding of the temporal connectives (Blything & Cain, 2016; Cain & Nash, 2011; Pyykkönen & Järvikivi, 2012). However, in line with previous studies we predicted that non-chronological sentences would be more difficult to process compared to chronological sentences because these sentences place greater demands on the reader's limited WM resources. These difficulties should be reflected in lower accuracy scores and longer reading times.

In addition, the current study is the first that aims to explore the differential roles of WM capacity and WM updating in comprehension of complex sentences with temporal connectives by directly examining individual differences in both types of WM ability. Therefore, we included a measure of WM capacity, a Sentence Span task (Swanson, Cochran, & Ewers, 1989), which measures participants' capacity to keep words in WM while performing a second task (i.e. answering a question), and a measure of WM updating, the Mental Counters task (Huizinga et al., 2006), which measures participants' ability to continuously update and monitor numerical information kept in WM. Following previous studies, our second and third hypotheses were that individual differences in WM updating would be related to comprehension of non-chronological sentences with a sentence-medial temporal connective (Pyykkönen & Järvikivi, 2012), because these sentences require the reader to update the order of information in their evolving mental representation of the sentences. In addition, we hypothesized that individual differences in WM capacity would be related to comprehension of sentences with a sentence-initial temporal connective (Blything et al., 2015) or of sentences in which the correct answer is situated in the subordinate clause, because both these factors place high demands on WM capacity as the reader has to actively retain the information in WM while also attending to other information.

Table 4.1.
Example sentences and comprehension questions for experiment 1 (a) and experiment 2 (b) and their relation to the factors that are hypothesized to facilitate comprehension of temporal relations.

	Main Clause	Connective position	Connective	Chronology
a. *Experiment 1. Comprehension question: "What happened first?"*				
Before Bart ate a cookie, **he drank milk**.	√	–	√	–
Bart ate a cookie before he drank milk.	√	√	√	√
After **Bart ate a cookie**, he drank milk.	–	–	–	√
Bart ate a cookie after **he drank milk**	–	√	–	–
b. *Experiment 2. Comprehension question: "What happened last?"*				
Before **Bart ate a cookie**, he drank milk.	–	–	√	–
Bart ate a cookie before **he drank milk**.	–	√	√	√
After Bart ate a cookie, **he drank milk**.	√	–	–	√
Bart ate a cookie after he drank milk	√	√	–	–

Note. For both experiments correct answers to the comprehension questions are shown in bold in the example sentence for each condition.
"√" indicates facilitation of comprehension and "-" indicate no facilitation by the correct response presented in the main clause, the position of the connective, the connective itself (*before* and *after*), or chronology (chronological vs. non-chronological sentences).

4.2 General Method

4.2.1 Materials

4.2.1.1 Reading Task

Comprehension of temporal relations was measured using a computerized reading task (E-prime version 2.0.8). On each trial, participants were asked to read a two-clause sentence describing two events and to determine which event occurred first (experiment 1) or which event occurred last (experiment 2). Eighty-four unique Dutch sentences were constructed, each representing two events that do not typically occur in a specific order. Thereby they preclude participants from relying on world knowledge so that they had to use the temporal connective to understand the sentences. Sentences contained one of three temporal connectives, *after* ('nadat'), *before* ('voordat'), or *while* ('terwijl'), which were presented in sentence-initial position or sentence-medial position. Our analyses focused on sentences with the connectives *after* and *before,* signaling a sequential order of the two events. The manipulation of these two connectives and their position in the sentence resulted in four sentence types (see Table 4.1a and 4.1b). Sentences with the connective *while,* indicating that events occurred simultaneously, were treated as filler trials. Participants were given 14 trials of each sentence type in a semi-randomized order, assuring the same type of sentences was not presented successively. The order of trials was counterbalanced between participants.

On each trial (see Figure 4.1), the sentence was followed by a screen on which the question "What happened first?" (Experiment 1) or "What happened last?" (Experiment 2) was presented at the top. Underneath the question, three pictures (from Microsoft Office ClipArt 2010) representing the response options were presented (see Figure 4.1). Text (in black), and pictures (in color) were presented on the screen against a light-grey background. The leftmost response option represented both events occurring simultaneously on each trial. The middle and right response options each represented one of the two events from the sentence. The position of the correct response in either the middle or the right picture was counterbalanced within conditions. The task was self-paced, after reading the sentence participants pressed the 'D' key on the keyboard with their left index finger to continue to the question. To choose the picture corresponding to the correct answer the participant pressed the 'A', 'S' or 'D' on the keyboard for response options 1, 2, or 3 with their left ring finger, middle finger or index finger, respectively. Each sentence and question remained on the screen for a maximum of 10 seconds. Between trials a fixation cross was presented for 500 ms.

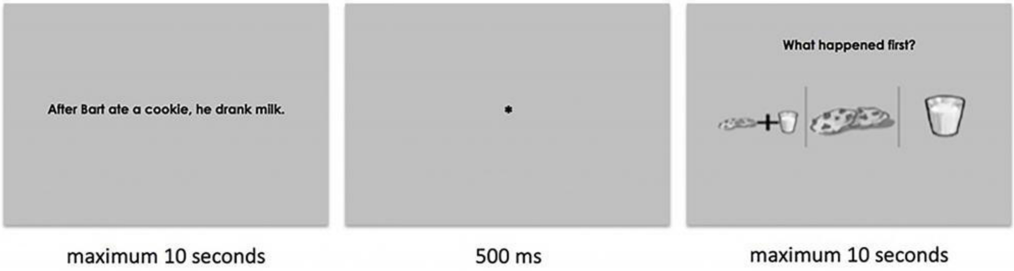

Figure 4.1. Schematic representation of a trial in the reading task. In this example two events (Bart ate a cookie, Bart drank milk) were shown with the sentence initial temporal connective after. Sentence presentation was self-paced with a maximum duration of 10 seconds. The sentence was followed by 500 ms fixation, which was followed by the comprehension question. Participants answered the comprehension question, which was presented at the top of the screen by selecting one of the three images that represent the choice options; "both events happened simultaneously" (which was always presented in the left panel), "Bart ate a cookie" (middle panel), and "Bart drank milk" (right panel). Within each condition, the position of the latter two response alternatives was counterbalanced between the middle and right panel.

A pilot study including 82 typically developing 9-year-old Dutch children (37 girls, M_{age} = 9 years, 4 months, SD_{age} = 4 months) was conducted to assess the age appropriateness of the reading task. Participants in the pilot study received a paper-and-pencil task in which they were asked to indicate the first event in each sentence by circling the correct answer. The overall mean accuracy score was fairly high: 80.17% (*SD* = 17.04), indicating the reading material was suitable for our test population. In addition, to test whether the pictures in the computerized task were easy to interpret a subset of the children who participated in experiment 1 (N = 49; 31 girls, M_{age} = 10 years, 7 months, SD_{age} = 1 year, 2 months) took part in a picture-word-matching test after completion of the main experiment. In this paper-and-pencil test, each of the 84 picture pairs used in the main experiment was shown to the participants, along with the corresponding words, and participants were asked to match each picture to the corresponding word. The overall mean accuracy of the picture-word-matching test was very high: 97.81% (*SD* = 4.92), indicating that the children understood the pictures.

4.2.1.2 Working-Memory Capacity

To assess participants' capacity to store items in WM while performing an interfering task we used a Dutch version of the Sentence Span Task by Swanson, Cochran, and Ewers (1989). The task contained four levels with two trials each. On each trial participants listened to a series of unrelated sentences and were asked to remember the last word of each sentence. After answering a comprehension question about one of the sentences, participants recalled the last word of each sentence. At the easiest level children listened to series of two sentences. At each level one sentence was added, increasing WM load, resulting in series of five sentences at the most difficult level. If the participant made an error on both trials within a level, the test was discontinued. Participants' accuracy scores, the sum of correctly remembered words for trials in which the question was answered correctly, were used in the analysis. This scoring method has gained a good internal consistency of .79 (Conway et al., 2005). The maximum score was 28.

4.2.1.3 Working-Memory Updating

To assess participants' ability to continuously update and monitor the stream of items held in WM we used the Mental Counters task. The Mental Counters task was developed by Huizinga and colleagues (2006) as part of a large executive function battery created to examine developmental changes in these functions. The task contained two sets of 20 trials each. On each trial participants were required to keep track of the score of counters presented visually as a black line on a white computer screen. Blocks appeared above or below the counters indicating if the score rose or fell with one point, respectively. The participant was required to indicate when the score for one of the counters reached above a certain number that was specified before the trial started. In the first set of trials participants had to track two counters, in the second set a third counter was added, increasing WM load. Participants' total accuracy score, the percentage of correct trials, was used in the analysis.

4.2.1.4 Raven Standardized Progressive Matrices

To obtain an estimate of general intellectual ability, participants completed a group-administered version of Raven's Standard Progressive Matrices (SPM: Raven, Raven, & Court, 1998). This measurement was included to test whether participants across the two experiments formed comparable groups of participants. Participants received a booklet with matrices, each one with a missing part, and were instructed to "solve as many puzzles" as possible within 30 minutes. The participants' task was to find the missing part out of six or eight options and write the number of the correct answer on an answering sheet. The items continuously increase in difficulty throughout the test. The estimated IQ scores were determined using international norms (Raven et al., 1998).

4.2.2 Procedure

Both experiments in the study were performed in two sessions. In the first session participants completed the Raven SPM in a classroom setting. In the month following the first session participants were tested individually at their schools. Each participant first performed the reading task, second the WM capacity task, and third the WM updating task. The reading task was explained using Powerpoint slides and six practice trials were completed in E-prime before the test trials began. The WM capacity task was explained verbally, and three practice trials were completed before the test trials began. The WM updating task was explained verbally with the use of visually presented examples, and three practice trials were completed before the test trials began. The reading task and the WM updating task were performed on 15-inch wide-screen Dell Latitude e6530 laptops running on windows 7 and an Intel core i5. The individual session took approximately 80 minutes to complete, including a break. The test battery included two further tasks (one in the first session and one in the individual session) which will, however, not be reported on in the current paper. At the end of the individual session, participants were thanked with a small gift. The study was carried out in accordance with the Ethical Declaration of Helsinki.

4.2.3 Analyses

To assess reading comprehension of temporal relations, sentence reading times and responses to the questions “What happened first?” (Experiment 1) and “What happened last?” (Experiment 2) were collected. The responses to the questions were binary scored as correct or incorrect. Generalized linear mixed-effects regression models were fitted for the accuracy data and linear mixed-effects regression models were fitted for the reading time data. The analyses were conducted with the statistical software R (version 3.5.1; R Core Team, 2018) using the package Lme4 (version 1.1-19; Bates, Mächler, Bolker, & Walker, 2015). Each model included the fixed factors Connective (*before* vs. *after*), Position (the connective appeared sentence-initial or sentence-medial), WM capacity, WM updating, and the full interactional terms of Connective by Position by WM capacity and Connective by Position by WM updating.[1] Participants and items were included as crossed random effects (Baayen, Davidson, & Bates, 2008). Sum coding was applied for the categorical independent variables (the connective *before* was coded

[1] R (Lme4) formula: Dependent variable ~ 1+Connective*Position*Sentence Span+ Connective*Position*Mental Counters+(1 | Participant)+ (1 | Item)

as -0.5 and the connective *after* was coded as 0.5; sentence-initial conditions were coded as -0.5 and sentence-medial conditions were coded as 0.5). The continuous predictors WM capacity and WM updating were centered to the mean. Fixed-effects estimates, standard errors, and the associated t-values (for the continuous dependent variable reading time) and z-values (for the categorical dependent variable accuracy) will be reported. To obtain p-values for the t-statistics, we follow the practice of Barr, Levy, Scheepers, and Tily (2013) and base those values on z-statistics as well.

4.3 Experiment 1

4.3.1 Method

4.3.1.1 Participants

Eighty-two typically developing children (50 girls) between the ages of 9 and 12 years (M = 10 years, 9 months, SD = 1 year, 1 month) were recruited from three primary schools located in middle-class neighborhoods in the Netherlands (Knol, 2012). Inclusion criteria were Dutch as mother tongue, and average to above-average scores on a standardized test on word reading developed by the Dutch Central Institute for Test Development (CITO; Krom, Jongen, Verhelst, Kamphuis, & Kleintjes, 2010). These scores were obtained from the schools. Children's scores on the Raven's SPM were in the average to above-average range (M = 113.13, SD = 11.06). Written parental consent was obtained for all children, and all participating children provided oral assent.

4.3.2 Results

Prior to all analyses, 8 participants' data were removed from the dataset because they were unable to complete one or more tasks (i.e., data was missing for the reading, WM capacity and/or WM updating tasks). For the remaining 74 participants (44 girls) we removed trials with reading times below 1000ms and trials with reading times above 10s. Similarly, trials in which the reaction time to the question was below 200ms or above 10s were removed from the analyses. As a result of this procedure 1.4% of the trials was removed. Table 4.2 and Figure 4.2 (top panel) report the mean accuracy scores (i.e., probability correct), reading times (in ms), and their standard errors (SE) as a function of the factors Connective and Position. On average, children obtained a score of 8.9 (range: 0-21) on the WM capacity task and a score of 81 (range: 58-100) on the WM updating task. The scores for the WM tasks did not show statistically significant correlations (r (72) = .13, p = 0.25).

Table 4.2

Mean accuracy scores (probability correct), reading times (in ms), and their standard errors (SE) in Experiments 1 and 2 as a function of Position and Connective.

		Experiment 1				Experiment 2			
Position	Connective	Accuracy	SE	Reading time	SE	Accuracy	SE	Reading time	SE
Initial	After	.74	.01	5563	65	.80	.02	5934	76
	Before	.93	.01	5100	63	.67	.02	5759	75
Medial	After	.85	.01	5289	63	.81	.02	5815	73
	Before	.93	.01	5201	59	.77	.02	5690	72

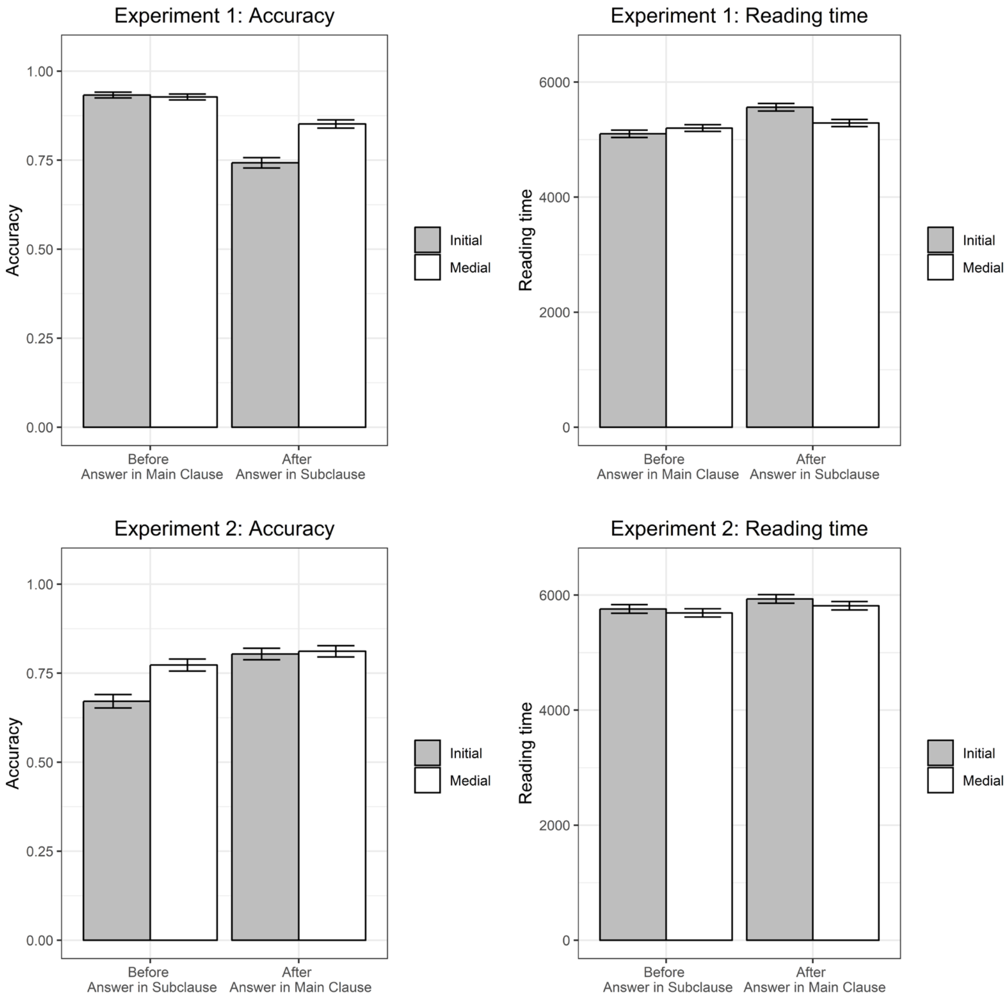

Figure 4.2. Mean accuracy scores (left) and reading times (right) in Experiments 1 (top panel) and 2 (bottom panel) as a function of Connective and Position. Error bars reflect standard errors (SE). Before-Medial and After-Initial sentences are chronological, Before-Initial and After-Medial sentences are non-chronological.

4.3.2.1 Accuracy

Table 4.3 reports the estimates (β-values), and the associated statistics for each fixed effect (excluding the Intercept) of the generalized linear mixed-effects regression model that was fitted to examine children's comprehension of temporal relations in Experiment 1. The analysis revealed a main effect of Connective, WM updating, and a Connective by Position interaction. For the fixed effect of Connective, the estimate carries a negative sign indicating that the accuracy scores were lower for sentences with the connective *after* (where the correct answer was situated in the subordinate clause) than for sentences with the connective *before* (where the correct answer was situated in the main clause). The main effect of WM updating carries a positive sign, indicating that children's comprehension accuracy scores increased as a function of how well they performed on the WM updating task. Figure 4.3 illustrates the Connective by Position interaction. Follow-up analyses (i.e., we fitted identical models, yet dummy-coded the categorical independent variables and adjusted the reference category to examine the relevant simple main effects) showed that there was no effect of Position for sentences with the connective *before* (when the correct answer was situated in the main clause) ($\beta = 0.16$, $SE = 0.32$, $z = 0.51$, $p = .61$). However, for sentences with the connective *after* (when the correct answer was situated in the subordinate clause), accuracy scores were lower for sentence-initial compared to sentence-medial conditions ($\beta = 0.81$, $SE = 0.29$, $z = 2.82$, $p < .01$). Furthermore, the simple main effects of Connective were significant for both sentence-initial and sentence-medial comparisons (*after-initial* vs. *before-initial*; $\beta = 1.91$, $SE = 0.31$, $z = 6.21$, $p <.001$; *after-medial* vs. *before-medial*: $\beta = 0.93$, $SE = 0.31$, $z = 3.05$, $p < 0.01$). To summarize, we did not find lower accuracy scores for non-chronological sentences, by contrast accuracy scores were particularly low for the condition in which the correct answer was situated in the subordinate clause and the connective appeared in the beginning of the sentence, even though in this condition events were presented chronologically.

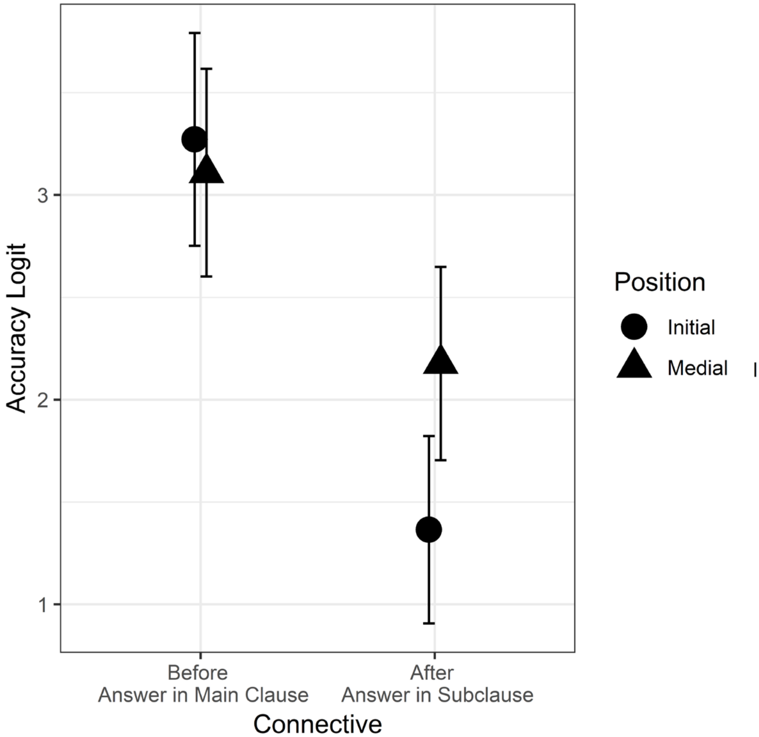

Figure 4.3. Fixed effects estimates (and their 95% confidence intervals) of the accuracy scores (logit scale) in Experiment 1 as a function of Connective and Position.

4.3.2.2 Reading Time

The linear mixed-effects regression analysis for the dependent variable reading time revealed main effects of Connective and WM capacity (Table 4.3). For the fixed effect Connective, the estimate carries a positive sign indicating that the reading times were longer for sentences with the connective *after* (where the correct answer was situated in the subordinate clause) than the reading times for sentences with the connective *before* (where the correct answer was situated in the main clause). The main effect of WM capacity carries a negative sign, indicating that children's reading times decreased as a function of how well they performed on the WM capacity task.

Table 4.3

Fixed effects estimates and associated statistics of the dependent variables in Experiment 1.

	Accuracy				Reading time			
Fixed Effects	*β*	*SE*	*z*	*p*	*β*	*SE*	*z*	*p*
Connective	-1.42	0.22	-6.53	<.01	282.51	134.38	2.10	.04
Position	0.32	0.22	1.50	.13	-76.93	134.38	-0.57	.57
WM capacity	0.04	0.03	1.14	.26	-52.70	24.07	-2.19	.03
WM updating	0.04	0.01	2.84	<.01	-12.97	10.63	-1.22	.22
Connective:Position	0.97	0.43	2.25	.02	-390.62	268.76	-1.45	.15
Connective:WM capacity	0.02	0.03	0.74	.46	-16.29	12.08	-1.35	.18
Position:WM capacity	0.02	0.03	0.60	.55	7.44	12.08	0.62	.54
Connective:WM updating	0.01	0.01	0.77	.44	1.57	5.43	0.29	.77
Position:WM updating	-0.01	0.01	-0.67	.50	-2.53	5.43	-0.47	.64
Connective:Position: WM capacity	-0.02	0.06	-0.43	.67	14.05	24.17	0.58	.56
Connective:Position: WM updating	0.04	0.02	1.63	.10	-2.72	10.85	-0.25	.80
Nr. of observations	3739				3739			
Nr. of participants	74				74			
Nr. of items	56				56			

4.3.3 Discussion

We tested how the temporal connectives *before* and *after* and the position of these temporal connectives affect 9-12-year-olds' comprehension of temporal relations between two events during reading. On average, comprehension scores were high, but there were performance differences between different sentence types. Our results differ from the results of previous studies because we did not find that processing and comprehension of non-chronological sentences was more challenging. By contrast, we found that children were most likely to give a wrong answer for sentences with the connective *after* – in which the correct answer was situated in the subordinate clause – especially when the connective was presented at the beginning of the sentence, even though this is a chronological sentence. Furthermore, reading times were longer for sentences where the correct answer was situated in the subordinate clause. These results indicate that clause salience could better explain task performance than sentence chronology. Interestingly, we found that on average, better WM updating was related to higher comprehension accuracy, whereas higher WM capacity was related to shorter reading times, suggesting that WM updating and WM capacity make separable contributions to sentence comprehension.

Experiment 1 did not allow us to rule out alternative explanations for these findings. First, facilitation of comprehension by the main clause was confounded with the familiarity of the connective. Therefore, our explanation that comprehension may have been compromised because the correct answer was situated in the less salient subordinate clause could be incorrect. An alternative explanation is that comprehension was compromised because the connective *after* is somewhat less common than the connective *before*. Second, our findings could be interpreted as a recency effect. Performance was best in the conditions in which the correct answer corresponded to the most recently read event (*before-initial* and *after-medial* sentences). To disentangle the effects of main clause salience and the familiarity of the connective, and to further examine recency effects on comprehension we conducted Experiment 2.

4.4 Experiment 2

In Experiment 2, we changed the comprehension question from "what happened first?" to "what happened last?". This modification changed the sentence types for which the correct response was facilitated by the main clause. Hence, in Experiment 2 the correct answer was located in the main clause for sentences with connective *after*, instead of the connective *before* as in Experiment 1. If comprehension is facilitated by the effect of salience of information in the main clause rather than the type of connective, one would expect that the sentence type for which most comprehension errors are made should shift from sentences with the sentence-initial connective *after* in Experiment 1 (see Table 4.1a), to sentences with the sentence-initial connective *before* in Experiment 2 (see Table 4.1b). If comprehension is facilitated by a recency effect, one would expect that performance would be better for sentences in which the correct answer corresponds to the most recently read event (*before-medial* and *after-initial* sentences).

4.4.1 Method

4.4.1.1 Participants

Fifty-three typically developing children (33 girls) between the ages of 9 and 12 years (*M* = 11 years, *SD* = 1 year, 1 month) were recruited from two primary schools located in middle-class neighborhoods in the Netherlands (Knol, 2012). Inclusion criteria were Dutch as mother tongue, and an average to above-average score on the standardized measure of word reading (CITO; Krom, Jongen, Verhelst, Kamphuis, & Kleintjes, 2010) provided by the school. Children's scores on the Raven's SPM were in the average to above-average range (M = 111.13, SD = 12.97) and did not differ significantly from the scores of children participating in Experiment 1 (t (130) = 0.94, p = .35). Written parental consent was obtained for all children, and all participating children provided oral assent.

CHAPTER 4

4.4.2 Results

Prior to all analyses, 2 participants' data were removed from the dataset because we were unable to collect the data for the WM tasks. Data from one additional participant was excluded because of failure to comply with the task instructions. For the remaining 50 participants (31 girls) we removed trials with reading times below 1000ms and trials with reading times above 10s. In addition, trials in which the reaction time to the question was below 200ms or above 10s were removed (13% of the trials was removed in total). Mean accuracy scores, reading times, and their standard errors are reported in Table 4.2 and Figure 4.2 as a function of the factors Position and Connective. On average, the children obtained a score of 8.0 (range: 0-21) on the WM capacity task and a score of 86 (range: 65-100) on the WM updating task. These scores did not correlate (r (48) = .17, p = 0.23).

4.4.2.1 Accuracy

The analysis revealed a main effect of Connective, a main effect of WM updating, a Connective by WM updating interaction, and a Connective by Position by WM updating interaction (see Table 4.4). The positive β-value for the fixed effect Connective indicates that the accuracy scores were higher for sentences with the connective *after* (where the correct answer was situated in the main clause) than for sentences with the connective *before* (where the correct answer was situated in the subordinate clause). This replicates the 'main-clause advantage' observed in Experiment 1. The positive β-value for the fixed effect WM updating shows that children's reading accuracy scores increased as a function of how well they performed on the WM updating task. Figure 4.4 depicts the significant two- and three-way interactions. The graph on the left shows that children with a higher score on the WM updating task are more sensitive to the main-clause advantage than children with a lower score on the WM updating task. However, as can be observed in the dual graph on the right (Figure 4.4), this effect interacts with the position of the connective as well. Two series of follow-up analyses were conducted to further interpret this three-way interaction.

First, we examined Position by WM updating interactions separately for the connectives *after* and *before* (i.e., we fitted models containing all the fixed effects of the original model, yet dummy coding was applied to obtain the relevant simple main effects). These analyses revealed that WM updating significantly interacts with the position of the connective *after* ($\beta = 0.052$, $SE = 0.019$, $z = 2.70$, $p < .01$), but not with the position of the connective *before* ($\beta = 0.016$, $SE = 0.016$, $z = 0.98$, $p = .33$). Moreover, only the *after-medial* condition showed a significant (positive) effect of WM updating, indicating that WM updating positively affects comprehension of a non-chronological sentence with a medial connective ($\beta = 0.083$, $SE = 0.020$, $z = 4.13$, $p < .001$; for all other conditions $z < 1.85$).

Second, to examine more directly how the factors Connective and Position affected children with lower and higher scores on the WM updating task, we adjusted the baseline value of this continuous predictor to the 1st (low) and 9th (high) decile (see vertical dotted lines in Figure 4.4). The analyses for children with a higher score on the WM updating task revealed a main effect of Connective only, indicating that children with a higher WM updating score performed better for sentences with the connective *after* (when the answer was situated in the main clause) ($\beta = 0.94$, $SE = 0.25$, $z = 3.75$, $p < 0.01$). The analyses revealed a different picture for children with a lower score on the WM updating measure. For these children there were no main effects of Connective and Position, yet the Connective by Position interaction was significant ($\beta = 1.32$, $SE = 0.50$, $z = 2.62$, $p < .01$). More specifically, for sentences with the connective *before* (when the correct answer was situated in the subordinate clause), performance was better for sentences with a sentence-medial connective than for sentences with a sentence-initial connective ($\beta = 0.76$, $SE = 0.35$, $z = 2.17$, $p = .03$). Numerically, the opposite pattern was present for sentences with the connective *after* (when the correct answer was part of the main clause), but this main effect was not significant ($\beta = 0.56$, $SE = 0.36$, $z = 1.54$, $p = .12$).

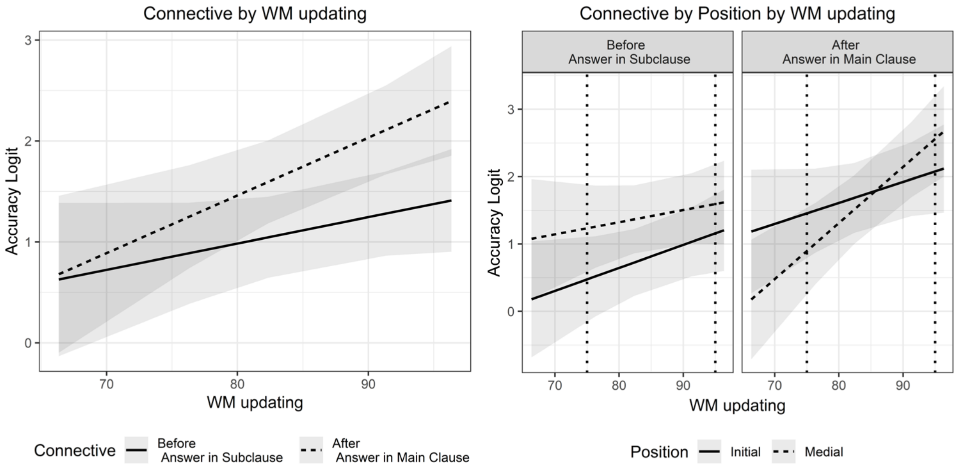

Figure 4.4. Fixed effects estimates (and their 95% confidence intervals) of the accuracy scores (logit scale) in Experiment 2 as a function of Connective, Position, and WM updating. The single graph on the left depicts the two-way interaction of Connective by WM updating and the dual graph on the right depicts the three-way interaction of Connective by Position by WM updating. The vertical dotted lines in the dual graph reflect WM updating scores at the 1st and 9th decile. Before-Medial and After-Initial sentences are chronological, Before-Initial and After-Medial sentences are non-chronological.

4.4.2.2 Reading Time

The analysis revealed a main effect of WM capacity, a Position by WM capacity interaction, and a Position by WM updating interaction (see Table 4.4). The main effect of WM capacity carries a negative sign, indicating that children's reading times decreased as a function of how well they performed on the WM capacity task. Figure 4.5 depicts the two-way interactions and reveals a clear picture. The WM capacity score is a better predictor of the reading times (i.e., a higher WM capacity score is associated with shorter reading times) when the connectives appear sentence-initially than when they appear sentence-medially (see graph on the left in Figure 4.5). The opposite holds for the WM updating task. The score for this task is a better predictor of the reading times when connectives occur sentence-medially, as opposed to sentence-initially (see graph on the right in Figure 4.5).

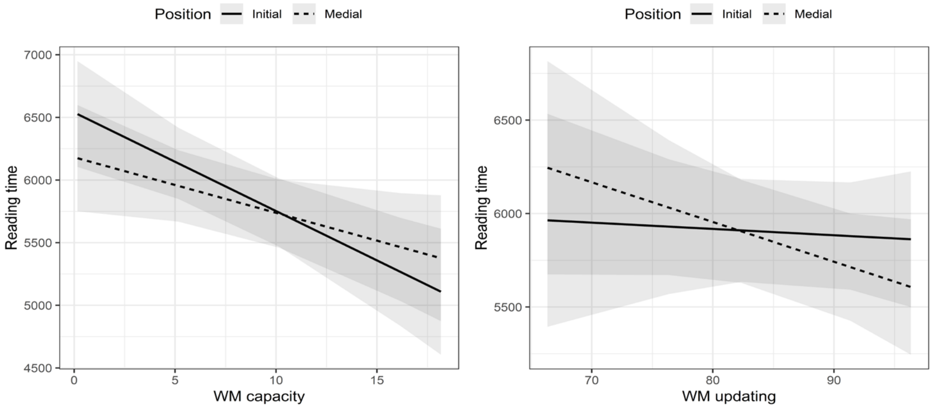

Figure 4.5. Fixed effects estimates (and their 95% confidence intervals) of the reading times in Experiment 2 as a function of WM capacity, WM updating, and Position. The graph on the left depicts the two-way interaction of WM capacity by Position and the graph on the right depicts the two-way interaction of WM updating by Position.

Table 4.4

Fixed effects and associated statistics of the dependent variables in Experiment 2.

	Accuracy				Reading time			
Fixed Effects	*β*	*SE*	*z*	*p*	*β*	*SE*	*z*	*p*
Connective	0.67	0.22	3.09	<.01	151.59	124.85	1.21	.22
Position	0.30	0.22	1.40	.16	-77.03	124.83	-0.62	.54
WM capacity	0.03	0.03	1.13	.26	-61.62	20.73	-2.97	<.01
WM updating	0.04	0.02	2.58	.01	-12.26	12.37	-0.99	.32
Connective:Position	-0.54	0.43	-1.25	.21	-71.17	249.70	-0.29	.78
Connective:WM capacity	-0.02	0.02	-1.09	.27	-11.77	13.36	-0.88	.38
Position:WM capacity	-0.03	0.02	-1.51	.13	34.34	13.36	2.57	.01
Connective:WM updating	0.03	0.01	2.45	.01	13.45	8.10	1.66	.10
Position:WM updating	0.02	0.01	1.42	.16	-18.01	8.11	-2.22	.03
Connective:Position: WM capacity	0.01	0.04	0.15	.88	-13.69	26.72	-0.51	.61
Connective:Position: WM updating	0.07	0.03	2.69	.01	-29.98	16.21	-1.85	.06
Nr. of observations	2443				2443			
Nr. of participants	50				50			
Nr. of items	56				56			

4.4.3 Discussion

The results of Experiment 2 replicated the results from Experiment 1. As in experiment 1, comprehension scores were high on average, but there were performance differences between different sentence types. Accuracy was a little lower in the second experiment. This is consistent with previous reports of slightly worse comprehension in response to the question "what happened last" reported by Blything and Cain (2016) in younger children. This increase in task difficulty may have caused individual differences between children to be more clear, and Experiment 2 revealed a more nuanced picture of how WM updating and WM capacity are differently related to processing and comprehension of two-clause sentences with temporal connectives.

We again did not find a clear effect of sentence chronology, in contrast, we replicated our finding that 9-12-year-old children were least likely to give the correct answer when it was presented in the subordinate clause. Hence, our prediction for Experiment 2 that the most difficult sentence type would shift and involve the connective *before* rather than *after* was correct. We did not find support for the explanation that a difference in familiarity between (the Dutch equivalents of) the words *before* and *after* affects comprehension in upper elementary school children (Blything & Cain, 2016; de Ruiter, Theakston, Brandt & Lieven, 2018; Evers-Vermeul & Sanders, 2009). Instead, our results support the interpretation that task-relevant information presented in the main clause facilitates comprehension across experiments. This finding is consistent with the interpretation that in older children comprehension of sentences containing temporal connectives cannot be explained by semantic differences between these sentences (Clark, 1971), and fits with a memory-capacity constrained framework (Just & Carpenter, 1992).

Interestingly, interactions with WM updating suggest comprehension differences between children as a function of their WM updating abilities. Taken together, the follow-up analyses for the interaction of the position of the connective with WM updating suggest that**:** First, WM updating is particularly relevant when the connective is in a sentence-medial position *and* the correct answer is part of the main clause. Second, children with a higher WM updating score show a main-clause advantage (i.e., they perform better when the correct answer is situated in the main clause). Third, children with a lower WM updating score do not display a main-clause advantage. Instead, they seem to perform better on chronological sentences, when the sentence structure follows the temporal sequence of events in the real world (which is the case in the *before-medial* and *after-initial* conditions, but not in the *before-initial* and *after-medial conditions*). At first glance this seems like an effect of sentence chronology, however it could also be interpreted as a recency effect. In sentences with a *before-medial* and *after-initial* connective the most recently read information is also the correct answer to the question "What happened last?".

Individual differences in WM ability were most apparent in our reading times results. While accuracy scores varied as a function of WM updating, reading times varied as a function of WM updating and WM capacity. Results showed both an interaction between the position of the connective and WM capacity and between the position of the connective and WM updating. WM capacity predicts reading times of sentences with sentence-initial connectives, supporting the hypothesis that these sentences place high demands on WM capacity as the reader has to actively retain the information in WM while also attending to other information. WM updating, on the other hand, predicts reading times of sentences with sentence-medial connectives, supporting the hypothesis that these require the reader to update the order of information in the evolving mental model of the sentence.

4.5 General Discussion

The aim of the two experiments in this study was to examine upper elementary school children's comprehension of the temporal relations between two events while reading two-clause sentences. As expected, sentence comprehension was relatively good for these 9-12-year-old children in both experiments. Even though our main results did not replicate previous findings that sentence chronology is an important factor influencing comprehension, our results are consistent with a memory capacity-constrained account of processing of these types of sentences in children (Blything & Cain, 2016; Blything, Davies & Cain, 2015) and adults (Münte, Schiltz, & Kutas, 1998; Ye et al., 2012a, 2012b). More specifically, we found that text factors that facilitate processing by reducing the demands on WM resources resulted in better comprehension. Interestingly, WM updating and WM capacity contributed differently to task performance. The subtle differences in accuracy between the two experiments, together with the finding in Experiment 2 that children with relatively poor WM updating abilities were most sensitive to task effects underline the importance of examining the interaction between textual factors and reader characteristics that influence comprehension in older children. In addition, our findings suggest that in upper elementary school comprehension of temporal relations in complex sentences is still not fully proficient, which fits with the results reported by Pyykkönen and Järvikivi (2012).

Children's comprehension of temporal connectives in our experiments was best explained by the effect of clause salience and the recency of the information that was the correct answer to the question in each task. Information in the subordinate clause proved to be especially difficult for children to comprehend. This finding is in line with previous studies in both adults (e.g. Baker & Wagner, 1987) and children (French & Brown, 1977; Trosborg, 1982), and could suggest that this syntactic factor contributes to comprehension in older children. However, it could also be interpreted in light of the demands these different types of sentences make on WM resources. In the current study the subordinate clause always starts with a temporal connective. As argued above, processing temporal relations and temporal connectives is an essential aspect of creating a coherent representation of a text (e.g. Claus & Kelter, 2006; Mann & Thompson, 1986; van den Broek, 1990; Van Silfhout et al., 2015; Zwaan, 1996). However, a connective may also be interpreted by a reader as a signal that the main idea is elsewhere and cause them to allocate more attention, and commit to deeper processing of the main clause rather than the subordinate clause (e.g. Sanford, 2002). In this context our finding that children with relatively good and poor WM updating abilities seem to approach these sentences differently is intriguing and adds new information to the literature on sentence comprehension in childhood. Children with good WM-updating abilities are sensitive to facilitating effects of the main clause and perform better when the correct answer coincides with the main clause, whereas comprehension in children with poor WM-updating abilities resembles the finding in younger children. For them chronological sentences seem easier to comprehend (e.g. Blything & Cain, 2016). However, the alternative explanation that this chronology effect is in reality a recency effect could also be true. Even though chronology and recency are confounded in Experiment 2, they are not in Experiment 1. Taken together, the results from both experiments suggest that recency effects might be a better explanation for the processing and comprehension of temporal connectives, at least in older children. Future work should explore these alternative explanations, as well as examine possible task-related effects.

Reading times largely mirrored the accuracy results, consistent with the interpretation that better comprehension is a consequence of reduced demands on WM resources. Thus, our reading times results support a memory capacity-constrained account as well. In both experiments better WM capacity was related to faster processing of sentences. In Experiment 2 – where slightly lower comprehension scores might have allowed us to better capture individual differences – WM capacity and WM updating both interacted with the position of the connective. On the one hand a bigger WM capacity facilitated processing of sentences with sentence-initial connectives, while on the other hand better WM updating facilitated processing of sentences with sentence-medial connectives. These findings reconcile previous seemingly contradictory findings that sentence-initial connective place high demands on WM resources because they require the reader to keep this information in mind while processing incoming information (Blything et al., 2015), and that sentence-medial connectives place high demands on WM resources because these require the reader to update the order of information in the evolving mental model of the sentence in (Pyykkönen & Järvikivi, 2012). These findings demonstrate the importance of using several measures of WM in developmental studies to unravel the relation between reading comprehension processes and individual differences.

Our study focused on 9-12-year-old readers in upper elementary school. In doing so we extended the literature by showing that in these children comprehension of connectives is not fully proficient which suggests that it continues to develop throughout adolescence. However, it should be noted that to further develop theoretical accounts of developmental change in readers comprehension of temporal relations, future studies should examine the influence of individual differences in both WM capacity and WM updating across a broader age range. In addition, in the context of this experiment readers encountered complex sentences with temporal connectives in isolation. It could be argued that demands on WM are different when these sentences are encountered in a text. Furthermore, we did not manipulate WM load directly in our experiment. An important next step would be to examine the effects of individual differences in response to changes in the demands made on WM capacity and WM updating abilities by the reading task. Nevertheless, our findings are an important initial step.

In conclusion, our findings show continued development of comprehension of temporal connectives in children in upper elementary school. Our findings are consistent with a memory resource-limited account and suggest that individual differences in WM updating and WM capacity make dissociable contributions to processing and comprehension of sentences with temporal order information. Therefore, our findings contribute to the further refinement of models of the development of the ability to comprehend temporal connectives during reading. It could be argued that the sentences that children encounter in everyday life are more complex than the sentences that we used in the present study. However, our findings suggest that even under these relatively simple circumstances comprehension of sentences with temporal connectives continues to show improvements until at least the age of 12.